11 (2012)

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ ПРИНЦИПА ПРЕДЕЛЬНЫХ ОБОБЩЕНИЙ

УДК 519+61:681.3

ПРОКОПЧУК Юрий Александрович
к.ф.-м.н., доцент, с.н.с. отдела системного анализа и проблем управления Института технической механики НАНУ и НКАУ.
Научные интересы: интеллектуальные и когнитивные системы, базы знаний.
e-mail: itk3@ukr.net

ВВЕДЕНИЕ И ПОСТАНОВКА ЗАДАЧ ИССЛЕДОВАНИЯ

Большинство сложных процессов обладают объективной неопределенностью и слабой формализацией, что обуславливает непрерывное развитие инструментария анализа и прогностики временных рядов (ВР). Наряду с традиционными методами (статистическими, нейросетевыми) все чаще используются интеллектуальные методы, которые в последние десятилетия оформились в виде отдельного направления, называемого Times-Series Data Mining. Большое развитие получили методы нечеткой регрессии, анализа данных нечетких временных рядов [1, 2].

Несмотря на значительное развитие указанных выше методов, многие важные вопросы остаются открытыми. Ясно, что для разных заключений, формируемых в результате анализа/прогноза, определяющую роль могут играть разные первичные и производные параметры многомерного ВР, измеряемые с разной степенью точности (заранее неизвестной). Вместе с тем статистические, нейросетевые и нечеткие методы фактически работают с одним уровнем общности описания ВР. Каким образом можно учесть в моделях множество уровней общности описания первичных и производных значений ВР? Существуют ли критические наброски ВР, которые минимально достаточны для решения целевой задачи? Нераскрыты процессы *самоорганизации* при формировании результирующих моделей интерпретации ВР. Требуемые модели, по-видимому, должны появляться в результате когнитивной самоорганизующейся критичности, следовательно, они являются существенно неравновесными моделями (система мониторинга открытая). Как запустить процессы самоорганизации моделей? Как формируется «опыт» решения целевых задач? Важным вопросом является обеспечение высокой *пластичности* решения целевых задач (в условиях возможной неполноты априорных данных). Эти и другие вопросы разрешаются в рамках применения Принципа предельных обобщений (ППО) [3].

Задачами настоящего исследования являются:

– выяснение сути когнитивного подхода к анализу ВР;

– описание схемы применения ППО к анализу и прогнозированию (многомерных) ВР;

– демонстрация с помощью программных прототипов возможности реализации предложенных структур в рамках интеллектуальных приложений.

ОНТОЛОГИЯ КОГНИТИВНОГО ПОДХОДА К АНАЛИЗУ ВР

При изучении феноменологического пространства наблюдателя базовым предположением является предположение о существовании единого *универсального принципа структурирования информации* (этот принцип закладывался в "устройство" мышления с эволюционным возникновением самого мышления). Именно благодаря этому общему принципу происходит сопряжение (непосредственное или опосредованное) любой мыслительной структуры с любой другой, и любого уровня информационной масштабности. Следствием универсального принципа структурирования информации являются модели двух взаимосвязанных информационных объектов: *орграфа набросков* (образов, моделей, ситуаций) и *динамического системопаттерна* (или сокращенно — *системопаттерна*). Разновидностью орграфа набросков является орграф доменов элементарного теста (домен — множество значений теста со связями наследования) [3].

Орграфы набросков разных типов и системопаттерны позволяют находить предельные по уровню обобщенности когнитивные структуры, однозначно решающие целевые задачи. Кроме того, орграфы набросков приводят к суперпозиции смыслов в интерпретации образов, а также квантово-семантической запутанности разных образов. В этом заключается суть «Принципа предельных обобщений» [3]. Основная гипотеза состоит в том, что ППО олицетворяет «встроенную» *оптимальность мышления*. Эволюционная обоснованность гипотезы заключается в стремлении когнитивной системы к глобальному энергетическому минимуму: если система оперирует меньшим адекватным объемом информации, ей требуется для этого меньше энергии, памяти и она может это делать с большей скоростью. Обсуждаемые теоретические представления созвучны с развиваемыми нейрофизиологами положениями об иерархическом принципе интеграции скоростных, медленных и сверхмедленных информационно-управляющих систем головного мозга, формируемых для обеспечения психических состояний и познавательной деятельности человека, включая вербальную ассоциативно-мыслительную деятельность [5]. Рассмотрим общую схему применения ППО к анализу (многомерных) ВР.

Пусть $\{\tau\}$ — множество элементарных тестов, с помощью которых описывается любая ситуация действительности, включая многомерный ВР. Элементарность теста означает, что его результат может быть представлен в виде «тест = значение». Конкретный результат теста τ будем обозначать $\underline{\tau}$. Значения тестов могут выбираться из разных доменов. Для фиксации того, что в качестве множества результатов теста τ используется домен T, будем использовать нотацию: $\underline{\tau}/T$. Используя разные домены, можно управлять общностью (масштабом) результата одного и того же теста. Правила пересчета значений теста из одного домена в другой задает *ориентированный граф доменов:* $G(\tau) = \{T \rightarrow T'\}_{\tau}$. Банк тестов представляет собой $\{G(\tau)\}$. Он может служить системой координат феноменологического пространства. Для элементов обобщающих доменов могут быть заданы функции принадлежности, что позволяет рассматривать нечеткие временные ряды [3]. Таким образом, в зависимости от контекста задачи на основе одного и того же банка тестов $\{G(\tau)\}$ могут быть реализованы нечеткие, интервальные, лингвистические, фрактальные, гранулярные операции.

Событием называется кортеж $<\underline{\tau}/T, \underline{t}/\Lambda>$, где $\underline{t}/\Lambda$ – время, задаваемое орграфом $G(t)$. Примеры доменов Λ:

$\Lambda 1$= 'Дата: время'; $\Lambda 2$= 'Дата: {утро, день, вечер, ночь}'; $\Lambda 3$ = 'Дата'; $\Lambda 4$ = 'Месяц, Год'; $\Lambda 5$ = 'Год'.

Примеры значений времени:

$\underline{t}/\Lambda 1$ = '10.01.09: 08.30', $\underline{t}/\Lambda 2$ = '10.01.09: утро', $\underline{t}/\Lambda 3$ = '10.01.09', $\underline{t}/\Lambda 4$ = 'январь, 2009', $\underline{t}/\Lambda 5$ = '2009'.

Вполне очевиден способ задания пересчета значений из одного домена в другой (с большим номером), т.е. определен орграф доменов: $G(t)=\{\Lambda 1 \rightarrow \Lambda 2 \rightarrow \Lambda 3 \rightarrow \Lambda 4 \rightarrow \Lambda 5\}$. Пример другого орграфа доменов $G(t)$:

$\Lambda 1$= 'Дата: время'; $\Lambda 2$= {1; 2;…; 24}; $\Lambda 3$= {1; 2;…; 7}. $G(t)=\{\Lambda 1 \rightarrow \Lambda 2; \Lambda 1 \rightarrow \Lambda 3\}$.

Правила пересчета: $\Lambda 1.t \rightarrow \Lambda 2.Hour(t)$ –текущий час; $\Lambda 1.t \rightarrow \Lambda 3.DayOfWeek(t)$ – текущий день недели. Можно объединить оба орграфа в один орграф.

Если через $\{p/P\}$ обозначить дополнительные параметры – уточняющие тесты, то расширенная трактовка элементарных событий примет вид составных событий: $e_1 = <\underline{\tau}/T, \{\underline{p}/P\}, \underline{t}/\Lambda>$. Например, если событие локализовано к какой-либо точке пространства $\{x/X, y/Y, z/Z\}$, то можно записать: $e = <\underline{\tau}/T, \{\underline{x}/X, \underline{y}/Y, \underline{z}/Z\}, \underline{t}/\Lambda>$.

Протяженным событием называется событие $e = <\underline{\tau}/T, \underline{\delta}/\Lambda>$, где $\underline{\delta}$ – произвольный временной интервал (возможно, не односвязный). Пример протяженного события: <Креатинин / {мг / 100 мл}? 1.3, [12.02.09; 14.02.09]/$\Lambda 3$>, где δ= [12.02.09; 14.02.09] (в терминах домена $\Lambda 3$). Ясно как определить нечеткое событие и нечеткое протяженное событие.

Произвольный временной ряд $\varphi(t)$ в дискретные моменты времени представим в виде множества событий $\{<\underline{\varphi}, \underline{t}>\}$, а учитывая, что значения теста φ обобщаются с помощью орграфа доменов $G(\varphi)$, можно записать: $\{<\underline{\varphi}//\Phi,\ \underline{t}/\Lambda>\}$ – обобщенный временной ряд. Моменты времени могут входить в название теста, поэтому любой временной ряд представим в виде множества $\{\underline{\tau}/T\}$. Многомерный временной ряд обозначим $\{\varphi(t)\}$. Для любого ВР $\{<\underline{\varphi}//\Phi,\ \underline{t}/\Lambda>\}$ может быть выполнена грануляция (разбиение или покрытие) в виде множества протяженных событий - тенденций: $\{<\underline{\psi}//\Psi,\ \underline{\delta}/\Lambda>\}$. Грануляция может быть выполнена нечеткими тенденциями [1]. Важно отметить, что грануляция может проводиться разными способами. Многообразие возможностей отражает множественность точек зрения на один и тот же процесс.

Рассмотрим пример. Зададим орграф доменов теста «Скорость» следующим образом (в формате конфигуратора [3]):

v – Скорость {

V5 #V2 {Большая ^c; *не* Большая ^a b}

V4 #V2 {Средняя ^b; *не* Средняя ^a c}

V3 {Маленькая ^a; *не* Маленькая ^b c}

V2 {Маленькая ^a [0; 3,0]; Средняя ^b (3,0; 7,0]; Большая ^c (7,0; 10,0]}

V1 {[0; 10,0]}

Орграф $G(v) = \{V1 \rightarrow V2 \rightarrow V3;\ V2 \rightarrow V4;\ V2 \rightarrow V5\}$.

Домен V1 является базовым (наиболее точным). Все остальные домены являются набросками базового домена. Орграф $G(v)$ является структурно-завершенным орграфом, в котором вершины V3, V4, V5 – домены-листья (рис. 1).

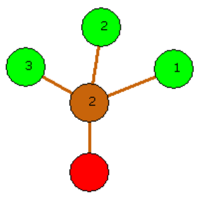


Рисунок 1 – Орграф доменов теста «Скорость»

Исходный временной ряд $v(t)$ зададим в терминах базового домена:

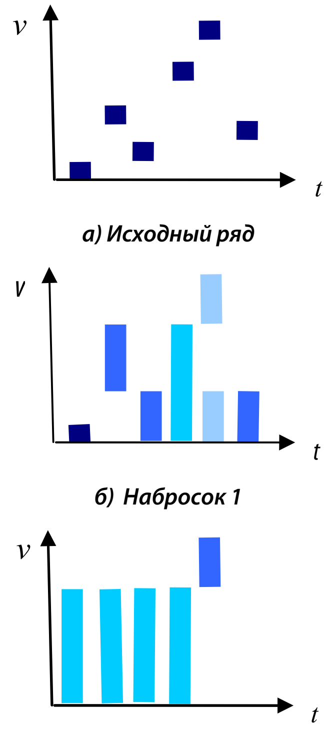
$v(t) = \{v(1)?1; v(2)?4; v(3)?2; v(4)?6; v(5)?8; v(6)?3\} \equiv \{<\underline{v}/V1,\ \underline{t}/\Lambda 1>\} \equiv \{<v/V1?1,\ t/\Lambda 1?1>;\ <v/V1?4,\ t/\Lambda 1?2>;\ <v/V1?2,\ t/\Lambda 1?3>;\ <v/V1?6,\ t/\Lambda 1?4>;\ <v/V1?8, t/\Lambda 1?5>; <v/V1?3, t/\Lambda 1?6>\}$.

Если время оставить неизменным, то ряд скоростей из шести значений можно представить 30-ю способами: $6 \times |G(v)| = 6 \times 5 = 30$. Приведем примеры двух обобщенных рядов (опуская время):

Набросок 1: <*v*/V1?1; *v*/V2?Средняя; *v*/V3?Маленькая; *v*/V5?*не* Большая; *v*/V4? *не* Средняя; *v*/V2? Маленькая>;

Набросок 2: <*v*/V5? *не* Большая; *v*/V5? *не* Большая; *v*/V5? *не* Большая; *v*/V5?*не* Большая; *v*/V5? Большая; *v*/V5? *не* Большая>.

Примеры двух обобщений одного и того же исходного ряда изображены на рис. 2 (чем темнее цвет, тем точнее значение ВР).



а) Исходный ряд

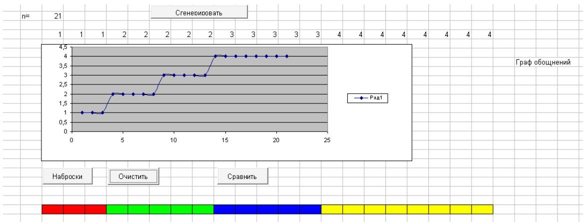
б) Набросок 1

в) Набросок 2

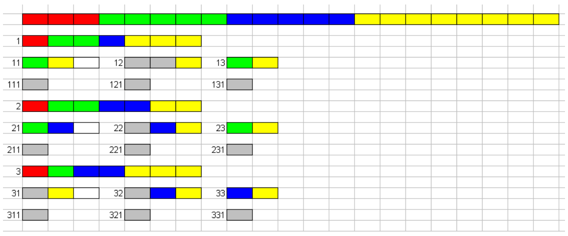
Рисунок 2 – Наброски временного ряда

Двадцать девять вариантов обобщения исходного временного ряда являются его набросками. Все 30 вариантов образуют одну из разновидностей орграфа набросков $Gs(v(t))$, которая опирается на автоматизмы обобщений в рамках орграфа доменов $G(v)$ [3]. Если учесть орграф $G(t)$, то пространство обобщающих описаний (набросков) значительно расширяется. Пример события: <v/V2? Средняя, $\underline{t}/\Lambda 2$? 10.01.09: утро>. Отметим, что при обобщении по времени может появиться неопределенное значение теста (морфология), например: <$\underline{\varphi}//\Phi$? Неопределенное, $\underline{t}/\Lambda 3$>.

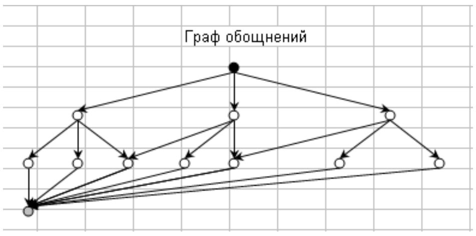
Помимо орграфа набросков $Gs(\varphi(t))$ на основе банка тестов для любого ВР могут строиться орграфы набросков $Gs_{\Sigma}(\varphi(t))$ на основе разнообразных схем Σ «грануляции – сжатия» (рис. 3, схема $\Sigma[3:2]$, серый цвет – морфология) [3]. К числу таких схем относятся и схемы с выделением *тенденций* (об этом речь пойдет ниже).



а) Задание и кодирование цветом ВР



б) Построение набросков ВР



в) Орграф набросков ВР

Рисунок 3 – Наброски временного ряда на основе схемы «грануляция-сжатие»

В работе [3] показано, что в рамках заданного контекста задачи (базы прецедентов и банка тестов) на основе орграфов набросков определяются экстремальные пограничные слои набросков, котрые содержат минимальный объем информации достаточный для однозначного решения целевой задачи.

Обозначим через $Z=\{1,\ldots, N\}$ множество заключений, имеющих отношение к временному ряду $\varphi(t)$ или многомерному временному ряду $\{\varphi(t)\}$. Примеры Z: Z = {Норма; Отклонение}; Z = {Хаотический режим (осцилляция); Стабильный режим; Качественная прерывность (разрушение)}; Z = {Рост; Падение; Стабилизация; Колебания; Хаос} – тенденции ВР; Z = {Резкий рост; Рост; Стабильность; Падение; Резкое падение} - тенденции линейных трендов; Z = {Низкий риск (осложнения, банкротства, аварии, отказа); Средний риск; Высокий риск}; $Z = \Phi$, где Φ – домен из $G(\varphi)$.

АНАЛИЗ/ПРОГНОЗИРОВАНИЕ ВР НА ОСНОВЕ ПРЕДЕЛЬНЫХ МОДЕЛЕЙ ЗНАНИЙ

Рассмотрим следующую задачу: для заданного ВР $\varphi(t)$ или многомерного ВР $\{\varphi(t)\}$ определить заключение $\underline{z}/Z$.

Данную задачу будем решать путем предварительного построения предельных синдромных и вероятностных моделей знаний. Предельные модели знаний применяются для решения исходной задачи. Опишем кратко процедуру построения предельных моделей [3, 4].

Пусть $\Omega(\{\tau/T\}) = \{\alpha(\{\underline{\tau}/T\}, \underline{z}/Z)\}$ – множество ситуаций действительности (база прецедентов) с известными заключениями. Частным случаем базы прецедентов является множество ситуаций $\{\alpha(\{\varphi(t)\}, \underline{z}/Z)\}$ или $\{\alpha(\{<\underline{\varphi}//\Phi, \underline{t}/\Lambda>\}, \underline{z}/Z)\}$. Будем говорить, что множество Ω не содержит *конфликтов* на уровне общности $\{\tau/T\}$, если нет двух ситуаций с разными заключениями, но совпадающими значениями тестов (совпадающими значениями временных рядов).

Под *формальным синдромом* (или просто *синдромом*) понимается неизбыточная совокупность значений тестов, позволяющая однозначно установить заключение: $S = (\{\underline{\tau}/T\} \rightarrow \underline{z}/Z)$. С каждым синдромом S связано «облако» предельных синдромов $\{S^*\}$, которое получается путем всех допустимых обобщений в рамках $\{G(\tau)\}$. *Предельный синдром* S^* является предельным в трех смыслах: его нельзя усилить, т.е. повысить ранг; его нельзя редуцировать и его нельзя обобщить ни по одному входящему тесту. В случае с ВР $v(t)$ имеем шесть однотипных тестов, задаваемых орграфом $G(v)$. В общем случае, многомерный ВР $\{\varphi(t)\}$, где $t = 1, \ldots, M$, задается $|\{\varphi(t)\}| \times M$ тестами с орграфами $\{G(\varphi)\}$.

Совокупность синдромов $\{S\}$ образует *синдромную модель знаний*, если она позволяет определить заключение, как минимум, для любой ситуации действительности из $\Omega(\{\tau/T_0\})$ – априорного описания. Синдромная модель знаний *минимальна*, если из нее нельзя удалить ни один синдром без потери полноты охвата прецедентов из $\Omega(\{\tau/T_0\})$. Для любой синдромной модели знаний $\{S\}$ можно построить *сопряженную предельную модель* $\{S^*\}$. Можно также найти все предельные синдромы на всех уровнях общности для каждой ситуации $\alpha \in \Omega$. Их объединение представляет собой *полную предельную синдромную модель знаний* в рамках контекста $<\Omega, \{G(\tau)\}>$, которую обозначим $\{S^*\}_{Full}$. На основе $\{S^*\}_{Full}$ могут быть построены (*абсолютно*) *минимальные предельные синдромные модели знаний* $\{S^*\}^*_{Min}$. Модель $\{S^*\}_{Full}$ доминирует все другие модели. Можно найти все $\{S^*\}^*_{Min}$, которые эквивалентны $\{S^*\}_{Full}$.

Помимо синдромов речь также может идти о вероятностных закономерностях $\{R\}$ и сопряженных предельных вероятностных закономерностях $\{R^*\}$, которые являются ранними предвестниками событий. *Вероятностной закономерностью* появления заключения $\underline{z}/Z$ назовем правило вида

$$R = (\{\underline{\tau}/T\} \rightarrow J_z\ \underline{z}/Z), \quad p(R) \geq p^*, \ \upsilon_R = \upsilon(R),$$

где $\{\underline{\tau}/T\}$ – неизбыточная совокупность значений тестов; J_z и $p(R)$ – ранг или вероятность получения заключения $\underline{z}/Z$ при условии $\{\underline{\tau}/T\}$; p^* – порог (например, 0.9); υ_R – вес правила, пропорциональный количеству прецедентов с заключением $\underline{z}/Z$, отвечающих правилу R. По аналогии с $\{S^*\}_{Full}$ строится предельная модель $\{R^*\}_{Full}$. Предельные модели знаний являются результатом когнитивной самоорганизующейся критичности.

Если необходимо повысить точность искомой закономерности, то либо расширяется само Z (исходный интервал значений целевого показателя разбивается на большее число подинтервалов), либо указанная процедура (формирование Z и построение предельных моделей знаний) повторяется для найденного подинтервала. Таким образом, получаем иерархическую процедуру определения требуемой закономерности заданного уровня точности.

Для ВР φ могут существовать ряды производных φ' и φ'', и соответственно, орграфы доменов тестов $G(\varphi')$, $G(\varphi'')$. Могут быть определены также орграфы набросков $Gs(\varphi)$, $Gs(\varphi')$, $Gs(\varphi'')$, $Gs_\Sigma(\varphi)$, $Gs_\Sigma(\varphi')$, $Gs_\Sigma(\varphi'')$, $Gs_\Sigma(\{\varphi\}, \{\varphi'\}, \{\varphi''\})$. Всю совокупность вторичных (вычисляемых) данных для ВР $\varphi(t)$ назовем *информационным множеством* $I(\varphi)$. Соответственно для многомерного ВР $\{\varphi(t)\}$ получим $I(\{\varphi\})$. Информационные множества позволяют перейти от $\{\alpha(\{\varphi\}, \underline{z}/Z)\}$ к $\{\alpha(\{\varphi\}, I(\{\varphi\}), \underline{z}/Z)\}$, что значительно обогащает возможности анализа и прогнозирования за счет построения более совершенных моделей знаний.

Определим скорость изменения значения произвольного теста τ. Пусть известны два события $<\underline{\tau}/T, \underline{t}/\Lambda>$ и $<\underline{\tau}'/T, \underline{t}'/\Lambda>$ таких, что $\underline{t}/\Lambda < \underline{t}'/\Lambda$. Если выполняется условие:

$$|\underline{t}' - \underline{t}|/\Lambda' \leq \Delta_\tau/\Lambda', \tag{1}$$

где Δ_τ/Λ' – параметр, зависящий от теста, то определим новое событие

$$e' = <(\underline{\tau} \uparrow \underline{\tau}')/T, (\underline{t}' - \underline{t})/\Lambda', \underline{t}/\Lambda>, \tag{2}$$

которое характеризует скорость изменения значения теста (упрощенное дифференцирование). Оператор изменения «$\uparrow$» зависит от теста. Во многих случаях удобно ввести производный тест «$\Delta\tau$», положив: $\Delta\underline{\tau}/T' = (\underline{\tau}' - \underline{\tau})/T'$. Выражение (2) перепишется следующим образом:

$$e' = <\Delta\underline{\tau}/T', \Delta\underline{t}/\Lambda', \underline{t}/\Lambda>. \tag{3}$$

Для теста $\Delta\tau$ должен быть определен орграф $G(\Delta\tau)$, относящийся к свойствам теста τ. Параметры $\{\Delta_\tau/\Lambda'\}$ также принадлежат к числу свойств теста τ. Приведем примеры [3]:

<Креатинин? 1.3, 12.02.11/Λ3>, <Креатинин? 1.7, 13.02.11/Λ3>;

e' = <Креатинин? 1.3$\uparrow$1.7, сутки, 12.02.11/Λ3>;

e' = <ΔКреатинин? 0.4, сутки, 12.02.11/Λ3>.

<АДс? 130, 12.02.11:12.00/Λ1>, <АДс? 155, 12.02.11:15.00/Λ1>;

e' = <АДс? 130$\uparrow$155, 3 часа, 12.02.11:12.00/Λ1>;

e' = <ΔАДс? 25, 3 часа, 12.02.11:12.00/Λ1>.

e' = <ΔАДс/D2? Большое, $\Delta\underline{t}/\Lambda$2? Малое, 12.02.11:12.00/Λ1>.

Не всегда изменение теста можно выразить в числовом виде. Пример:

<МОКРОТА.Цвет? белая, 17.04.11/Λ3>; <МОКРОТА.Цвет? желтоватая, 18.04.11/Λ3>;

e' = <МОКРОТА.Цвет? белая$\uparrow$желтоватая, сутки, 17.04.11/Λ3>;

e' = <ΔМОКРОТА.Цвет/{Есть; Нет]? Есть, сутки, 17.04.11/Λ3>.

С помощью производного события «Изменение» можно устанавливать качественные прерывности в динамике тестов. Такой «рекурсивный» подход открывает новое пространство признаков (тестов), явно не присутствующих в предметной области, что расширяет границы информационных множеств.

Один из самых ярких результатов теории самоорганизации состоит в понимании существования некой грани, до которой мы можем предвидеть будущее состояние процесса. За этой гранью мы вынуждены оставаться в области вероятных сценариев. Важнейшее положение анализа ВР на основе ППО заключается в том, что для каждого наброска процесса $\{\underline{x}(t)/X\}$ эта грань своя.

АНАЛИЗ ТЕНДЕНЦИЙ

Анализ тенденций может быть частью построения информационного множества $I(\varphi)$. Действительно, анализ тенденций порождает протяженные события вида $<\underline{\tau}/T,\ \underline{\delta}/\Lambda>$, где $\underline{\delta}$ – произвольный временной интервал. Последние играют важную роль в поиске закономерностей и предвестников угрожающих состояний. С методологической точки зрения важно определить легко масштабируемый тест «Тенденция».

Бинарное отношение $<$ на множестве R называют отношением строгого порядка, если оно транзитивно ($x < y \wedge y < z \Rightarrow x < z$) и антисимметрично ($x < y \wedge y < x \Rightarrow x = y$). Обобщим данное определение на произвольный домен T орграфа $G(\tau)$.

Элементы домена T могут содержать любые подмножества базового домена T_0, а именно: множества чисел, интервалов, термов. Будем говорить, что на домене T орграфа $G(\tau)$ задано бинарное отношение (строгого) порядка $\{<; >\}$, если оно транзитивно и антисимметрично. Для любых элементов $a, b \in T$ можно записать: $a < b$ или $a > b$ (элементы домена не могут совпадать). Примеры. Пусть $a = \{1\ 2\ 5\}$, $b = \{8\ 9\ 10\}$, $c = \{4\ 7\}$. Ясно как трактовать: $a < b$ (любой элемент из a меньше любого элемента из b) и $c < b$, но между a и c нет отношения порядка. Пусть $a = \{[1, 3]\ [5, 8]\}$, $b = \{[9, 11]\ (11, 15]\}$. Ясно как трактовать: $a < b$. Пусть a = «слабый рост (показателя)», b = «сильный рост». Ясно как трактовать: $a < b$ (сравниваются интенсивности качества).

***Предложение* 1**. Все домены произвольного орграфа $G(\tau)$ делятся на два каласса: домены, на которых можно задать отношение порядка, и домены, на которых нельзя задать отношение порядка. Один из классов может быть пустым.

Рассмотрим два примера орграфов доменов.

Пример 1. Зададим орграф $G(\tau) = \{D0 \rightarrow D1 \rightarrow D2\}$ следующим образом:

$D0\ \{1; 2; 3; 4; 5\} \rightarrow D1\ \{2; 1\ 3; 4; 5\} \rightarrow D2\ \{1\ 2\ 3; 4; 5\}$

Легко видеть, что на доменах D0 и D2 можно задать порядок, а на промежуточном домене D1 нет. Данный пример показывает, что если на домене D определено отношение порядка, то это не означает, что на любом более точном домене D′: D′ $\rightarrow$ D также определено отношение порядка.

Вместе с тем, даже если на всем домене нельзя задать отношение порядка, то вполне возможно, что такое отношение можно задать на его подмножестве, например, на подмножестве {1 3; 4; 5} домена D1 можно задать отношение порядка. Зададим временной ряд из трех значений: $\{\varphi(t)/D1\} = \{\varphi(1)/D1?\ 1\ 3;\ \varphi(2)/D1?\ 4;\ \varphi(3)/D1?\ 5\}$. ВР принимает значения не из всего домена, а только той его части, на которой можно определить отношение порядка. В результате устанавливается тенденция «рост» на D1. Данный пример иллюстрирует общее правило.

Пример 2. Зададим орграф $G(\varphi) = \{T0\}$ следующим образом: $T0 = \{1; 2; 3\}$. Построим структурно-завершенный орграф $G^{+}(\varphi) = \{T0 \rightarrow T1; T0 \rightarrow T2; T0 \rightarrow T3\}$:

$G^{+}(\varphi) = \{T3\ \#\ T0\ \{3; 1\ 2\}\ T2\ \#\ T0\ \{2; 1\ 3\}\ \ T1\ \{1; 2\ 3\}\ T0\ \{1; 2; 3\}\}$

Домены-вершины T1, T2, T3 являются листьями. Легко убедиться, что на доменах T0, T1, T3 можно задать отношение порядка, а на домене T2 нельзя.

Зададим временной ряд из трех значений: $\{\varphi(t)/T0\} = \{\varphi(1)/T0?\ 1;\ \varphi(2)/T0?\ 2;\ \varphi(3)/T0?\ 3\}$. Ряд $\{\varphi(t)/T0\}$ показывает четкую тенденцию «роста». Ту же монотонную тенденцию показывают наброски {$\varphi(t)$/T1} и {$\varphi(t)$/T3}, в частности, {$\varphi(t)$/T3} = {$\varphi(1)$/T3? 1 2; $\varphi(2)$/T3? 1 2; $\varphi(3)$/T3? 3}. Однако про тенденцию ряда {$\varphi(t)$/T2} = {$\varphi(1)$/T2? 1 3; $\varphi(2)$/T2? 2; $\varphi(3)$/T2? 1 3} практически ничего определенного сказать нельзя: может быть «рост», «падение» и «колебания». На основе ряда {$\varphi(t)$/T2} можно сделать вывод только о том, что «нет стабильности».

Зададим несколько иной временной ряд: {$\varphi(t)$/T0} = {$\varphi(1)$/T0? 2; $\varphi(2)$/T0? 1; $\varphi(3)$/T0? 3}. На базовом уровне имеет место «колебание» в то время как обобщенный набросок {$\varphi(t)$/T3} = {$\varphi(1)$/T3? 1 2; $\varphi(2)$/T3? 1 2; $\varphi(3)$/T3? 3} демонстрирует «рост». Набросок {$\varphi(t)$/T1} = {$\varphi(1)$/T1? 2 3; $\varphi(2)$/T1? 1; $\varphi(3)$/T1? 2 3} также демонстрирует «колебание», а на основе наброска {$\varphi(t)$/T2} = {$\varphi(1)$/T2? 2; $\varphi(2)$/T2? 1 3; $\varphi(3)$/T2? 1 3} можно сделать вывод только о наличии качественной прерывности. В итоге получаем «спектр заключений» на всех уровнях общности описания ВР, который включает «рост», «колебание», «нестабильно» и т.д. Важно отметить, что все заключения справедливы *одновременно*. Другими словами, имеет место суперпозиция смыслов (значений).

Разные уровни обобщения ВР позволяют более точно определить понятия «сходства» и корреляции между рядами, включая сходство тенденций. Данные понятия имеют смысл только в рамках фиксированного уровня общности описания ВР. Сколько уровней общности описания ВР столько и мер сходства. Так ряды {1, 2, 3} и {2, 1, 3} полностью совпадают в рамках домена T3.

Пусть тест x задан орграфом доменов $G(x)$. Для любого домена X можно рассмотреть временной ряд $\{\underline{x}/X\ (t)\} \equiv \{\underline{x}(t)/X\}$. Пусть на X определено отношение порядка. *Элементарная тенденция* (ЭТ) между двумя соседними значениями ВР $\{\underline{x}(t)/X\}$ на базовом уровне общности орграфа G(ЭТ) принимает только три значения: «рост» - ЭТ↑ ($\underline{x}'/X > \underline{x}/X$), «падение» - ЭТ↓ ($\underline{x}'/X < \underline{x}/X$), «стабильно» ($\underline{x}'/X = \underline{x}/X$). Может быть построен структурно-завершенный орграф G^+(ЭТ), который по виду совпадает с орграфом $G^+(\varphi)$.

Любая *локальная тенденция* (ЛТ) определяется на основе известной последовательности более мелких локальных тенденций, в частности элементарных тенденций. Если локальная тенденция охватывает весь временной интервал, то она становится *общей тенденцией*. Определим шесть базовых значений ЛТ ВР $\{\underline{x}(t)/X\}$ на отрезке $[t, t']$ следующим образом ($\underline{x} = \underline{x}(t)$, $\underline{x}' = \underline{x}(t')$):

«стабильно» $\Leftrightarrow$ ($\underline{x}'/X = \underline{x}/X$) & (все ЛТ «стабильно»);

«рост» $\Leftrightarrow$ ($\underline{x}'/X > \underline{x}/X$) & (все ЛТ «стабильно» или «рост»);

«падение» $\Leftrightarrow$ ($\underline{x}'/X < \underline{x}/X$) & (все ЛТ «стабильно» или «падение»);

«колебания - стабильно» $\Leftrightarrow$ ($\underline{x}'/X = \underline{x}/X$) & $\neg$ (все ЛТ «стабильно»);

«колебания - рост» $\Leftrightarrow$ ($\underline{x}'/X > \underline{x}/X$) & $\neg$ (все ЛТ «стабильно» или «рост»);

«колебания - падение» $\Leftrightarrow$ ($\underline{x}'/X < \underline{x}/X$) & $\neg$ (все ЛТ «стабильно» или «падение»).

Определение ЛТ является рекурентным. Справедливо предложение.

***Предложение* 2**. Характер локальной тенденции инвариантен относительно разбиения ВР $\{\underline{x}(t)/X\}$ на более мелкие локальные тенденции.

Введем орграф доменов теста «Тенденция»:

Тенденция ^*tr* {

3 #1 {Итог-стабильно ^1 4; Итог-рост ^2 5; Итог-падение ^3 6}

2 {Стабильно ^1; Монотонное изменение ^2 3; Колебания ^4 5 6}

1 {Стабильно ^1; Рост ^2; Падение ^3; Колебания - стабильно ^4; Колебания - рост ^5; Колебания - падение ^6}}

G(tr) = {1 $\rightarrow$ 2; 1 $\rightarrow$ 3}.

На рис. 4 показан предельный структурно-завершенный орграф G^{++}(tr). Общее число вершин – 14 (зеленым или светлым цветом обозначены листья, одна вершина блуждающая). Все они участвуют в анализе тенденций, обеспечивая суперпозицию смыслов (даже те домены, на которых не определено отношение порядка).

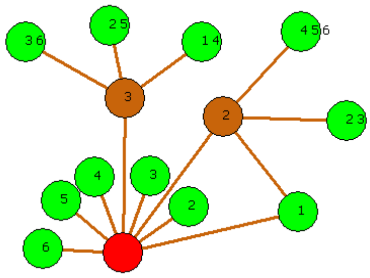


Рисунок 4 – Предельный структурно-завершенный орграф $G^{++}(tr)$

Предложение 3. Если на интервале ВР [*t*, *t'*] существуют ЭТ↑ и ЭТ↓, то tr/2 = «Колебания».

Предложение 4. Пусть тест *x* задан орграфом доменов *G*(*x*) и на домене *X* определено отношение порядка. Если на интервале [*t*, *t'*] для ВР {$\underline{x}(t)/X$} имеет место tr([*t*, *t'*])/2 = «Колебания», то для любого домена *X'* такого, что $X' \rightarrow X$ и на *X'* определено отношение порядка, верно: для ВР {$\underline{x}(t)/ X'$} имеет место tr([*t*, *t'*])/2 = «Колебания».

Другими словами, повышение уровня точности описания ВР сохраняет «колебания». Выше было показано, что не на каждом домене *X'* может быть определено отношение порядка.

Предложение 5. Общее количество масштабных уровней в смысловой суперпозиции тенденций при анализе ВР {$\underline{x}(t)$} оценивается величиной $|G(x)|\cdot|G^{+}(\mathrm{tr})|$ при условии, что все элементы ряда преобразовываются одновременно к одному уровню общности (домену).

Таким образом, при минимальных требованиях к домену (наличие отношения порядка хотя бы на части домена) удалось ввести масштабируемый тест «Тенденция». Ясно, что орграф *G*(tr) может иметь другой вид, нежели тот, что определен выше.

ЗАДАЧИ ПОЗНАНИЯ, ПОИСК ЗАКОНОМЕРНОСТЕЙ

Задачи познания предметной области сводятся к задачам усиления (в логическом смысле) логической эмпирической теории за счет обнаружения аксиом на множестве эмпирических данных Ω [3, 6]. В практическом плане задача состоит в автоматическом выделении семантических образов произвольных причинно-связанных процессов из слабоструктурированных информационных потоков, отображающих реальные физические и когнитивные процессы.

Причинно-следственные связи, в частности, позволяют: определять условия и время попадания траекторий процессов в поглощающее или инвариантное множество; строить информационные множества; предсказывать катастрофы и бифуркации процессов (например, инфаркты, инсульты и т.д.); выявлять параметры порядка при анализе путей наиболее безопасного выхода из критического состояния после катастрофы; проводить анализ асимптотического поведения системы; выполнять оценку устойчивости процесса; определять дрейф структурных параметров системы.

Определим базовые структуры закономерностей (по аналогии с [6]). Под ситуацией действительности в контексте данной работы понимается информационное множество многомерного ВР – $I(\{\varphi\})$.

Утверждение «в ситуации действительности α вслед за множеством событий {*c*} на временном интервале δ/Λ наблюдается событие *e*» запишем следующим образом: J_{+1} Exp (α, {*c*}, δ/Λ, *e*).

Утверждение «в ситуации действительности α вслед за множеством событий {*c*} на временном интервале δ/Λ не наблюдается событие *e*» запишем следующим образом: J_{-1} Exp (α, {*c*}, δ/Λ, *e*).

Введем *J*-атомарные формулы с предикатным символом PC (возможная причина):

J_{+1} PC ({*c*}, δ/Λ, {$\underline{\tau}/T$}, *e*) будем интерпретировать как «{*c*} является возможной причиной наступления *e* на интервале δ/Λ при условии выполнения {$\underline{\tau}/T$}» (+Правило для PC).

$J_{\vee}$ PC ({*c*}, δ/Λ, {$\underline{\tau}/T$}, *e*) будем интерпретировать как «Хотя бы одно из событий {*c*} является возможной причиной наступления *e* на интервале δ/Λ при условии выполнения {$\underline{\tau}/T$}» (∨Правило для PC).

J_L PC (L({*c*}), *e*) будем интерпретировать как «Факт L({*c*}) является возможной причиной наступления *e*» (*L*Правило для PC).

J_{-1} PC ({*c*}, δ/Λ, {$\underline{\tau}/T$}, *e*) будем интерпретировать как «{*c*} является возможной причиной не наступления *e* на интервале δ/Λ при условии выполнения {$\underline{\tau}/T$}» (-Правило для PC).

J_0 PC ({*c*}, δ/Λ, {$\underline{\tau}/T$}, *e*) будем интерпретировать как «сведения о причинно-следственных отношениях между $\{c\}$ и e на интервале δ/Λ при условии выполнения $\{\underline{\tau}/T\}$ противоречивы» (0Правило для РС).

$J_?$ РС ($\{c\}$, δ/Λ, $\{\underline{\tau}/T\}$, e) будем интерпретировать как «сведений о причинно-следственных отношениях между $\{c\}$ и e на интервале δ/Λ при условии выполнения $\{\underline{\tau}/T\}$ недостаточно» (?Правило для РС).

Условия $\{\underline{\tau}/T\}$ действуют на множестве ситуаций действительности Ω. Условия $\{\underline{\tau}/T\}$ могут быть фильтром, как ситуаций α, так и событий $\{c\}$. Условия $\{\underline{\tau}/T\}$ могут отсутствовать (в таком случае получим безусловные правила).

Пример фильтра только для ситуаций действительности (из медицинской практики): $\{\underline{\tau}/T\}$ = {Пол?Ж; Ds? Бронхит; Возраст/В2? 40-55}. Пример фильтра только для событий: $\{\underline{\tau}/T\}$ = {Тип тестов в событиях $\{c\}$? Биохимия крови}.

Тот факт, что ситуация действительности α удовлетворяет условиям $\{\underline{\tau}/T\}$ будем записывать с помощью нотации: $\alpha\nabla\{\underline{\tau}/T\}$. Аналогично будем записывать сведения о том, что события $\{c\}$ удовлетворяют условиям $\{\underline{\tau}/T\}$: $\{c\}\nabla\{\underline{\tau}/T\}$

Дадим неформальное описание «+Правила» для РС.

Пусть

- существует хотя бы одна ситуация действительности, удовлетворяющая $\{\underline{\tau}/T\}$, где вслед за событиями $\{c\}$, удовлетворяющими $\{\underline{\tau}/T\}$, на интервале δ/Λ наблюдается событие e;

- не существует ни одной ситуации действительности, удовлетворяющей $\{\underline{\tau}/T\}$, где вслед за событиями $\{c\}$, удовлетворяющими $\{\underline{\tau}/T\}$, на интервале δ/Λ не наблюдается событие e.

Тогда события $\{c\}$ являются возможной причиной наступления события e на интервале δ/Λ при условии выполнения $\{\underline{\tau}/T\}$.

Формальная запись «+Правила» будет иметь следующий вид:

$\exists\alpha\nabla\{\underline{\tau}/T\}$ $(J_{+1}\mathrm{Exp}$ $(\alpha, \{c\}\nabla\{\underline{\tau}/T\}, \delta/\Lambda, e))$ $\&\neg\exists\beta\nabla\{\underline{\tau}/T\}$ $(J_{-1}\mathrm{Exp}$ $(\beta, \{c\}\nabla\{\underline{\tau}/T\}, \delta/\Lambda, e))$ $\rightarrow$ J_{+1}РС ($\{c\}$, δ/Λ, $\{\underline{\tau}/T\}$, e).

Дадим неформальное описание «$\vee$Правила» для РС.

Пусть

– существует хотя бы одна ситуация действительности, удовлетворяющая $\{\underline{\tau}/T\}$, где вслед за каким либо событием из $\{c\}$, удовлетворяющим $\{\underline{\tau}/T\}$, на интервале δ/Λ наблюдается событие e;

– не существует ни одной ситуации действительности, удовлетворяющей $\{\underline{\tau}/T\}$, где вслед за каким либо событием из $\{c\}$, удовлетворяющим $\{\underline{\tau}/T\}$, на интервале δ/Λ не наблюдается событие e.

Тогда множество событий $\{c\}$ содержит возможные причины наступления события e на интервале δ/Λ при условии выполнения $\{\underline{\tau}/T\}$.

Формальная запись «$\vee$Правила» будет иметь следующий вид:

$\exists\alpha\nabla\{\underline{\tau}/T\}$ $\exists c'\in\{c\}\nabla\{\underline{\tau}/T\}$ $(J_{+1}\mathrm{Exp}$ $(\alpha, c', \delta/\Lambda, e))$ $\&\neg\exists\beta\nabla\{\underline{\tau}/T\}$ $\forall c'\in\{c\}\nabla\{\underline{\tau}/T\}$ $(J_{-1}\mathrm{Exp}$ $(\beta, c', \delta/\Lambda, e))$ $\rightarrow$ $J_{\vee}$РС ($\{c\}$, δ/Λ, $\{\underline{\tau}/T\}$, e).

Неформальное описание «-Правила» для РС.

Пусть

– существует хотя бы одна ситуация действительности, удовлетворяющая $\{\tau/T\}$, где вслед за событиями $\{c\}$, удовлетворяющими $\{\tau/T\}$, на интервале δ/Λ не наблюдается событие e;

– не существует ни одной ситуации действительности, удовлетворяющей $\{\tau/T\}$, где вслед за событиями $\{c\}$, удовлетворяющими $\{\tau/T\}$, на интервале δ/Λ наблюдается событие e.

Тогда события $\{c\}$ являются возможной причиной *не* наступления события e на интервале δ/Λ при условии выполнения $\{\underline{\tau}/T\}$.

Формальная запись «-Правила» будет иметь следующий вид:

$\exists\alpha\nabla\{\underline{\tau}/T\}$ $(J_{-1}\mathrm{Exp}$ $(\alpha, \{c\}\nabla\{\underline{\tau}/T\}, \delta/\Lambda, e))$ $\&\neg\exists\beta\nabla\{\underline{\tau}/T\}$ $(J_{+1}\mathrm{Exp}$ $(\beta, \{c\}\nabla\{\underline{\tau}/T\}, \delta/\Lambda, e))$ $\rightarrow$

J_{-1}РС ($\{c\}$, δ/Λ, $\{\underline{\tau}/T\}$, e).

Неформальное описание «0Правила» для РС.

Пусть

– существует хотя бы одна ситуация действительности, удовлетворяющая $\{\tau/T\}$, где вслед за событиями $\{c\}$, удовлетворяющими $\{\tau/T\}$, на интервале δ/Λ не наблюдается событие e;

–существует хотя бы одна ситуация действительности, удовлетворяющая $\{\tau/T\}$, где вслед за событиями {c}, удовлетворяющими {τ/T}, на интервале δ/Λ наблюдается событие e.

Тогда сведения о причинно-следственных отношениях между {*c*} и *e* на интервале δ/Λ при условии выполнения $\{\underline{\tau}/T\}$ противоречивы.

Формальная запись «0Правила» будет иметь следующий вид:

$\exists\alpha\nabla\{\underline{\tau}/T\}$ (J_{-1}Exp (α, $\{c\}\nabla\{\underline{\tau}/T\}$, δ/Λ, e)) & $\exists\beta\nabla\{\underline{\tau}/T\}$ (J_{+1}Exp (β, $\{c\}\nabla\{\underline{\tau}/T\}$, δ/Λ, e)) $\rightarrow$

J_0 PC ({*c*}, δ/Λ, $\{\underline{\tau}/T\}$, *e*).

Рассмотрим теперь неформальное описание «?Правила» для PC.

Пусть

– не существует ни одной ситуации действительности, удовлетворяющей {τ/T}, где вслед за событиями {c}, удовлетворяющими {τ/T}, на интервале δ/Λ наблюдается событие e;

– не существует ни одной ситуации действительности, удовлетворяющей {τ/T}, где вслед за событиями {c}, удовлетворяющими {τ/T}, на интервале δ/Λ не наблюдается событие e.

Тогда сведений о причинно-следственных отношениях между {*c*} и *e* на интервале δ/Λ при условии выполнения $\{\underline{\tau}/T\}$ недостаточно.

Формальная запись «?Правила» будет иметь следующий вид:

$\neg\exists\alpha\nabla\{\underline{\tau}/T\}$ (J_{-1}Exp (α, $\{c\}\nabla\{\underline{\tau}/T\}$, δ/Λ, e)) &$\neg\exists\beta\nabla\{\underline{\tau}/T\}$ (J_{+1}Exp (β, $\{c\}\nabla\{\underline{\tau}/T\}$, δ/Λ, e)) $\rightarrow$

$J_?$ PC ({*c*}, δ/Λ, $\{\underline{\tau}/T\}$, *e*).

Все операции необходимо рассматривать с учетом банка тестов $\{G(\tau)\}$.

ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ МЕТОДИКИ

Пусть изучаются факторы, влияющие на наступление некоторого события, например, банкротства предприятия [3]. Положим *Z* = {событие+; событие-} = {+; -} = {1, 2}. Сформируем первичную базу прецедентов следующего вида: $\Omega = \{\alpha(\{\underline{\tau}/T\}, \underline{z}/Z)\}$, где *Z* = {1, 2}. Если речь идет о банкротстве предприятий, то $\Omega^+ = \{\alpha(\{\underline{\tau}/T\}, \underline{z}=1)\}$ – множество обанкротившихся предприятий, соответственно – $\Omega^- = \{\alpha(\{\underline{\tau}/T\}, \underline{z}=2)\}$ – множество стабильных (успешных) предприятий.

Эффективным методом построения орграфов набросков является расчет всевозможных индикаторов (показателей, индексов), характеризующих течение сложного процесса. Индикаторы применяются практически во всех областях знаний: экономике, медицине, экологии, политике, социальных исследованиях. Каждый индикатор *p* является тестом и, следовательно, задается орграфом доменов G(*p*). Среди экономических индикаторов первостепенное значение имеют показатели состояния и результатов функционирования экономики в целом, которые часто называют *агрегированными показателями* $\{p/P\}^*$. В контексте построения орграфа набросков первичным этапом является определение *компонентов* (агрегированных) экономических показателей {*c/C*}. Орграф набросков позволяет перейти от первичного описания базы прецедентов к расширенному описанию следующего вида: $\Omega' = \{\alpha(\{\underline{\tau}/T\}, \{\underline{c}/C\}', \{\underline{c}/C\}, \{\underline{p}/P\}, \{\underline{p}/P\}^*, \underline{z}/Z)\}$, где *Z* = {1, 2}. Фактически речь идет о построении информационных множеств $\Omega' = \{\alpha(\{\varphi\}, I(\{\varphi\}), \underline{z}/Z)\}$. Мониторинг показателей за определенный период позволяет сформировать многомерные ВР.

Переход от описания Ω к описанию Ω' позволяет задействовать все виды обобщений, как по горизонтали (в рамках одного прецедента), так и по вертикали (пример комбинации нескольких видов орграфов набросков). Далее, в рамках контекста $K = \langle\Omega', \{G(\tau)\}\rangle \equiv \langle\Omega', \{G(\tau)\},\{G(c)\},\{G(p)\}\rangle$ строятся синдромные модели знаний {*S*}, в частности, предельные синдромные модели знаний $\{S^*\}=\{S^*\}^+\cup\{S^*\}^-$, а также модели (предельных) вероятностных закономерностей $\{R^*\}=\{R^*\}^+\cup\{R^*\}^-$.

Множество синдромов $\{S^*\}^+$ определяет все возможные наборы параметров порядка (в рамках контекста K), которые приводят к наступлению «события+» (например, банкротству), а закономерности из $\{R^*\}^+$ являются предвестниками события. Напротив, множество синдромов $\{S^*\}^-$ определяет все возможные наборы параметров порядка, которые обеспечивают благоприятный режим функционирования. На основе знания указанных множеств синдромов может быть реализован *синдромный принцип управления*, целью которого является выход на благоприятный режим функционирования и его стабилизация [3].

Конкретизируем приведенную выше схему применения орграфа набросков и многомерных ВР для изучения факторов банкротства предприятия.

Текущий риск банкротства предприятия будем оценивать с помощью пяти тестов: X1 – общая оборачиваемость активов предприятия; X2 – отношение чистой прибыли к активам предприятия; X3 – отношение краткосрочных и долгосрочных заемных средств к активам предприятия; X4 – отношение долгосрочных обязательств к активам предприятия; X5 – натуральный логарифм выручки предприятия. Собственно риск P задается некоторой функцией от параметров X1, …, X5, т.е. P = P(X1, …, X5). Банк тестов $\{G(\tau)\} = \{G(X1), \ldots, G(X5), G(P)\}$. Без потери общности базовый домен во всех орграфах обозначим «1». Текущий риск банкротства является примером интегрального показателя (индикатора).

В базе прецедентов по каждому предприятию будем хранить данные за несколько лет предшествовавших банкротству, т.е. множество кортежей вида $\{<X1, \ldots, X5>_1, \ldots, <X1, \ldots, X5>_N\}$, где N – количество лет, включая год банкротства. Для успешных предприятий также берутся данные за N лет. Следовательно, первичная база прецедентов имеет вид:

$\Omega(\{\tau/T_0\}) = \{\alpha(\{<X1/1, \ldots, X5/1>_1, \ldots, <X1/1, \ldots, X5/1>_N\}, \underline{z}/Z)\}$,

где $Z=\{1, 2\}$. Расширенная база прецедентов, включающая индикаторы (риски), имеет вид:

$\Omega'(\{\tau/T_0\}) = \{\alpha(\{<X1/1, \ldots, X5/1, P/1>_1, \ldots, <X1/1, \ldots, X5/1, P/1>_N\}, \underline{z}/Z)\}$.

На основе контекста $K = <\Omega', \{G(\tau)\}>$ строятся предельные синдромные модели знаний $\{S^*\}$ и множества предельных вероятностных закономерностей $\{R^*\}$. Каждый предельный синдром из $\{S^*\}^+$ описывает, по сути, набор параметров порядка (сценарий) банкротства предприятия и, соответственно, характер изменения финансовых показателей в ретроспективной динамике. Каждая предельная вероятностная закономерность из $\{R^*\}^+$ является предвестником события. На основе синдромных моделей $\{S^*\}$ и моделей $\{R^*\}$ может быть выполнено построение комплекса динамических моделей прогнозирования банкротства путем выделения синдромов и вероятностных закономерностей разной степени удаленности от интересуемого события, в частности, ранних предвестников события.

Пример из медицины. Пусть изучаются факторы, влияющие на наступление некоторого события, например, летального исхода, инвалидизации, неблагоприятного течения заболевания (после инсульта, инфаркта и т.д.) или неэффективности лечения. Положим Z = {событие+; событие-} = {+; -} = {1, 2}. Сформируем первичную базу прецедентов следующего вида: $\Omega = \{\alpha(\{\underline{\tau}/T\}, \underline{z}/Z)\}$, где $Z = \{1, 2\}$. Если речь идет о летальном исходе, то $\Omega^+ = \{\alpha(\{\underline{\tau}/T\}, \underline{z}=1)\}$ – множество умерших пациентов, соответственно – $\Omega^- = \{\alpha(\{\underline{\tau}/T\}, \underline{z}=2)\}$ – множество выживших пациентов. Далее, в рамках контекста $K = <\Omega', \{G(\tau)\}>$ строятся предельные синдромные модели знаний $\{S^*\}=\{S^*\}^+ \cup \{S^*\}^-$ и предельные вероятностные модели $\{R^*\}=\{R^*\}^+ \cup \{R^*\}^-$. Множество синдромов $\{S^*\}^+$ определяет все возможные наборы параметров порядка (в рамках контекста K) приводящие к наступлению «события+», а закономерности из $\{R^*\}^+$ являются предвестниками события. Напротив, множество синдромов $\{S^*\}^-$ определяет все возможные наборы параметров порядка, которые обеспечивают благоприятное течение процесса.

По аналогии строятся синдромные и вероятностные модели знаний для произвольных экономических, социальных, политических и иных событий [3, 7]. Модели включают многомерные ВР.

На рис. 5 приведена диалогая форма партнерской информационной системы, предназначенной для автоматизированной оценки качества медицинской помощи [8].

Графическое представление динамики основных симптомов (тестов) позволяет эксперту и ИС все анализируемые симптомы разделить на появившиеся сразу (лихорадка, ознобы) и появившиеся позже (желтуха, одышка, гематурия, спутанность сознания). Иными словами, эти группы симптомов находятся во временной зависимости.

Заключение. Когнитивный подход на основе ППО позволяет обобщить постановку и решение традиционных задач анализа временных рядов, например, таких как: сегментация – разбиение ВР на значимые сегменты; кластеризация – поиск группировок ВР или их паттернов; классификация – назначение ВР или их паттернам одного их заранее определенных классов; индексирование – построение индексов для эффективного выполнения запросов к базам данных ВР; резюмирование (summarization) – формирование краткого описания ВР, содержащего существенные черты с точки зрения решаемой задачи; обнаружение аномалий – поиск новых, не типичных паттернов; частотный анализ – поиск часто проявляющихся паттернов; прогнозирование – прогноз очередного значения на базе истории ВР; извлечение ассоциативных правил – поиск правил, относящихся к паттернам ВР.

Рисунок 5 – Диалоговая форма –Динамика симптомов {$\underline{\tau}$(t)}

Частью общей методологии анализа ВР на основе ППО являются: алгоритмы построения предельных моделей знаний, алгоритмы построения информационных множеств, алгоритмы познания предметной области (извлечения обобщенных закономерностей), алгоритмы построения орграфов набросков и экстремальнеых пограничных слоев набросков (для фиксированной целевой задачи), а также методы когнитивной бикластеризации [3].

При создании когнитивных агентов способных анализировать ВР предельные модели знаний преобразуются в *среды радикалов*, где каждый радикал отвечает за установление отдельного синдрома или вероятностной закономерности. Связка «Модель знаний – среда радикалов» образует функциональную систему когнитивно-поведенческого уровня (в общем случае следует рассматривать иерархию функциональных систем). Функциональные системы обеспечивают высокую *пластичность* решения целевых задач [3, 7].

ЛИТЕРАТУРА:

1. Ярушкина, Н.Г. Интеллектуальный анализ временных рядов: учебное пособие /Н.Г. Ярушкина, Т.В. Афанасьева, И.Г. Перфильева. – Ульяновск: УлГТУ, 2010. – 320 с.
2. Song Q., Chisson B. Fuzzy Time Series and its Models //Fuzzy Sets and Systems. – 2004. – №54. – P.269-277.
3. Прокопчук Ю.А. Принцип предельных обобщений: методология, задачи, приложения. Монография. – Дн-вск: ИТМ НАНУ и НКАУ, 2012. – 384 с.
4. Прокопчук Ю.А. Предельные синдромные и вероятностные модели знаний //Научный вестник Херсонской государственной морской академии. – 2011. – №2 (5). – С.322-333.
5. Илюхина В.А. Иерархический принцип организации и феномен интеграции разных по скоростям мозговых процессов в механизмах познавательной деятельности человека /В.А. Илюхина //IV-й конференция по когнитивным наукам, июнь 2010 г., г. Томск: тезисы докладов. – Томск: Томский государственный университет, 2010. – Т.2. – С.300.
6. Аншаков О.М. ДСМ-метод и модификационные вычисления //Искусственный интеллект и принятие решений. – 2008. – №1. – С.55-79.
7. Прокопчук Ю.А. Реализация когнитивных и метакогнитивных технологий в информационных системах, системах управления и образовании //Вестник Херсонского национального технического университета. – 2012. – №1 (44). – С.27-39.
8. Прокопчук Ю.А. Интеллектуальные медицинские системы: формально-логический уровень. – Дн-ск: ИТМ НАНУ и НКАУ, 2007. – 259 с.