УДК 004.8:004.89

ОЦЕНИВАНИЕ КАЧЕСТВА ИДЕНТИФИКАЦИОННЫХ И ПРОГНОЗНЫХ РЕШЕНИЙ В ИНЖЕНЕРИИ КВАНТОВ ЗНАНИЙ

И.Б. Сироджа

Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», г. Харьков, Украина

Рассматриваются плохо формализованные задачи знаниеориентированного принятия идентификационных и прогнозных решений средствами инженерии квантов знаний в условиях неопределенности. Благодаря введенной логической модели экстраполяции наблюдений в виде системы импликативных закономерностей предметной области как базы знаний, задача прогнозирования в частном случае сводится к задаче идентификации ситуаций. На этой основе предложена единая количественная мера качества идентификационных и прогнозных решений, дедуктивно выводимых из предварительно построенной импликативной базы квантов знаний (БkЗ).

ИНЖЕНЕРИЯ КВАНТОВ ЗНАНИЙ, ИМПЛИКАТИВНЫЕ ЗАКОНОМЕРНОСТИ, БАЗА ЗНАНИЙ, ИДЕНТИФИКАЦИОННЫЕ И ПРОГНОЗНЫЕ РЕШЕНИЯ, ДЕДУКТИВНЫЙ ВЫВОД

Введение

Квантовый подход к инженерии знаний в искусственном интеллекте привел к образованию нового научного направления – *инженерии квантов знаний* (ИКЗ) для решения плохо формализованных задач принятия *идентификационных* и *прогнозных* решений в условиях неопределенности [1, 2]. В отличие от существующих подходов, основанных на фреймах, продукциях, семантических сетях и предикатных исчислениях [4, 5], ИКЗ базируется на представлении знаний посредством алгоритмических структур 0-*го*, 1-*го* и 2-*го* уровней сложности, имеющих три составляющих – *семантическую*, *информационную* и *операторную*. Эти структуры названы *разноуровневыми алгоритмическими квантами знаний* (РАКЗ). Условно *кванты* знаний (k-*знания*) относятся к 0-*му уровню*, если они представимы *числом* или символом; к 1-*му уровню*, если – числовым или символьным *вектором* и ко 2-*му уровню*, если – числовой или символьной *матрицей*. Важнейшим преимуществом РАКЗ-моделей представления и преобразования *k-знаний* является возможность их описания в множественной, векторно-матричной и аналитической (конечно-предикатной) формах, что позволяет манипулировать знаниями посредством машинных алгебр и процедур логического вывода. Основы теории, практические приложения ИКЗ, а также оценивание качества знаниеориентированных решений посредством вычисления риска ошибочного решения по контрольной выборке подробно изложены в монографии [2].

В данной работе предложена единая количественная *мера качества идентификационных* и *прогнозных* решений, определяемая эмпирической степенью корреляции между наблюдаемыми и модельными прогнозными процессами. Такая мера позволяет обоснованно судить о потенциальной *предсказуемости* динамических процессов и об основных причинах неудачных прогнозов.

1. Особенности средств инженерии квантов знаний

Главная особенность математических РАКЗ-моделей и методов ИКЗ состоит в представлении базы *k*-знаний (БkЗ) системой *импликативных* и/или *функциональных* закономерностей, которые находятся индуктивно по выборочным обучающим k-знаниям из предметной области. *Импликативной* закономерностью или *запретом* r-го ранга согласно [2, 5], называется такая устойчивая логическая связь между r двоичными признаками из общего их числа n $(2 \le r \le n)$ $x_1, x_2, ..., x_n$, характеризующих объект принятия решений (ОПР), когда для реальных объектов недопустима (запрещена) хотя бы одна комбинация $\binom{n}{r}$ значений признаков. *Функциональная* связь как частный случай вытекает из *импликативной*. Их отличие состоит в том, что при *функциональной* связи значения признаков-аргументов всегда однозначно определяют значение признака-функции, а при *запретной* связи – не всегда, только при определенных комбинациях значений аргументов. Все эти связи и характеристики ОПР представляются квантами знаний соответствующих уровней. Если указанные признаки $x_1, x_2,..., x_n$ принимают значения из конечных двоичных множеств $X_1, X_2, \cdots, X_n$, то декартово произведение $X_1 \times X_2 \times \cdots \times X_n$ образует бинарное пространство B, элементами которого служат всевозможные наборы элементов, взятых по одному из множеств X_i, $i = \overline{1,n}$. Наборы значений признаков, отвечающие *реальным* объектам, образуют подмножество $B_r \subset B$, а отвечающие выборочным *обучающим* объектам – подмножество $B_0 \subset B_r$. Очевидно, мощности множеств B_0, B_r и B находятся в отношении «<<» – «намного меньше»:

$$|B_0| << |B_r| << |B|. \qquad (1)$$

Кванты знаний об ОПР и связях между их признаками удобно интерпретировать как *интервалы* пространства B или соответствующие элементарные конъюнкции. Из соотношения (1) следует, что *импликативные* закономерности для искомой БкЗ можно обнаруживать путем нахождения *запретных интервалов* во множестве B_0 как *обучающей* выборке. Действительно, если существует связь между признаками, которой отвечает *запретный* интервал J в пространстве B, то должно выполняться отношение $J \cap B_r = \varnothing$, а значит, и отношение $J \cap B_0 = \varnothing$, то есть интервал J – пустой для B_0, так как в обучающей выборке отсутствует соответствующая запретная комбинация значений признаков. Таким образом, если $J \cap B_0 = \varnothing$, то можно выдвинуть гипотезу о существовании запретной связи определенного ранга r, то есть числа связанных переменных в соответствующей элементарной конъюнкции. Достоверность гипотезы оценивается величиной математического ожидания $M_z\{m,n,r\}$ [2] как средним числом пустых интервалов ранга r, возникающих в пространстве B при отсутствии каких-либо закономерностей, когда B_0 объемом $m \times n$ (m – количество квантов-наблюдений, n – число признаков ОПР) выбирается из B случайно с равновероятным выбором элементов. Очевидно, что гипотеза принимается лишь при достаточно малом значении величины $M_z\{m,n,r\}$ из интервала [0, 1], не большего от заданного допустимого порога M_z^* при ранге r запрета, не превышающего $r_{\max}$. Минимизированную и безизбыточную БкЗ представляют в виде матричного запретного кванта знаний 2-го уровня $k_2\overline{\Sigma}_{BM}$, строки которого описывают определенные запреты и могут интерпретироваться как интервалы пространства B либо как отвечающие им элементарные конъюнкции [1, 2]. Следовательно, *импликативные* k-знания $k_2\overline{\Sigma}_{BM}$ в целом определяют запретную область $\overline{\Sigma}_B$ в пространстве B, которая описывается характеристической функцией запрета (ХФЗ) $\overline{\Phi}$ в виде конечного бинарного предиката, принимающего значение «1» только на элементах области $\overline{\Sigma}_B$. Предикат $\overline{\Phi}$ удобно выразить в дизъюнктивной нормальной форме (ДНФ) как дизъюнкции элементарных конъюнкций, отвечающих найденным *запретам*. Именно представление БкЗ в виде ХФЗ $\overline{\Phi}$ позволяет решать задачу прогнозирования динамических процессов как задачу экстраполяции интересуемых свойств ОПР по имеющимся наблюдениям и в частном случае получить решение задачи идентификации состояний процессов, то есть распознавания объектов. Специфику предлагаемой экстраполяции поясним геометрическими соображениями с неявным учетом времени прогноза с $\tau_{пр}$ в отличие от общепринятого использования явной зависимости наблюдаемого процесса *Y(t)* от времени *t*.

Пусть известна некоторая БкЗ = $k_2\overline{\Sigma}_{BM}$ в форме ХФЗ $\overline{\Phi}$, устойчивость закономерностей которой гарантируется на заданное время прогноза τ. Результаты частичных наблюдений заданы векторным квантом знаний 1-го уровня $k_1 y_\omega$, представляющим интервал y_ω пространства $B_r \subset B$, в котором локализован ОПР ω. Очевидно, что если этот интервал пересекается с запретной областью $\overline{\Sigma}_B$ данной БкЗ, то путем удаления из y_ω области пересечения получим остаток $\hat{y}_\omega = y_\omega \setminus \left(\overline{\Sigma}_B \cap y_\omega\right)$, объем и форма которого определяют возможный результат экстраполяции. Если интервал y_ω не пересекается с $\overline{\Sigma}_B$, то имеем результат «не знаю» в силу недостатка знаний в БкЗ. В случае полного пересечения y_ω с $\overline{\Sigma}_B$ приходим к результату, свидетельствующему о противоречивости БкЗ. Формально такое использование системы запретов БкЗ=$\overline{\Phi}$ при экстраполяции данных, представленных квантом наблюдений $k_1 y_\omega$, полностью реализуется оператором редукции БкЗ $k_1 y_\omega$ вида $RED\left(k_2\overline{\Sigma}_{BM}, k_1 y_\omega\right)$. Получаемые при этом редуцированные k-знания $k_1\hat{y}_\omega = RED\left(k_2\overline{\Sigma}_{BM}, k_1 y_\omega\right)$ содержат необходимую информацию для принятия *идентификационных* (*С-задача*) и *прогнозных* (*В-задача*) решений при анализе текущих ситуаций. *В-задача* состоит в определении (предсказании) неизменных значений s признаков ОПР по известным значениям *(n-s)* оставшихся признаков из общего числа n путем редукции БкЗ=$\overline{\Phi}$ по кванту наблюдений $k_1 y_\omega$ в данной предметной области. В *С-задаче* как в частном случае *В-задачи* требуется определить (предсказать) неизвестное значение i-го идентификационного признака $\left(i=\overline{1,n}\right)$ по измеренным значениям (n-1)-го признака распознаваемого класса объектов. Отсюда следует возможность построения единой количественной меры качества принимаемых прогнозных и идентификационных решений как некоторой степени сходства между наблюдаемым процессом $Y(t)$ и модельным прогнозным процессом $Z(t)$. Процессы Y и Z связаны с не регистрируемым напрямую реальным процессом $X(t)$ при знаниеориентированном решении проблемы предсказуемости. Количественно степень сходства между $Y(t)$ и $Z(t)$ будем представлять показателем корреляции между наблюдением и прогнозом, спустя время τ после начала наблюдения. Далее сформулируем рассматриваемую задачу оценивания качества знаниеориентированных решений и приведем методику ее решения средствами ИКЗ с обсуждением вопросов, связанных с предсказуемостью динамических процессов.

2. Постановка задачи

Согласно изложенным в пункте 1 посылкам, содержательная постановка задачи оценивания качества знаниеориентированных идентификационных и прогнозных решений состоит в определении единой количественной меры W их качества, опираясь на возможность решения B-, C-задач путем редуцирования характеристической функции $\overline{\Phi}$ импликативной БкЗ с неявным учетом времени. При этом мера W должна характеризовать потенциальную предсказуемость наблюдаемого (регистрируемого) процесса Y с использованием модельного процесса Z и знаний о реальном процессе X, который не подвергается прямой регистрации, а также позволять судить об основных причинах некачественного прогноза. Заметим, что динамический процесс $Z(t)$ в нашем случае представляется РАКЗ-моделью в виде ХФЗ $\overline{\Phi}$, где время t не является явным аргументом, а лишь определяет временной период гарантированной устойчивости системы закономерностей БкЗ. При этом ХФЗ $\overline{\Phi}$ одновременно выполняет роль БкЗ и служит аксиоматической основой дедуктивного механизма вывода искомых решений.

Формальная постановка задачи заключается в следующем. Пусть наблюдаемый процесс $Y(t)$ связан с реальным процессом $X(t)$ соотношением

$$Y(t) = X(t) + \sigma_\nu + \sigma_f + \sigma_{\Delta Q}\,, \qquad (2)$$

где величины σ_ν, σ_f и $\sigma_{\Delta Q}$ отражают соответственно искажения от инструментальных шумов при измерениях, от флуктуационных физических шумов и от влияния неадекватности модели прогнозирования $\Delta Q = Q_x - Q_z$ как от «шумов незнания». Реальный динамический процесс $X(t)$ может быть описан реальным оператором Q_x в символическом виде:

$$Q_x\left(S_x, x; \alpha_k, f_l\right) = 0\,, \qquad (3)$$

где S_x – математический язык представления процесса X; x – набор существенных переменных на принятом уровне описания; α_k – параметры процесса X; f_l – несущественные и флуктуационные переменные. Например оператор (3) для многих динамических систем часто имеет вид дифференциального уравнения

$$Q_x\left(d/dt, x; \alpha_k, f_l\right). \qquad (4)$$

В зависимости от того, как решаются вопросы выбора принципа прогнозирования, существенных и несущественных переменных, возникает потребность в упрощенном, идеализированном представлении реального процесса модельным процессом $Z(t)$ с оператором Q_z, аналогичным оператору (3):

$$Q_z\left(S_z, z, \hat{\alpha}_k, \varnothing\right) = 0\,. \qquad (5)$$

Здесь $\hat{\alpha}_k$ – оценки параметров процесса Z, а символ «$\varnothing$» означает пренебрежение несущественными факторами f_l, что и составляет идеализации. Поскольку в нашем случае Z представляется РАКЗ-моделью в виде ХФЗ $\overline{\Phi}(z)$, модельный оператор (5) с использованием квантов знаний зададим выражением:

$$Q_z\left(RED\left(k_2\overline{\Sigma}_{BM}, k_1 y_\omega, k_1\hat{y}_\omega\right),\ \overline{\Phi}(z),\ \hat{\alpha}_k\right) = 0 \qquad (6)$$

В развернутом виде выражение (6) является конечно-предикатным уравнением относительно набора существенных переменных z и оценок $\hat{\alpha}_k$ параметров процесса Z, а в квантовом представлении – дедуктивным преобразователем матричного кванта $k_2\overline{\Sigma}_{BM}$ = БкЗ по векторному кванту наблюдаемой ситуации $k_1 y_\omega$ (то есть ОПР ω) в результирующий прогнозный квант знаний $k_1\hat{y}_\omega$ посредством оператора редукции (RED). При этом значения наборов z разнотипных признаков ОПР измеряются как в количественных, так и в качественных шкалах, но для машинного манипулирования k-знаниями преобразуются в булеву форму (значение признака z_i=1, если ОПР обладает этим признаком и значение z_i=0, если не обладает). Таким образом, согласно квантовой прогнозной модели (6) принятие знаниеориентированных решений при условиях неопределенности в сформулированных B-и C-задачах реализуются с помощью редуцирования базы квантов знаний по текущей ситуации наблюдаемого процесса Y для получения прогнозируемой ситуации модельного процесса Z. Поэтому мы будем использовать общепринятую характеристику качества прогноза [3] в виде эмпирического усреднения квадрата ошибки $\eta = Y - Z$:

$$\left\langle \eta^2 \right\rangle = \left\langle (Y - Z)^2 \right\rangle, \qquad (7)$$

где угловые скобки ⟨, ⟩ упрощенно означают выбранный исследователем способ эмпирического усреднения. Например, если обозначить через $y_i = Y\left(t_j^0 + \tau\right)$ значение процесса $Y(t)$, спустя время τ после момента t_j^0 начала j-го наблюдения, а через $z_j = Z\left(t_j^0 + \tau_{\text{пр}}\right)$ – соответствующий прогноз на интервале $\left(t_j^0, t_j^0 + \tau_{\text{пр}}\right)$, удовлетворяющий начальному условию $z_j^0 = y_j^0$, то средний квадрат ошибки (7) можно определить по формуле

$$\left\langle \eta^2 \right\rangle = \frac{1}{N}\sum_{j=1}^{N}\left[Y\left(t_j^0 + \tau_{\text{пр}}\right) - Z\left(t_j^0 + \tau_{\text{пр}}\right)\right]^2, \qquad (8)$$

где N – число наблюдений.

Очевидно, чем больше N, тем надежнее оценка погрешности (8). Для того, чтобы характеризовать при прогнозировании потенциальную предсказуемость, в роли единой количественной меры качества знаниеориентированных решений будем использовать безразмерную величину

$$W(\tau) = \frac{\left\langle Y \cdot Z \right\rangle}{\sqrt{\left\langle Y^2 \right\rangle \cdot \left\langle Z^2 \right\rangle}}\,, \qquad (9)$$

которая является эмпирическим коэффициентом корреляции между наблюдением и прогнозом через время τ после наблюдения.

Руководствуясь приведенными выше соображениями, принимаем в качестве исходных следующие данные:

— заблаговременно синтезированная

$$\text{БkЗ} = k_2 \overline{\Sigma}_{BM} = \overline{\Phi}_\tau$$

с гарантированной устойчивостью ее закономерностей на заданном интервале времени прогнозирования τ для рассматриваемой предметной области;

— выборочные N пар значений (реализаций) наблюдаемого процесса $Y(t)$ (2) и модельного процесса $Z(t)$, вычисляемых по формуле (6) для данной предметной области.

Требуется на основании использования исходных данных и формул (7)-(9) разработать методику определения единой меры качества $W(\tau)$ принимаемых знаниеориентированных и прогнозных решений независимо от природы предметной области. Учитывая корреляционный характер количественной меры $W(\tau)$, установить ее связь с абсолютной погрешностью прогноза $\langle \eta^2 \rangle$ (7), с причинами неуспешного прогнозирования и с некоторыми характеристиками предсказуемости динамических процессов.

3. Решение поставленной задачи: методика вычисления меры качества $W(\tau)$

Изложим методику определения единой меры качества идентификационных и прогнозных решений $W(\tau)$ по заданным $\text{БkЗ} = k_2 \overline{\Sigma}_{BM} = \overline{\Phi}_\tau$ и N выборочным парам реализаций двух процессов $Y(t)$ и $Z(t)$ как эмпирического распределения двух случайных величин Y и Z. Статистические оценки характеристик такой двумерной генеральной совокупности при неизвестном теоретическом распределении будем вычислять по рекомендациям в [6]. Заметим, что трактовка случайности как непредсказуемости является лишь одним из многих соглашений о случайности, принятых в современной науке. Мы будем следовать соглашению, принятому в экспериментальной физике, где случайность (непредсказуемость) связывают с потерей корреляции. Например для процесса $Y(t)$, характеризуемого коэффициентом корреляции

$$K_y(\tau) = \frac{\langle Y(t) \cdot Y(t-\tau) \rangle}{\langle Y^2 \rangle}, \qquad (10)$$

значения $Y(t)$, которые разделены временем, превышающим время корреляции τ_c, считаются непредсказуемыми [3].

Итак, методику определения меры качества $W(\tau)$ по формуле (9) изложим в виде последовательности действий и формул статистического оценивания необходимых эмпирических характеристик распределения значений случайных величин Y и Z, обоснованных в справочнике [6]. Сначала для N выборочных пар реализаций Y и Z выбирают размеры и положение разрядов $\left(y'_{i-1}, y'_i\right)$, $\left(z'_{j-1}, z'_j\right)$, на которые делят область эмпирического распределения Y и Z в отдельности. Число разрядов для Y и Z рекомендуется брать не более 10, так как при увеличении их числа точность оценок изменяется несущественно, а объем вычислений быстро возрастает. Полученные данные сводят в корреляционную таблицу (табл. 1), в каждую клетку которой заносится частота N_{ij} — число пар измерений, удовлетворяющих условиям

$$y'_{i-1} < y < y'_i, \quad z'_{j-1} < z < z'_j. \qquad (11)$$

Этим реализациям приписываются значения:

$$y_i = \frac{1}{2} \cdot \left(y'_{i-1} + y'_i\right), \quad z_j = \frac{1}{2} \cdot \left(z'_{j-1} + z'_j\right). \qquad (12)$$

В нижнюю строки табл. 1 заносят частоты N_{yi}, а в крайний правый столбец — частоты N_{zj}, определяемые по формулам:

$$N_{yi} = \sum_{j=1}^{m_z} N_{ji}, \quad N_{zj} = \sum_{i=1}^{m_y} N_{ij}, \qquad (13)$$
$$\left(i = 1, 2, \ldots, m_y\right), \quad \left(j = 1, 2, \ldots, m_z\right),$$

где m_y — число разрядов для величины Y; m_z — число разрядов для величины Z.

По данным табл. 1, используя формулы (11)-(13), определяют условные выборочные средние $\overline{Y}(z_j)$ и $\overline{Z}(y_i)$:

$$\overline{Z}(y_i) = \frac{1}{N_{yi}} \cdot \sum_{j=1}^{m_z} N_{ji} \cdot z_j, \quad (i = 1, 2, \ldots, m_y); \qquad (14)$$

$$\overline{Y}(z_j) = \frac{1}{N_{zj}} \cdot \sum_{i=1}^{m_y} N_{ij} \cdot y_i, \quad (j = 1, 2, \ldots, m_z). \qquad (15)$$

Таблица 1

		y_i					
z_j	i / j	1	2	…	m_y	N_{zj}	$\overline{Y}(z_j)$
	1						
	2						
	…						
	m_z						
N_{yi}							
$\overline{Z}(y_i)$							

Далее, на основании использования формул (12)–(15) и данных табл. 1 вычисляют выборочные средние $\overline{Y}$ и $\overline{Z}$ по формулам:

$$\overline{Y} = \frac{1}{N} \cdot \sum_{i=1}^{m_y} N_{yi} \cdot y_i = \frac{1}{N} \cdot \sum_{j=1}^{m_z} N_{zj} \cdot \overline{Y}(z_j); \qquad (16)$$

$$\overline{Z} = \frac{1}{N} \cdot \sum_{j=1}^{m_z} N_{zj} \cdot z_j = \frac{1}{N} \cdot \sum_{i=1}^{m_y} N_{yi} \cdot \overline{Z}(y_i). \qquad (17)$$

Используя формулы (12), (13), (16) и (17), а также соответствующие данные табл. 1, определяют эмпирические дисперсии S_{1y}^2 и S_{1z}^2 по формулам:

$$S_{1y}^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{m_y} N_{yi} \left(y_i - \overline{Y}\right)^2; \qquad (18)$$

$$S_{1z}^2 = \frac{1}{N-1} \cdot \sum_{j=1}^{m_z} N_{zj} \left(z_j - \overline{Z} \right)^2 . \tag{19}$$

Наконец, на основании использования выражений (12), (13), (16)-(19) приходим к соотношению для вычисления эмпирического коэффициента корреляции $\bar{r}$:

$$\bar{r} = \frac{\sum_{i=1}^{m_y} \sum_{j=1}^{m_z} N_{ij} \cdot y_i \cdot z_j - N \cdot \overline{Y} \cdot \overline{Z}}{(N-1) \cdot \sqrt{S_{1y}^2 \cdot S_{1z}^2}} = W(\tau) , \tag{20}$$

которое и является рабочей формулой для вычисления единой количественной меры качества знаниеориентированных решений $W(\tau)$ (9). Численное значение меры $W(\tau)$ получаем путем подстановки в формулу (20) значений вычисленных величин N_{ij}, y_i, z_j, $\overline{Y}$, $\overline{Z}$, S_{1y}^2 и S_{1z}^2 для заданной эмпирической N – выборки по указанным формулам. При $N > 70 \div 80$ величина $\bar{r} = W(\tau)$ (20) приближенно подчиняется нормальному распределению, что позволяет оценивать меру качества $W(\tau)$ с помощью доверительной вероятности

$$\gamma = P\left\{ \bar{r} - \beta \frac{1-\bar{r}^2}{\sqrt{N}} < W(\tau) < \bar{r} + \beta \frac{1-\bar{r}^2}{\sqrt{N}} \right\} \approx 2\Phi_0(\beta) ,$$

используя методику расчетов [6] (см. § 14.3, пример 14.3.1, табл. III).

4. Связь меры качества $W(\tau)$ с потенциальной предсказуемостью

Условимся называть сроком (временем) прогнозирования интервал времени $\tau = t - t^0$ после начала наблюдения t^0. При исследовании проблемы предсказуемости мы всегда вынуждены рассматривать три процесса**:** реальный процесс $X(t)$, наблюдаемый процесс $Y(t)$ и модельный прогнозный процесс $Z(t)$, которые схематично изображены на рис. 1.

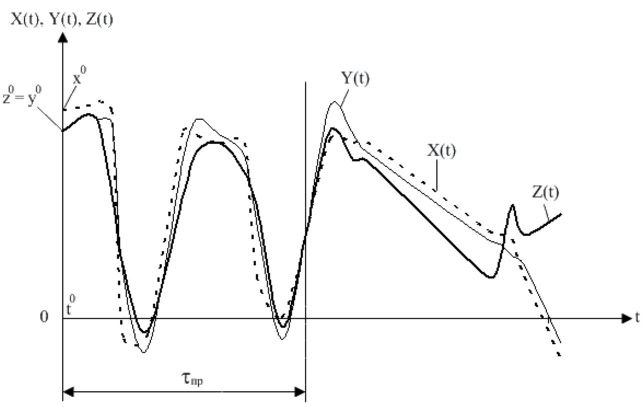

Рис. 1. Процессы $X(t)$, $Y(t)$, $Z(t)$

В силу корреляционного характера величина меры качества прогнозирования $W(\tau)$ равна 1 при $\tau = t - t^0 = 0$ поскольку начальное значение прогноза $z_0 = y_0$, а с течением времени уменьшается до 0. Общий вид зависимости $W(\tau)$ показан на рис. 2. Близкие к 1 значения $W(\tau)$ отвечают удовлетворительному прогнозу, а малые значения соответствуют неудовлетворительной предсказуемости из-за слабой согласованности наблюдения и прогноза. Поэтому величину меры $W(\tau)$ будем называть *степенью предсказуемости*. Время, за которое величина $W(\tau)$ снижается до уровня 0,5, назовем *временем предсказуемости* (см. рис. 2) с обозначением $\tau_{\text{пр}}$. Степень предсказуемости $W(\tau)$ однозначно связана с абсолютной погрешностью прогноза $\langle \eta^2 \rangle$ (7):

$$W(\tau) = \frac{\langle Y^2 \rangle + \langle Z^2 \rangle}{2 \cdot \sqrt{\langle Y^2 \rangle \cdot \langle Z^2 \rangle}} \cdot \left[1 - \frac{\langle \eta^2 \rangle}{\langle Y^2 \rangle + \langle Z^2 \rangle} \right] . \tag{21}$$

Времени предсказуемости $\tau_{\text{пр}}$ отвечает абсолютная погрешность $\langle \eta^2 \rangle$ порядка

$$0{,}5 \cdot \left(\langle Y^2 \rangle + \langle Z^2 \rangle \right) \approx \langle Y^2 \rangle ,$$

то есть порядка дисперсии наблюдаемого процесса *Y(t)*. Если высокую степень предсказуемости ($W(\tau) \approx 1$) рассматривать как свидетельство предсказуемой (детерминированной) динамики процесса, а малые значения $W(\tau)$ приписать непредсказуемому, случайному поведению, то $W(\tau)$ можно трактовать как степень детермированности процесса $Y(t)$ по отношению к модельному процессу $Z(t)$ [3].

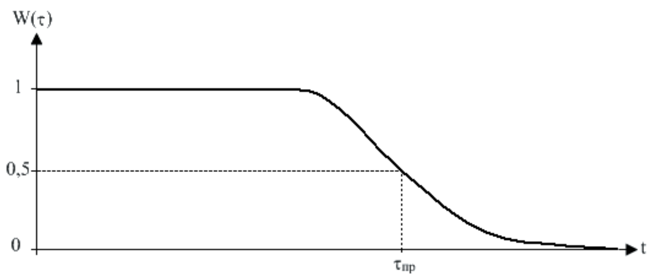

Рис. 2. Зависимость степени предсказуемости W от времени $\tau = t - t^0$, протекшего после начала наблюдения

В экспериментальной физике установлено, что коэффициент корреляции $K_Y(\tau)$ (10) наблюдаемого процесса $Y(t)$ является наименьшей (наихудшей) степенью предсказуемости, а время корреляции τ_c – наименьшим (наихудшим временем прогнозируемого поведения $Y(t)$. В интервале $\tau_c < \tau < \tau_{\text{пр}}$ один и тот же процесс $Y(t)$ следует классифицировать как случайный, если судить в рамках соглашения, отождествляющего случайность с потерей корреляции и как детерминированный с точки зрения отождествления случайности с непредсказуемостью [3]. Нижний предел $\tau_{\text{пр}} = \tau_c$ проявляется при условиях, когда исследователь не располагает динамическим уравнением (4) и для прогноза использует принцип «завтра будет как сегодня». В этом случае в качестве прогноза $z(t) = z(t^0 + \tau)$ принимает значение $y(t^0)$, наблюдаемое в данный момент t:

$$z(t+\tau) = y(t^0) \text{ или } z(t) = y(t-\tau). \qquad (22)$$

При этом степень предсказуемости $W(\tau)$ (9) превращается в коэффициент корреляции $K_Y(\tau)$ (10).

Результат сопоставимости $\tau_{пр}$ со временем корреляции τ_c проиллюстрируем на примере зашумленного динамического хаоса из [3]. Рассматривается наблюдаемый хаос $Y(t)$ с наибольшим положительным показателем Ляпунова λ_+ со средним квадратом ошибки прогноза, нарастающем по экспоненциальному закону

$$\left\langle \eta^2 \right\rangle = \left(\sigma_\nu^2 + \sigma_f^2 + \sigma_{\Delta Q}^2\right) \cdot \exp^{(2\cdot\lambda_+\cdot\tau)}.$$

Здесь величины σ_ν^2, σ_f^2 и $\sigma_{\Delta Q}^2$ отражают соответственно вклад инструментальных шумов, физических шумов и «шумов незнания» подобно (2). Полученная оценка времени предсказуемого поведения $Y(t)$ имеет вид:

$$\tau_{пр} \approx \frac{1}{2\cdot\lambda_+} \cdot \ln \frac{\left\langle Y^2 \right\rangle}{\sigma_\nu^2 + \sigma_f^2 + \sigma_{\Delta Q}^2}. \qquad (23)$$

Поскольку порядок величины $1/(2\lambda_+)$ отвечает времени корреляции процесса $\tau_c \approx 1/2\lambda_+$, отношение $\tau_{пр}/\tau_c$ характеризуется значением величины, стоящей под. знаком логарифма в (23). Если, например, шумовые слагаемые σ_ν^2, σ_f^2 и $\sigma_{\Delta Q}^2$ составляют одну тысячную (10^{-3}) от среднеквадратичной амплитуды $A_Y \approx \left\langle Y^2 \right\rangle^{1/2}$, то $\tau_{пр} \approx \tau_c \cdot \ln 10^3 \approx 7\tau_c$, то есть $\tau_{пр}$ в 7 раз превышает τ_c. Здесь вскрывается важность понятия *горизонта предсказуемости* τ_{hor}, под которым будем понимать предельное время предсказуемого поведения процесса $Y(t)$, которое нельзя превзойти ни усовершенствованием измерений, ни улучшением предсказательной модели $Z(t)$. Иными словами, τ_{hor} отвечает пределу пренебрежимо малых измерительных шумов и шумов незнания, что равносильно в нашем примере хаоса пренебрежению вкладами σ_ν^2 и $\sigma_{\Delta Q}^2$ по сравнению с вкладом физических шумов σ_f^2, влияние которых невозможно устранить ни при каких условиях:

$$\tau_{hor} = \frac{1}{2\lambda_+} \ln \frac{\left\langle Y^2 \right\rangle}{\sigma_f^2}.$$

Таким образом, изложенные выше соображения позволяют представить на рис. 3 общий характер изменения степени предсказуемости $W(\tau)$ по мере совершенствования прогностической модели Z(t).

Кривая ① показывает, что на малых временах, не превышающих время корреляции τ_c, делается удовлетворительный прогноз по принципу «завтра как сегодня», когда $W(\tau)$ (9) совпадает с коэффициентом корреляции $K_Y(\tau)$ (10). Кривые ② и ③ отвечают более или менее удовлетворительным прогнозным моделям $Z_1(t)$ и $Z_2(t)$, когда при $\tau_{пр} > \tau_c$ по мере совершенствования моделей степень предсказуемости $W(\tau)$ вместе со временем $\tau_{пр}$ увеличиваются. На этом этапе посредством использования РАКЗ-модели (6) процесс $Y(t)$ можно превратить из непредсказуемого (точка w_1 на кривой ②, отвечающей модели $Z_1(t)$) в предсказуемый (детерминированный) (точка w_2 на кривой ③, отвечающей модели $Z_2(t)$). Переход $w_1 \to w_2$ от модели Z_1 к более совершенной РАКЗ-модели Z_2 переводит процесс из разряда случайных (непредсказуемых) к детерминированным (предсказуемым) и возможен только в рамках концепции, отождествляющей случайное с непредсказуемым. В рамках аксиоматической теории вероятностей такой переход невозможен в принципе, так как процесс остается случайным независимо от наличия либо отсутствия модели для его предсказания. Наконец, кривая ④ на рис. 3 отвечает предельно достижимой степени предсказуемости $W(\tau)$ за предельное время предсказания τ_{hor} («горизонт предсказуемости»), начиная с которого никакая модель не может обеспечить удовлетворительное предсказание.

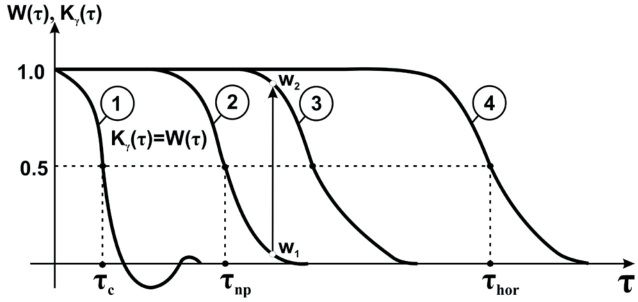

Рис. 3. Характер изменения степени предсказуемости $W(\tau)$ по мере совершенствования прогнозной модели $Z(t)$

Следовательно, предложенные оценки степени предсказуемости как количественной меры $W(\tau)$ (20) вместе с горизонтом предсказуемости τ_{hor} имеют принципиальное значение – нет смысла строить прогнозы на сроки, превышающие τ_{hor}, в силу неоправданных потерь времени и средств.

Выводы

1. На основе использования прогнозных РАКЗ-моделей вида (6) и меры качества решений $W(\tau)$ (20) в инженерии квантов знаний выделены и исследованы основные причины, препятствующие успешному прогнозированию. Это – локальная неустойчивость исследуемых процессов, действие флуктуационных сил и шумов измерения, эволюция динамики процессов, неадекватность (неточность) модельного оператора Q_z (5) относительно реального оператора Q_x (3) и финансовые ограничения.

2. Любой прогнозный процесс имеет ограниченное время предсказуемого поведения, которое зависит от указанных в п. 1 основных причин. Существует максимальное значение этого времени – так называемый «горизонт предсказуемости», обусловленный действием только неустранимых флуктуаций самого процесса, за пределы которого выйти невозможно.

3. Успешный прогноз достигается главным образом за счет удачного выбора модельного прогнозного оператора Q_z (6), что послужило мотивацией для разработки знаниеориентированной РАКЗ-модели принятия идентификационных и прогнозных решений в виде характеристической функции запретов $\overline{\Phi}$ с неявным учетом времени.

Эффективность предложенной меры качества $W(\tau)$ знаниеориентированных решений и приведенные в заключении выводы подтверждены результатами решения многочисленных практических задач [2].

Список литературы. 1. *Сироджа И.Б.* Математическое и программное обеспечение интеллектуальных компьютерных систем: Учеб. пособ. – Харьков: ХАИ, 1992. – 100 с. 2. *Сироджа И.Б.* Квантовые модели и методы искусственного интеллекта для принятия решений и управления. – К.: Наукова думка, 2002. – 490 с. 3. *Кравцов Ю.А.* Фундаментальные и практические пределы предсказуемости // Пределы предсказуемости: Науч.-техн. сб. – М.: Наука, 1996. С. 170-1999. 4. *Люгер Дж. Ф.* Искусственный интеллект: стратегии и методы решения сложных проблем, 4-е издание: Пер. с англ. – М.: Издательский дом «Вильямс», 2003. – 863 с. 5. *Закревский А.Д.* Выявление импликативных закономерностей в булевом пространстве признаков и распознавние образов // Кибернетика. – 1982. – №1. – С. 1-6. 6. *Абезгауз Г.Г., Тронь А.П., Копенкин Ю.Н., Коровина И.А.* Справочник по вероятностным расчетам. – М.: Воен. издат. МО СССР, 1970. – 532 с.

Поступила в редколегию 22.09.2008

УДК 004.89

Оцінювання якості ідентифікаційних та прогнозних рішень в інженерії квантів знань / І.Б. Сіроджа // Біоніка інтелекту: наук.-техн. журнал – 2008. – № 2 (69). – С. 77-83.

Розроблено логічну модель екстраполяції спостережень у вигляді предикатної системи імплікативних закономірностей предметної області як бази квантів знань. Вдалося звести задачу прийняття ідентифікаційних рішень до частого випадку прийняття прогнозного рішення. При цьому загальна задача прийняття прогнозних рішень запропонована по новому, з неявним обліком змінної часу (на відміну від традиційної постановки з явним обліком часу), завдяки використанню предикатної характеристичної функції для екстраполяції стійких закономірностей динаміки процессу, що прогнозується у фазовому просторі.

Іл. 3. Бібліогр.: 6 найм.

UDK 004.89

Estimating qualities identification and prognostic decisions in engineering of quantes of knowledge / I. Sirodzha // Bionics of Intelligence: Sci. Mag. – 2008. – № 2 (69). – P. 77-83.

The logic model are develops of extrapolation observation in the form of predicate system implicational laws of a object domain, as bases of quantes of knowledge (БкЗ). It was possible to reduce a problem of acceptance the identification decision to an acceptance special case prognostic decisions. Thus the general problem of acceptance prognostic decisions is posed with the implicit account of a time variable (unlike traditional posed problem with the obvious account of time), thanks to using of predicate characteristic function for extrapolation of steady laws of dynamics predictable process in phase space.

Fig. 3. Ref.: 6 items