

Информационные технологии

УДК 519.816+519.876.5

ПРОДУКЦИОННЫЕ МОДЕЛИ ФОРМАЛИЗАЦИИ ДАННЫХ И ЗНАНИЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ АНАЛИЗА ГЕОЛОГОРАЗВЕДОЧНЫХ ПРОЕКТОВ

В.И. Батищев, Н.Г. Губанов, Ю.В. Тышковская

Самарский государственный технический университет

443100, г. Самара, ул. Молодогвардейская, 244

E-mail: Nick_G_Gubanov@mail.ru

Рассмотрены вопросы представления данных в информационных системах анализа. На основе синтеза фреймовой и продукционной моделей предложен формальный аппарат формирования баз знаний геологоразведочных проектов.

Ключевые слова: геологоразведочные проекты, продукционная система, структурная таксономия.

Методы формирования систем анализа геологоразведочных предприятий.

Проблема эффективного управления системами обеспечения проведения геологоразведочных работ охватывает широкий круг задач – от проектирования и производства работ до обеспечения безопасной эксплуатации потенциально опасных производственных объектов. Первоочередной задачей управления геологоразведочными работами является создание такой стратегии, которая была бы направлена на формирование оптимальных производственных показателей и ориентирована на работу в условиях динамично меняющейся внешней среды [1]. Решение подобной задачи требует применения системного подхода, предполагающего разработку инструментария прогнозирования, принятия решений, оценку результатов принимаемых решений, корректировку управленческих решений и направленного на создание интегрированной системы обеспечения комплексной безопасности предприятия. Задачами такой системы являются упорядочение, гармонизация и интеграция различных средств организационного, методического, технического и программного обеспечения комплексов средств мониторинга и обеспечения безопасности основных процессов предприятия при воздействии на них внутренних и внешних угроз, планирования общих и частных мероприятий по идентификации угроз безопасности и минимизации ущерба в деятельности предприятия [2].

В настоящее время особое внимание уделяется вопросам создания интегрированной информационной среды предприятия, содержащей актуальные знания и включающей методическое, алгоритмическое, информационное, техническое, программное обеспечение информационных систем предприятия [3].

Виталий Иванович Батищев – д.т.н., профессор.

Николай Геннадьевич Губанов – к.т.н., доцент.

Юлия Владимировна Тышковская – преподаватель.

Структура и функционирование аналитических систем S зависит от следующих информационных сущностей: объекта анализа (ГРП) Q ; цели функционирования аналитической системы G , определяемой конкретной задачей принятия решения; полимодельного комплекса, задающего структуру системы, M ; среды, определяющей параметры системы, C , а также отношений между данными структурами:

$$R = (r_{Q,M}, r_{Q,C}, r_{Q,G}, r_{M,C}, r_{Q,M}, r_{Q,C}).$$

Воспользовавшись общесистемным подходом, систему оценки геологоразведочного проекта (ГРП) представим в виде структуры, состоящей из системы управления СУ, объекта управления ОУ и среды [4]. В нашем случае с учетом специфики предметной области данная формализация представится в следующем виде:

$$S = \langle G, P, L \rangle,$$

где G – система управления – представлена соответствующими государственными органами в структурах управления минерально-сырьевой базы РФ, осуществляющими поддержку инвестиционной программы P^{inv} , где множество инвестиционных проектов p_i^{inv} .

$$P^{inv} = \{p_1^{inv}, p_2^{inv}, \dots, p_n^{inv}\} = \bigcup_{i=1}^n p_i^{inv}.$$

Для определенности будем считать, что инвестиционная программа ГРП направлена на достижение глобальной цели – перспективного развития процесса накопления знаний и оценки запасов минерально-сырьевой базы страны, которую будем считать оптимальной.

Система управления характеризуется совокупностью целей C и управляющих воздействий U :

$$G = C \cup U.$$

Согласно алгоритму оценка эффективности ГРП осуществляется в два этапа. На первом этапе оценивается эффективность проекта в целом при допущении, что финансирование идет только за счет бюджетных средств; результаты первого этапа являются основой для формирования схемы финансирования с привлечением частных инвесторов. На первом этапе осуществляется последовательное решение задач: оценки общественной значимости, оценки общественной эффективности и оценки коммерческой эффективности.

Управляющие воздействия проявляются принятием решений по заданному инвестиционному проекту на различных этапах оценки эффективности $U = (U_{oz}, U_{oэ}, U_{кэ})$, где U_{oz} – решения на этапе оценки общественной значимости проекта; $U_{oэ}$ – решения на этапе оценки общественной эффективности проекта; $U_{кэ}$ – решения на этапе оценки коммерческой эффективности проекта.

Среда характеризует рынок и включает в себя следующие субъекты геологоразведочного проекта: подрядные организации, инвестиционные институты, транспорт, энергетиков, осуществляющих поддержку процесса геологоразведочных работ на всех этапах жизненного цикла. Данные субъекты на основании рыночных условий, сложившихся в данном регионе, формируют величины затратных и доходных статей геологоразведочного проекта.

Идентификация состояния и характера среды осуществляется в соответствии с этапом принятия решения:

- на первом этапе, когда осуществляется оценка эффективности ГРП в целом, среда является пассивной и нецеленаправленной подсистемой;

– на этапе формирования схемы финансирования среда проявляется как активная и целенаправленная подсистема и наравне с управляющей подсистемой вырабатывает управляющие воздействия.

Объектом управления является непосредственно сам геологоразведочный проект P . Комплексной характеристикой ГРП является описание его жизненного цикла (ЖЦ). Жизненный цикл описывает наименование и стоимостные характеристики параметров ГРП в период времени от момента принятия решения о необходимости инвестирования средств до завершения ЖЦ объекта. Весь жизненный цикл обычно разбивают на три фазы: $ГРП = (P_{пред}, P_{инв}, P_{эксп})$, где $P_{пред}$ – параметры ГРП предынвестиционной фазы, $P_{инв}$ – параметры ГРП инвестиционной фазы, $P_{эксп}$ – параметры ГРП эксплуатационной фазы.

Инвестиционный проект P описывается тремя классами параметров: затратами, эффектом и возможными рисками.

Система экспертизы ГРП основывается на определенной модели инвестиционного проекта. Создание математической модели затруднительно в силу наличия множества неопределенных параметров, сложных взаимосвязей между ними, а также наличия латентных – трудно идентифицируемых – характеристик ГРП, например рисков.

В то же время накоплен достаточно большой опыт проведения геологоразведочных работ. Существуют информационные базы внедренных инвестиционных проектов, даны оценки их последствий, рисков и эффективности. Предпочтительно было бы воспользоваться накопленными знаниями для поддержки принятия решений реализации текущих инвестиционных проектов.

Для этой цели необходимо построить прецедентную модель на основе накопленных данных, сформировав базу знаний инвестиционных проектов путем выявления некоторых скрытых закономерностей.

Обосновать данный методологический подход можно на основе гипотезы компактности, которая состоит в том, что реализации одного и того же образа обычно отражаются в признаковом пространстве в геометрически близкие точки, образуя «компактные» сгустки. Если принять, что информативные признаки образов близки, то и целевые признаки, указывающие имя образа, тоже близки. Данная гипотеза равнозначна предположению о наличии закономерной связи между признаками. Необходимо определенными методами сформировать базу знаний (БЗ), содержащую информацию о закономерностях развития промышленных объектов.

Одной из ключевых научных задач при формировании информационных систем анализа геологоразведочных проектов является создание формального аппарата представления данных и знаний. Решение данной задачи является комплексной методологической основой для проектирования подсистем сбора и обобщения данных, а также отбора и адаптации моделей геологоразведочного проекта [5].

В качестве формального аппарата представления знаний в информационных системах анализа геологоразведочных проектов использован синтез фреймовой и продукционной моделей представления знаний, где в продукционной системе в качестве множества входных и множества выходных литералов используется иерархическая фреймовая структура, а множество продукций с функциями присвоения формируют правила формирования и обработки данных структур.

Информацию об инвестиционном геологоразведочном проекте можно структурировать через описание его жизненного цикла. Таким образом, базу данных будем

рассматривать как фрейм $D = (D_1, D_2, \dots, D_k) = \bigcup_{i=1}^k D_i$ где D_i – данные i -того инвестиционного геологоразведочного проекта предприятия $D_i = \{(\langle D_i^n \rangle, H), (\langle D_i^u \rangle, F), (\langle D_i^o \rangle, E)\}$, где D_i^n – имя фрейма (освоенные геологоразведочные проекты на предынвестиционном этапе); D_i^u – имя фрейма (освоенные геологоразведочные проекты на инвестиционном этапе); D_i^o – имя фрейма (освоенные геологоразведочные проекты на эксплуатационном этапе); $\langle H \rangle$ – имя слота = (параметры геологоразведочных проектов на предынвестиционном этапе); $\langle F \rangle$ – имя слота = (параметры геологоразведочных проектов на инвестиционном этапе); $\langle E \rangle$ – имя слота = (параметры геологоразведочных проектов на эксплуатационном этапе).

В качестве заполнителей слотов содержится информация о трех классах параметров: затраты, эффект и возможные риски на разных этапах создания геологоразведочного проекта.

Базу знаний будем рассматривать как фрейм

$$Z = (Z_1, Z_2, \dots, Z_g) = \bigcup_{i=1}^g Z_i,$$

где Z_i – знания i -того геологоразведочного проекта,

$$Z_i = \{(\langle Z_i^n \rangle, Y), (\langle Z_i^u \rangle, T), (\langle Z_i^o \rangle, J)\},$$

где Z_i^n – имя фрейма (знания геологоразведочного проекта на предынвестиционном этапе); Z_i^u – имя фрейма (знания геологоразведочного проекта на инвестиционном этапе); Z_i^o – имя фрейма (знания геологоразведочного проекта на эксплуатационном этапе); $\langle Y \rangle$ – имя слота = (параметры геологоразведочного проекта на предынвестиционном этапе) $\langle T \rangle$ – имя слота = (параметры геологоразведочного проекта на инвестиционном этапе); $\langle J \rangle$ – имя слота = (параметры геологоразведочного проекта на эксплуатационном этапе).

Рассмотрим более подробно эти массивы

$$Y = \{Y_1, Y_2, \dots, Y_l\} = \bigcup_{j=1}^l Y_j \quad Y_j = \bigcup_{i=1}^u \{(\langle y_{ij} \rangle, y_{ij})\},$$

где Y_j – j -тое знание предынвестиционного этапа геологоразведочного проекта; $\langle y_{ij} \rangle$ – имя слота (i -тая характеристика знания геологоразведочного проекта), y_{ij} – заполнитель слота, где $y_{ij} = (H_{ij}, r_{ij}^n) r_{ij}^n$ – i -тый результат предынвестиционного этапа $T = \{T_1, T_2, \dots, T_k\} = \bigcup_{j=1}^k T_j \quad T_j = \bigcup_{i=1}^u \{(\langle t_{ij} \rangle, t_{ij})\}$, где T_j – j -тое знание инвестиционного этапа геологоразведочного проекта; $\langle t_{ij} \rangle$ – имя слота (i -тая характеристика знания

геологоразведочного проекта \rangle ; t_{ij} – заполнитель слота, где $t_{ij} = (F_{ij}, r_{ij}^u)$ r_{ij}^u – i -тый результат инвестиционного этапа, $J = \{J_1, J_2, \dots, J_n\} = \bigcup_{j=1}^n J_j$,

$J_m = \bigcup_{i=1}^n \{(j_{im}, \langle j_{im} \rangle)\}$, где J_m – m -тое знание инвестиционного этапа ГРП; $\langle j_{im} \rangle$ – имя слота $\langle i$ -тая характеристика знания геологоразведочного проекта \rangle ; j_{im} – заполнитель слота, где $j_{im} = (E_{im}, r_{ij}^o)$ r_{ij}^o – i -тый результат эксплуатационного этапа.

Модель геологоразведочного проекта рассматривается как фрейм $M = (M_1, M_2, \dots, M_v) = \bigcup_{i=1}^v M_i$, где M_i – модель i -того инвестиционного проекта, $M_j = \bigcup_{i=1}^h \{(m_{ij}, \langle m_{ij} \rangle)\}$, где $\langle m_{ij} \rangle$ – имя слота $\langle i$ -тая характеристика модели геологоразведочного проекта \rangle , $m_{ij} = (v_{ij}^{in}, v_{ij}^{out})$, v_{ij}^{in} , v_{ij}^{out} – входные и выходные характеристики модели геологоразведочного проекта соответственно. Фрейм эффективности геологоразведочного проекта представлен в виде

$$Q = \{Q_1, Q_2, \dots, Q_k\} = \bigcup_{j=1}^k Q_j \quad Q_j = (q_j^{os}, q_j^{os}, q_j^{k3}),$$

где q_j – показатель эффективности j -того геологоразведочного проекта; q_j^{os} – показатель общественной значимости j -того геологоразведочного проекта; q_j^{os} – показатель общественной эффективности j -того геологоразведочного проекта; q_j^{k3} – показатель коммерческой эффективности j -того геологоразведочного проекта. Схема финансирования геологоразведочного проекта представлена в виде: $N = \{kf, gf\}$, где kf и gf – величины коммерческого и бюджетного финансирования соответственно.

Такое наполнение фреймовой модели позволяет говорить о ГРП как о структурном объекте. Рассмотрим базу знаний формирования ГРП. При построении баз знаний (БЗ) ключевыми вопросами становятся вопросы формализации знаний, методов формирования и алгоритмов обработки знаний [6]. Вначале необходимо определить, что понимать под знаниями инвестиционного проекта. В качестве обоснованного с точки зрения предметной области, применяемого подхода в моделировании с учетом перспективы проектирования выбран следующий подход к определению знаний. Знаниями в данном случае являются нетривиальные практически полезные и доступные интерпретации сведения, необходимые для принятия решений в инвестиционном проектировании, заключающиеся в определенных закономерностях и систематических взаимосвязях между переменными, которые затем можно применить для исследования новых совокупностей данных.

Таким образом, знания (БЗ) геологоразведочных проектов можно представить в виде: $P=(D, F, R)$, где D – данные о проектах; F – характеристики последствий проекта; R – правила формирования знаний инвестиционного геологоразведочного проекта. При таком подходе возникает проблема необходимости формирования огромного числа продукций, в связи с чем предпочтительно увеличить представительность выборки для более точного формирования модели, что приводит к существенному

увеличению вычислительной мощности задачи. Данная проблема имеет ряд решений, заключающихся в применении методов снижения размерности задачи. Однако сложность применения «чистых методов» прикладной статистики заключается в практическом формировании обучающей выборки. Нет гарантированных условий достаточной представительности обучающей выборки. Поэтому видится целесообразным использование методов, основанных на ряде эмпирических гипотез – уже упомянутой гипотезе компактности, гипотезе монотонности решений – и, если позволяют условия, более сильных гипотез: гипотезы λ -компактности и т.п. На основании данных допущений можно применять методы кластеризации или таксономии, делая некоторые выводы о закономерностях параметров ГРП по достаточно бедной выборке.

Базу правил P можно представить:

$$P = (P_z, P_m, P_o, P_f),$$

где P_z – правила формирования базы знаний геологоразведочного проекта; P_u – правила формирования модели геологоразведочного проекта; P_k – правила комплексной оценки геологоразведочного проекта; P_w – правила формирования схемы финансирования геологоразведочного проекта.

Рассмотрим более подробно каждую систему продукций

$$P_z = \langle D, Z, R_z, O_z \rangle,$$

где D – фрейм: множество параметров геологоразведочного проекта, Z – фрейм: знания ИП, R_z – множество продукций выявления знания Z с помощью данных

$$D : R_z = \{R_z^n, R_z^u, R_z^o\},$$

где R_z^n – продукция (правило формирования знаний геологоразведочного проекта на предынвестиционном этапе)

$$R_z^n; D^n \cup Z^n \exists H_j \cup O_z \rightarrow Z_j^n, j = j+1, 1 \leq j \leq l;$$

R_z^u – продукция (правило формирования знаний геологоразведочного проекта на инвестиционном этапе) $R_z^u; D^u \cup Z^u \exists F_j \cup O_z \rightarrow Z_j^u, j = j+1, 1 \leq j \leq l;$

R_z^o – продукция (правило формирования знаний геологоразведочного проекта на эксплуатационном этапе) $R_z^o; D^o \cup Z^o \exists E_j \cup O_z \rightarrow Z_j^o, j = j+1, 1 \leq j \leq l$, где O_z – процедура формирования знаний геологоразведочного проекта;

$P_m = \langle Z, M, R_m, O_m \rangle$, где Z – фрейм: знания геологоразведочного проекта, M – фрейм: модель геологоразведочного проекта, R_m – множество продукций построения модели геологоразведочного проекта с помощью знаний:

$$Z : R_m; Z \cup M_j \exists Z_j \cup O_m \rightarrow m_j, j = j+1, 1 \leq j \leq l,$$

где O_m – процедура формирования модели геологоразведочного проекта.

$$P_o = \langle M, Q, R_o, O_o \rangle,$$

где M – фрейм: модель геологоразведочного проекта, Q – фрейм: последствия геологоразведочного проекта, R_o – множество продукций комплексной оценки модели геологоразведочного проекта;

$$R_o = \{R_o^{oz}, R_o^{oz}, R_o^{kz}\},$$

где R_o^{oz} – продукция ⟨правило формирования показателя общественной значимости геологоразведочного проекта⟩;

$$R_o^{oz}; M_j \cup Q \exists m_j \cup O_o^{oz} \rightarrow q_j^{oz} j = j+1, 1 \leq j \leq l,$$

где R_o^{oz} – продукция ⟨правило формирования показателя общественной эффективности геологоразведочного проекта⟩;

$$R_o^{oz}; M_j \cup Q \exists m_j \cup O_o^{oz} \rightarrow q_j^{oz} j = j+1, 1 \leq j \leq l,$$

где R_o^{kz} – продукция ⟨правило формирования показателя коммерческой эффективности геологоразведочного проекта⟩;

$$R_o^{kz}; M_j \cup Q \exists m_j \cup O_o^{kz} \rightarrow q_j^{kz} j = j+1, 1 \leq j \leq l, \text{ где } O_o = (O_o^{oz}, O_o^{oz}, O_o^{kz}),$$

где O_o^{oz} – процедура формирования общественной значимости геологоразведочного проекта; O_o^{oz} – процедура формирования параметров общественной эффективности геологоразведочного проекта; O_o^{kz} – процедура формирования параметров коммерческой эффективности геологоразведочного проекта.

$$P_f = \langle Q, N, R_f, O_f \rangle,$$

где Q – фрейм: последствия ИП, N – фрейм: схема финансирования геологоразведочного проекта, R_f – продукция формирования схемы финансирования геологоразведочного проекта:

$$R_f; Q \cup N \exists Q_j \cup O_f \rightarrow N; j = j+1, 1 \leq j \leq l,$$

где O_f – процедура формирования схемы финансирования геологоразведочного проекта.

Данный формальный аппарат является методической основой формирования базы знаний информационной системы анализа геологоразведочных проектов. Для программной реализации алгоритмов обработки знаний были разработаны алгоритмы индуктивного вывода на основе двухэтапной процедуры структурной таксономии: на первом этапе осуществляется предварительная классификация в декартовом пространстве; на втором этапе – структурная классификация в компактном пространстве.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Тышковская Ю.В. Информационное обеспечение систем управления геологоразведочными предприятиями / Ю.В. Тышковская // Современное общество: актуальные проблемы и перспективы. Всерос. науч.-практ. конф., апрель 2009 г. – Волгоград – М.: Глобус, 2009. – 189-192 с
2. Батищев В.И. Аппроксимационные методы и системы промышленных измерений, контроля, испытаний, диагностики / В.И. Батищев, В.С. Мелентьев. – М.: Машиностроение-1, 2007. – 393 с.
3. Батищев В.И. Методология оперативной реструктуризации информационных систем анализа состояния сложных технических объектов / В.И. Батищев, Н.Г. Губанов // Проблемы управления и моделирования в сложных системах: Тр. IX Международ. конф. – Самара: СНЦ РАН, 2008. – С. 187-193.
4. Батищев В.И. Категорное представление сложных технических объектов в индуктивных системах логического вывода / В.И. Батищев, Н.Г. Губанов // Проблемы управления и моделирования в сложных системах: Тр. IX Международ. конф. – Самара: СНЦ РАН, 2008. – С. 185-191.

5. Загоруйко Н.Г. Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. – Новосибирск: Изд-во Ин-та математики, 1999. – 270 с.
6. Губанов Н.Г. Категорный подход при формировании полимодельных комплексов сложных систем [Текст] / Н.Г. Губанов // Вестник Самар. гос. техн. ун-та. Сер. Технические науки. 2008. Вып. 1 (21) С. 183 - 185.

Статья поступила в редакцию 24 февраля 2011 г.

UDK 519.816+519.876.5

PRODUCTIONAL MODELS OF DATA AND KNOWLEDGE FORMALIZING IN IT-SYSTEMS OF GEOLOGICAL RECON PROJECTS ANALYSIS

V.I. Batishchev, N.G. Gubanov, U.V. Tishkovskaya

Samara State Technical University
244, Molodogvardeyskaya st., Samara, 443100

Problems of data representation in the information analysis systems are evaluated. The formal scheme of geological recon knowledge based is presented by means of frame and production models synthesis.

Keywords: exploration projects, production system, structural taxonomy.

*Vitaliy I. Batishchev – Doctor of Technical Sciences, Professor.
Nikolay G. Gubanov – Candidate of Technical Sciences, Associate professor.
Ulia V. Tishkovskaya – Senior Lecture.*