

Программный комплекс для интеллектуального анализа данных в горном производстве

Р.Ю. Замараев, С.Е. Попов

Аннотация. Описываются принципы создания и реализация распределенного программного комплекса (РПК) для интеллектуального анализа показателей функционирования сложных систем горного производства. Для решения задач многокритериального выбора и системного аудита в широком наборе показателей различной природы и размерности применяется энтропийный метод. Приводится пример использования комплекса для решения задач ранжирования по агрегированным характеристикам участков угледобычи Кузнецкого угольного бассейна на основе базы данных показателей углепроб.

Ключевые слова: распределенный программный комплекс, анализ данных, энтропийный метод, многокритериальный выбор, системный аудит, функциональные показатели, сложные системы, горное производство.

Введение

Совокупность данных, характеризующих состояние объектов горного производства, отличается значимой неоднородностью и паллиативностью. При ранжировании объектов используются экономические и технические показатели самих объектов, внешние горно-геологические условия, петрографические показатели полезных ископаемых, географические особенности расположения в контексте экологической безопасности, социальной значимости и др.

В таком же широком наборе ищутся и исследуются декомпозиции сложных систем – корпоративных, территориальных и технологических объединений. При этом многие показатели, значимые для принятых в горном производстве форм отчетности, являются связанными коэффициентами или грубо ранжированными характеристиками.

Дополнительные трудности для анализа создает существенная изменчивость показателей функционирования, которая определяется характерной для горного производства многостадийностью технологических процессов. Фик-

сируемые в произвольные моменты времени значения объема производства, производительности или горно-геологическая ситуация могут сильно отличаться от среднегодовых или среднемесячных значений.

Стандартные пакеты прикладных программ предлагают различного методического уровня инструменты для обработки данных, бухгалтерского и экономического анализа, DataMining. С их помощью удастся решать многие важные задачи, но при наличии устоявшихся оценок для типичных объектов и ситуаций в стационарных условиях.

В быстро меняющихся условиях при существенной неоднородности и разнообразии свойств объектов интерес и практическую значимость, в конечном счете, приобретают заключения с использованием инвариантных критериев об отличительных особенностях выборочного объекта на фоне функционально подобных ему. Такой подход к анализу данных был сформулирован как «диагностика состояния уникальных объектов» и реализован в энтропийном методе [1-5].

Однако эргономика применения нового метода оказалась довольно сложной. Основные инструменты для построения агрегированных моделей и отображений данных требуют четкой логики и понимания смысла промежуточных процедур, модулей системного профиля (в контексте РПК). При большом количестве агрегируемых показателей для результирующих отображений отсутствуют, как правило, прямые аналоги и интерпретации из предметных областей, а ресурсоемкость расчетов множественных системных профилей не позволяет своевременно и оперативно решать поставленные задачи.

Зачастую анализируемые данные, получаемые из информационной инфраструктуры горного предприятия, являются конфиденциальными, вследствие чего прямого доступа к ним экспертные группы научного учреждения не имеют, усложнение ситуации происходит по причине разнородности систем хранения данных и их форматов.

Поэтому возникает необходимость разработки программного комплекса, в котором были бы реализованы основные и вспомогательные алгоритмы энтропийного метода; функциональность удаленного доступа к разнообразным типам данным; имелись бы средства контроля логики и синтаксиса расчетных схем; имелась бы возможность интерактивной отладки расчетных схем и интерпретации результатов в виртуальных экспертных сообщениях через сеть Internet.

1. Энтропийный метод анализа данных

Рассматривается вектор $\mathbf{x} = \{x_i\}_{i=1, \dots, m}$, образованный положительными значениями некоторого показателя X для совокупности R мощностью m функционально подобных автономных объектов.

Функциональное подобие здесь понимается как тождественность значимости, размерности и направленности избранных для анализа показателей для всех объектов совокупности. Автономность подразумевает априорную независимость формирования показателей объектов,

отсутствие среди избранных показателей очевидного и доказанного аргумента.

Для перехода к вектору безразмерных величин в унифицированном масштабе используется преобразование

$$\mathbf{q} = \{\ln x_i^P\}, \quad i = 1, \dots, m, \quad (1)$$

где $P = 1 / \ln \prod_i x_i$.

Учитывая внутренние ограничения энтропийного преобразования исходных данных, использующего логарифмирование, для значений элементов вектора $\mathbf{x}$ может потребоваться масштабирование с тем, чтобы $\min(\mathbf{x}) \geq e$.

Важнейшими свойствами этого преобразования являются:

- аддитивность элементов вектора $\mathbf{q}$, т.к. для любого подмножества W рассматриваемой совокупности R правило вычисления суммы имеет вид

$$\sum_{i \in W} q_i = \ln \prod_{i \in W} x_i^P; \quad (2)$$

- аддитивность набора векторов $\mathbf{q}_1, \dots, \mathbf{q}_n$, т.к. правило вычисления суммы имеет вид

$$\sum_{j=1}^n \mathbf{q}_j = \left\{ \ln \prod_j x_{i,j}^{P_j} \right\}, \quad i = 1, \dots, m, \quad j = 1, \dots, n. \quad (3)$$

В обоих случаях как для суммы элементов вектора, так и для суммы векторов получаем мультипликативные функции.

Значение i -го элемента вектора $\mathbf{q}$ численно равно доле i -го объекта в величине производственной функции $\ln \prod_i x_i$ совокупности R по

показателю x . отождествление значения i -го элемента вектора $\mathbf{q}$ с весом или вероятностью выбора i -го объекта по показателю X из совокупности позволяет перейти к отображению энтропийного типа

$$\mathbf{E} = -\mathbf{q} \otimes \ln \mathbf{q} = \{-q_i \ln q_i\}. \quad (4)$$

Формула (4) подобна определению информационной энтропии по К.Шеннону, что и определило название метода. Вектор $\mathbf{E}$ является моделью генерации информации (информативности) системы i -ыми объектам по показателю

X . Сумма элементов вектора $\mathbf{E}$ является количеством информации в системе по показателю X . Элементы E_i вектора являются мерами неопределенности выбора i -го объекта: чем больше E_i , тем больший вес имеет i -ый объект в совокупности и меньше неопределенность (риск) его выбора.

Для вектора $\mathbf{q}$ можно также ввести логарифмическое отображение, производное от энтропийного, и получать, соответственно, с точностью до константы (-1) производные модели информативности

$$\mathbf{L} = \frac{d\mathbf{E}}{d\mathbf{q}} = -\ln \mathbf{q} = \{-\ln q_i\}. \quad (5)$$

Матрица, составленная несколькими векторами $\mathbf{E}$, является матрицей информативности системы, заданной выбранными для анализа объектами и их показателями. Преобразование исходных данных к единой информационной мере позволяет в такой матрице формировать произвольные модели информативности системы – по показателям и по объектам, частные и обобщенные. Ограничения здесь накладывают только здравый смысл и интерпретируемость конечной модели.

Например, для решения задач ранжирования, многокритериального выбора и декомпозиции системы строятся комбинированные модели показателей с учетом их направленности:

$$\begin{aligned} \hat{\mathbf{E}} &= \sum_{j \in U} a_j \mathbf{E}_j - \sum_{k \in V} b_k \mathbf{E}_k = \left\{ \sum_{j \in U} a_j E_{i,j} - \sum_{k \in V} b_k E_{i,k} \right\}, \quad (6) \\ \hat{\mathbf{L}} &= \sum_{j \in U} a_j \mathbf{L}_j - \sum_{k \in V} b_k \mathbf{L}_k = \left\{ \sum_{j \in U} a_j L_{i,j} - \sum_{k \in V} b_k L_{i,k} \right\}, \end{aligned}$$

где U и V – подмножества положительно и отрицательно направленных показателей соответственно; a_j и b_k – коэффициенты значимости показателей. В системном аудите анализируется информативность моделей, упорядоченных по индексу объекта, а также их комбинации. Примеры представления таких моделей приведены в разделе 4.

Для эволюционирующих систем, представленных показателями одного объекта на различных этапах развития, открывается возмож-

ность анализировать их как системы самостоятельных объектов с индивидуальными функциональными состояниями.

Поскольку разброс частных информативностей объектов и, в особенности, их производных может быть очень велик, то для моделей используется единый масштаб долей среднеквадратического отклонения

$$\overset{\circ}{\mathbf{E}} = \frac{\hat{\mathbf{E}} - M[\hat{\mathbf{E}}]}{\sigma[\hat{\mathbf{E}}]}, \quad \overset{\circ}{\mathbf{L}} = \frac{\hat{\mathbf{L}} - M[\hat{\mathbf{L}}]}{\sigma[\hat{\mathbf{L}}]}, \quad (7)$$

где $M[\dots]$ и $\sigma[\dots]$ – операторы среднего значения и среднеквадратического отклонения элементов вектора соответственно.

Ключевая методическая новизна энтропийного метода заключается в отказе от традиционного построения регрессий и симплексов при достигнутой аддитивности операндов. Для анализа состояния и ранжирования объектов используются свойства их изображающих точек с координатами $(\hat{E}_i, \hat{L}_i)$, полагая, что пара векторов $\hat{\mathbf{E}}$ и $\hat{\mathbf{L}}$ определяет аналог фазовой плоскости $\hat{\mathbf{L}}(\hat{\mathbf{E}})$ дискретно заданной системы. При этом оказываются доступными для использования строгие границы и критерии видов состояний систем и их элементов на фазовой плоскости, известные из динамики (теории колебаний).

Для разделения фазовой плоскости на области используются следующие границы:

- главная ось совокупности изображающих точек как предельная линия прямой связи фазовых координат, разделяющая элементы системы по критерию консервативности

$$r = \frac{1}{m} \sum_i \overset{\circ}{E}_i \overset{\circ}{L}_i, \quad (8)$$

где r – тангенс угла наклона главной оси, проходящей через начало координат фазовой плоскости $\overset{\circ}{\mathbf{L}}(\overset{\circ}{\mathbf{E}})$;

- эллипс как предельная линия гармонического решения уравнения связи фазовых координат, разделяющая элементы по критерию устойчивости

$$\frac{u^2}{A^2} + \frac{v^2}{B^2} = 1, \quad (9)$$

$$A = \sqrt{2 \left(\frac{m-1}{m} + \frac{2r^2}{1+r^2} \right)}, \quad B = \sqrt{2 \left(\frac{m-1}{m} - \frac{2r^2}{1+r^2} \right)},$$

где u и v – фазовые координаты, A и B – большая и малая полуоси эллипса;

- гиперболы с идентичными полуосями

$$\frac{u^2}{A^2} - \frac{v^2}{B^2} = \pm 1, \quad (10)$$

что задает одинаковую жесткость двух модельных систем.

Интерпретация вводимых границ определяется показателями, включенными в определение фазовых осей. Неизменным остается качественное разделение фазовой плоскости на области с различными видами состояния и тенденциями поведения объектов. Параметры границ не зависят от предметного содержания показателей, вовлеченных в построение фазовых осей, следовательно, границы являются фундаментальными и универсальными.

2. Постановка задачи

Процесс разработки комплекса рассмотрим в контексте расширения области применения и развития правил интерпретаций моделей энтропийного анализа, унификации и описания аналитических схем метода, использования ресурсоемких вычислительных алгоритмов метода на примере задач ранжирования по агрегированным характеристикам участков угледобычи Кузнецкого угольного бассейна.

Этапы разработки комплекса:

1. разработать единый формат системного профиля на базе языка XML, полностью описывающего логику расчета модели, обеспечивающего независимость программного кода от количества, типа и природы расчетных данных, возможность индивидуальной настройки графического интерфейса пользователя и полного контроля над его функционалом;

2. создать консоль управления расчетными процессами, пользователями и системными профилями РПК;

3. разработать конструктор, позволяющий создавать, редактировать и производить вери-

фикацию системных профилей, их хранение и извлечение из СУБД;

Упрощенная модель потоков данных в процессе функционирования представлена на Рис. 1. На ее основе будем строить логическую модель и программную реализацию системы.

3. Распределенный программный комплекс «Энтропия Плюс»

3.1. Краткая характеристика компонентов системы

Основными функциональными свойствами программного комплекса были заявлены: распределенное n-уровневое исполнение; возможность интерактивного изменения и применения аналитических схем к непрерывно пополняемым данным; возможность получения решений и графических образов, синхронизированных в плоскости клиент-сервер; отслеживание состояния пользовательских задач; возможность работы в сетях Internet и Intranet.

При реализации вся необходимая функциональность была разделена между следующими модулями: административным, серверным, клиентским, экспертным и конструктором расчетных схем.

Центральным модулем РПК является серверный. На нем производится подготовка и обработка данных в соответствии с аналитической схемой, активируемой по запросу из списка доступных согласно профилю пользователя. Архитектура серверного модуля и правила составления аналитических схем позволяют быстро дополнять РПК новыми процедурами и функциями. Оформлен этот модуль как набор динамических библиотек, содержащих маршалируемые объекты, которые обеспечивают загрузку исходных данных, расчет схем пользователя согласно профилю, поддержку и исполнение управляющих сообщений (Рис. 2).

Экспертный модуль содержит опробованные экспертные заключения о состоянии объектов, подсистем и системы в целом для известных типов задач и наборов данных. И, если это предусмотрено расчетной схемой, сопровождается заключениями визуализацию обработанных данных на стороне клиента. Оформлен модуль как пополняемый набор динамических библиотек решающих правил.

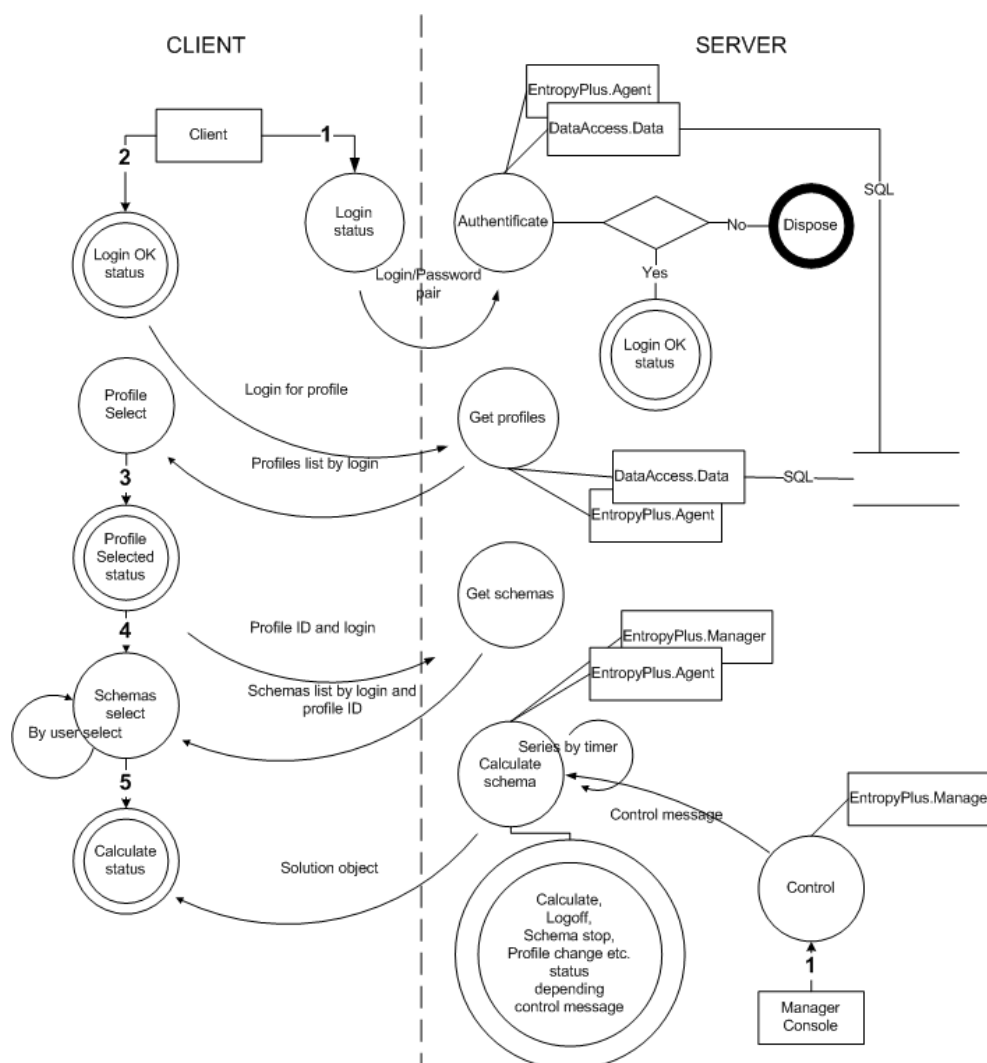


Рис. 1. Упрощенная диаграмма поток данных РПК

Superfetch	Поддерж...	Отключена
VCL.CalculationCluster.EntropyPlus.Agent.Server	Работает	Авто
VCL.DataCluster.DataAccess.Server	Работает	Авто
VCL.ManageCluster.EntropyPlus.Agent.Server	Работает	Авто
VCL.ManageCluster.EntropyPlusManager.Server	Работает	Авто
Windows Audio	Управлен...	Работает Авто

Рис. 2. Перечень системных сервисов серверного модуля

Административный модуль (Рис. 3) создан как графическое приложение (GUI), функционал которого обеспечивает контроль и управление пользователями, соединениями системы, системными профилями, подключениями внешних источников данных и данными во временных локальных таблицах.

Разработка и отладка расчетных схем ведутся через модуль конструктора (Рис. 4). Это гра-

фическое приложение имеет инструменты интерактивного конструирования и выполнения аналитических схем с указанием источников данных и способов визуализации на клиентском модуле. В нем же ведется заполнение шаблонов экспертного модуля.

Конструктор рассматривается как инструмент эксперта-аналитика для интерактивной разработки, отладки и интерпретации расчет-

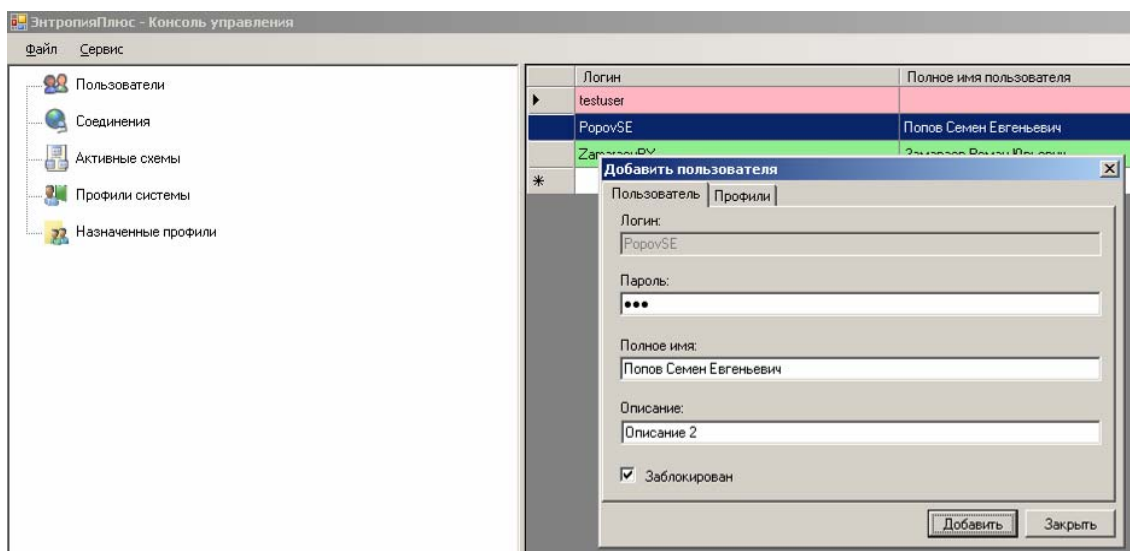


Рис. 3. Консоль управления административного модуля (фрагмент)

ных схем. Обновление существующей или внесение новой расчетной схемы в системный профиль фиксируется в базе системных профилей и специальным системным сообщением вызывает исключение на стороне пользователя. При перезапуске клиентского модуля пользователю открывается обновленный перечень доступных расчетных схем и соответствующих графических представлений.

Клиентский модуль является графическим приложением (GUI), в котором по выбору пользователя активизируются аналитические схемы из списка доступных для профиля пользователя, визуализируются обработанные данные и экспертные заключения. Клиентское приложение отображает фазовые плоскости, линейные графики, диаграммы, служебную информацию, позволяет задавать расчетные диапазоны в матрице, «проигрывать» графические представления в заданном диапазоне данных, эмулируя развитие системы (Рис. 5).

3.2. Функционал программного комплекса

Программный комплекс разработан на базе технологии .NET Remoting (Microsoft), что позволяет размещать модули в различных узлах сети Internet и Intranet и обеспечивать совместную работу модулей и удаленных объектов через межсетевые экраны [6].

Для унификации записи расчетной схемы и удобного контроля ее логики была предложена

модель метаописания алгоритмов энтропийного анализа на базе языка XML (таблица). Соответственно каждый пользовательский профиль структурируется и формализуется заданными тэгами-описателями на базе DTD-модели (Рис. 6,а).

Профиль содержит описательную часть расчета (название профиля, схем и параметров). Расчетная матрица обозначена как элемент **<matrix>**, включает элементы **<column>** (показатели), строку соединения с БД и дополнительные параметры расчета, необходимые для различной функциональности клиента. Каждый элемент **<module>** формально описывает процедуру преобразования данных, содержит атрибуты **id** и **parentid**, соответственно, номер преобразования в последовательности и ссылку на результат предыдущего преобразования.

Атрибут **parentid** может содержать список номеров, идущих не обязательно по порядку, что означает применение комбинации результатов предыдущих процедур. Схема разбита на два потока **<streamX>** и **<streamY>**, описывающих элементарные преобразования показателей по осям фазовой плоскости. Схема содержит системную информацию: интервал обновления расчета, применение процедур подготовки данных, способ именования конечного результата в соответствии с последовательностью преобразований (Рис. 6,б).

Экспертные модули могут «вклиниваться» между любыми двумя узлами **<module>**, обра-

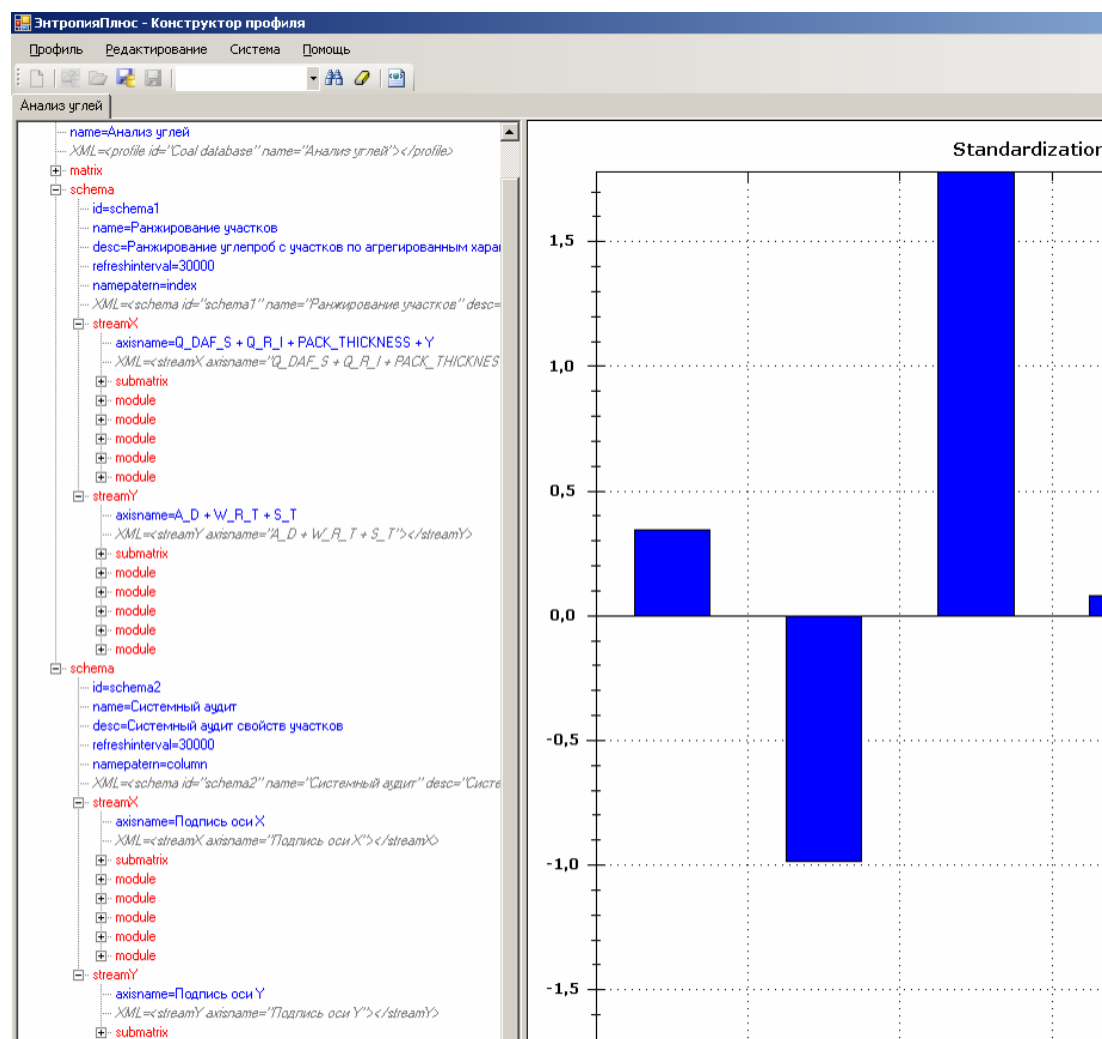


Рис. 4. Интерфейс конструктора системных профилей (фрагмент)

батьвать промежуточную матрицу и добавлять текстовое сообщение в переменную **ExpertConclusion**, аккумулируя простейшие экспертные заключения (Рис. 5).

Таким образом, серверный модуль оперирует системными профилями по запросу клиентского, создавая экземпляры объектов-вычислителей по количеству активируемых схем, «прогоняя» начальную матрицу «объект-параметр» по дереву, видоизменяя ее к конечному результату – структуре типа **<X,Y,Objects,ExpertConclusion>**.

В конструкторе пользовательский профиль формируется в виде «дерева» (Рис. 4), «движение» по которому задает действия серверного модуля для получения конечного графического представления, что сводит построение профиля

к конструированию иерархии объектов-тэгов согласно DTD-модели. Отдельные процедуры конструктора отслеживают изменения структуры «дерева» и контролируют правила вложенности процедур и целостность схемы.

На каждом этапе наращивания дерева процедур имеется возможность посмотреть результат того или иного элементарного математического преобразования в виде таблицы данных или графического образа для визуальной оценки логики работы расчетной схемы.

Таким образом, разработанная технология расчетных схем и системных профилей обеспечивает единые и прозрачные правила работы серверного и клиентского модулей и снимает проблемы контроля логики анализа и целостности профиля при внесении изменений.

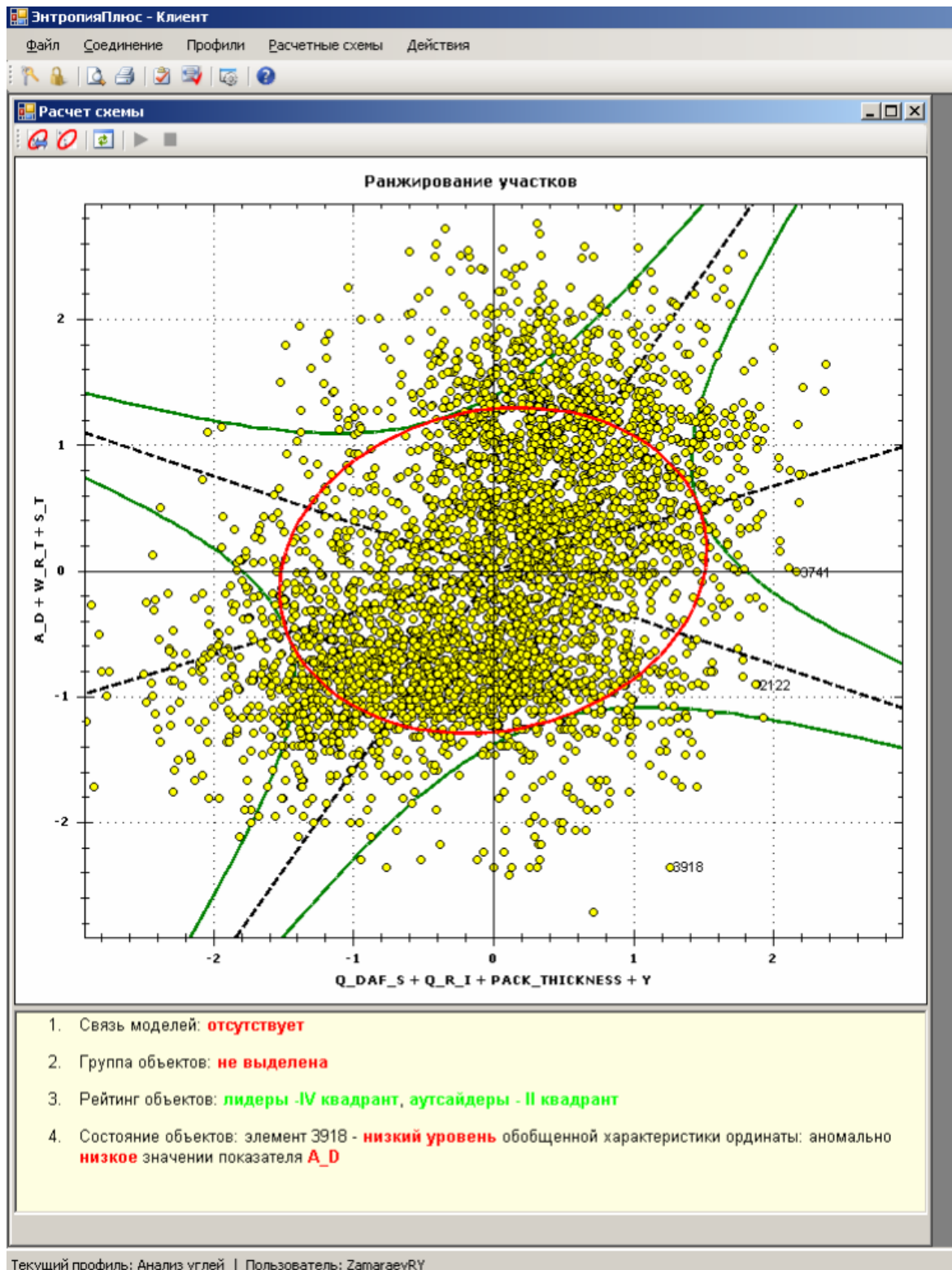


Рис. 5. Интерфейс клиентского компонента (фрагмент)

```

<!DOCTYPE profile [
<!ELEMENT profile (matrix, schema+)>
<!ATTLIST profile
    id ID #REQUIRED
    name CDATA #REQUIRED>
<!ELEMENT matrix (column+)>
<!ATTLIST matrix
    refreshinterval CDATA #REQUIRED
    connectionstring CDATA #REQUIRED
    remotingtable CDATA #REQUIRED
    startdatabound CDATA #REQUIRED
    step CDATA #REQUIRED>
<!ELEMENT submatrix (column+)>
<!ATTLIST submatrix
    startdatabound CDATA #REQUIRED
    step CDATA #REQUIRED>
<!ELEMENT column EMPTY>
<!ATTLIST column
    id CDATA #REQUIRED
    name CDATA #IMPLIED
    isObject CDATA #IMPLIED>
<!ELEMENT schema (streamX, streamY)>
<!ATTLIST schema
    id ID #REQUIRED
    name CDATA #REQUIRED
    desc CDATA #IMPLIED
    refreshinterval CDATA #REQUIRED
    namepatern CDATA #REQUIRED>
<!ELEMENT streamX (submatrix,module*)>
<!ATTLIST streamX
    isNull CDATA #REQUIRED
    axisname CDATA #IMPLIED>
<!ELEMENT module EMPTY>
<!ATTLIST module
    id CDATA #REQUIRED
    mask CDATA #IMPLIED
    direction CDATA #IMPLIED
    parameterindex CDATA #IMPLIED
    parentid CDATA #IMPLIED
    step CDATA #IMPLIED
    name CDATA #REQUIRED>
<!ELEMENT streamY (submatrix,module*)>
<!ATTLIST streamY
    isNull CDATA #REQUIRED
    axisname CDATA #IMPLIED>
-

```

а) DTD-модель профиля

```

<profile id="Profile1" name="Общая модель информ. системы (сейсмика)">
  <matrix refreshinterval="30000" connectionstring="
    Provider=OraOLEDB.Oracle.1;
    User Id=TEST_USR;
    Password=password1;
    Data Source=(DESCRIPTION=(CID=GTU_APP) (ADDRESS_LIST=(ADDRESS=
      (PROTOCOL=TCP) (HOST=localhost) (PORT=1521)))
      (CONNECT_DATA=(SID=ORCL) (SERVER=DEDICATED))) );
    DistribTX=0;Persist Security Info=True;"
    remotingtable="MATRIX2" startdatabound="12" step="300">
    <column id="C1" name="C1" isObject="False" />
    <column id="C2" name="C2" isObject="False" />
    <column id="C3" name="C3" isObject="False" />
    <column id="C4" name="C4" isObject="False" />
  </matrix>
  <schema id="schema1" name="Сейсмические события (сигнал)"
    desc="Краткое описание схемы"
    refreshinterval="2000" namepattern="row">
    <streamX axisname="Номер отсчета">
      <submatrix startdatabound="" step="">
        <column id="C1" name="C1" isObject="False" />
        <column id="C2" name="C2" isObject="False" />
        <column id="C3" name="C3" isObject="False" />
        <column id="C4" name="C4" isObject="False" />
      </submatrix>
      <module id="1" name="Shift" parentid="0" />
      <module id="2" name="Weighting" parentid="1" direction="column" />
      <module id="3" name="Entropy" parentid="2" />
      <module id="4" name="Accumulate" parentid="3" />
      <module id="5" name="Standardization" parentid="4" direction="row" />
    </streamX>
  </schema>
</profile>

```

б) фрагмент XML-описания профиля пользователя согласно DTD-модели

Рис. 6.

Соответствия элементарных преобразований метода энтропийного анализа
элементам системного профиля

Элементарные преобразования		XML вариант
Название	Выражение	
Взвешивание	${}^1R = \frac{R_{ij}}{\sum_{i=1}^m \sum_{j=1}^n R_{ij}}$	<code><module id="[№]" parentid="[№]([№],[№]...)" name="Weighting" direction="matrix column row"/></code>
Энтропия	$E_{ij} = -\frac{R_{ij}}{\sum_{j=1}^n R_{ij}} \ln\left(\frac{R_{ij}}{\sum_{j=1}^n R_{ij}}\right)$	<code><module id="[№]" parentid="[№]([№],[№]...)" name="Entropy"/></code>
Стандартизация	${}^sR = \frac{R_{ij} - M}{\sigma}$	<code><module id="[№]" parentid="[№]([№],[№]...)" name="Standardization" direction="matrix column row"/></code>

В таблице представлена часть методов. Общее количество элементарных преобразований соответствует 16 элементам **<module>** системного профиля.

Система управления экземплярами удаленных объектов построена по технологии Remoting Events и позволяет обмениваться сообщениями между экземплярами классов, подписавшихся на удаленные события маршалируемого объекта.

Любой из объектов может генерировать событие, которое перехватывают и обрабатывают все остальные объекты – его потомки, и производить делегирование удаленного события за границы домена приложений. Клиентские приложения работают каждое со своим экземпляром серверного модуля, следовательно, управление полностью разграничено.

В качестве формата управляющего сообщения был выбран язык XML, а в серверный модуль внедрен соответствующий обработчик. Структурированный и вложенный формат записи позволяет в одном сообщении инкапсулировать управление сразу несколькими объек-

тами (Рис. 7), что снижает нагрузку на аппаратные ресурсы физического сервера и его сетевой трафик.

Каждый элемент **<message>** содержит атрибуты **timemark** и **username**, по которым происходит дифференциация сообщений клонами серверного модуля. Клон выделяет свою часть управляющей команды, соответственно программному классу (элементы **<login>**, **<schema>**, **<profiledataupdater>**) и выполняет действие (атрибут **action**).

Автоматизированная часть административного модуля взаимодействует с серверным, получает и обрабатывает информацию о пользователях системы и их профилях, соединениях и запущенных расчетных схемах, соответствующих каждому пользователю в определенный момент времени. Для работы с информационно-эволюционирующими системами и попол-

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE messages [
  <!ELEMENT messages (message+)>
  <!ELEMENT message (profile? | profiledataupdater? | login? | schema*) >
  <!ATTLIST message
    timemark ID #REQUIRED
    username CDATA #REQUIRED>
  <!ELEMENT profile EMPTY>
  <!ATTLIST profile
    action CDATA #REQUIRED>
  <!ELEMENT profiledataupdater EMPTY>
  <!ATTLIST profiledataupdater
    action CDATA #REQUIRED>
  <!ELEMENT schema EMPTY>
  <!ATTLIST schema
    id CDATA #REQUIRED
    action CDATA #REQUIRED>
  <!ELEMENT login EMPTY>
  <!ATTLIST login
    action CDATA #REQUIRED>
]>
<messages>
  <message timemark="tm2009040213400020" username="user1">
    <login action="off"/>
  </message>
  <message timemark="tm200904021450090" username="ep_system1_usr2">
    <profiledataupdater action="stop"/>
  </message>
  <message timemark="tm200904021550080" username="ep_system1_usr1">
    <schema id="schema1" action="change"/>
  </message>
</messages>
```

Рис. 7. XML-описание управляющего сообщения

няемыми базами данных в DTD-модели профиля предусмотрены независимые параметры периодичности обновления **refreshinterval** (в миллисекундах) для схем и матрицы исходных данных. Согласно установленным значениям происходит автоматическое обращение к источнику данных для заполнения локальной таблицы в оперативной памяти и перерасчет схемы с отправкой обновленных результатов всем ассоциированным профилям и активным пользовательским модулям.

Главная функциональность консоли управления заключается в автоматической генерации и отправке управляющего сообщения экземпляру расчетного модуля на определенные действия пользователя (отключить пользователя от системы, остановить/запустить обновление данных и т.п.), а также обновление соответствующих им таблиц.

Маршалируемые объекты доступа к данным серверного модуля отвечают за взаимодействие с системной БД, состоящей из пяти таблиц **USERS**, **LOGINS**, **SCHEMAS**, **PROFILES**, **USER_TO_PROFILES** (Рис. 8). Они содержат, соответственно, имя/пароль пользователей и его описание, текущие соединения и активные схемы пользователя, расчетные профили, связку пользователь-профиль (один пользователь может иметь несколько независимых профилей).

При входе пользователя в систему в таблицах **LOGINS** и **SCHEMAS** создаются соответствующие записи согласно комбинации логин/пароль. При выходе пользователя соответствующие кортежи удаляются. Остальные таблицы заполняются в процессе конструирования профиля (таблица **PROFILES**) и при помощи консоли управления (таблицы **USERS** и **USERS_TO_PROFILES**).

Объекты серверного модуля имеют также функции доступа к внешним базам данных для формирования расчетной матрицы. В настоящий момент программный комплекс может работать с локальными и удаленными хранилищами ORACLE, MS SQL Server и MySQL. Это достигнуто путем применения стандартных классов ODBC архитектуры .NET Framework и объектов сторонних разработчиков. Правила извлечения данных хранятся в элементе профиля **<matrix>** в виде строки подключения. Фак-

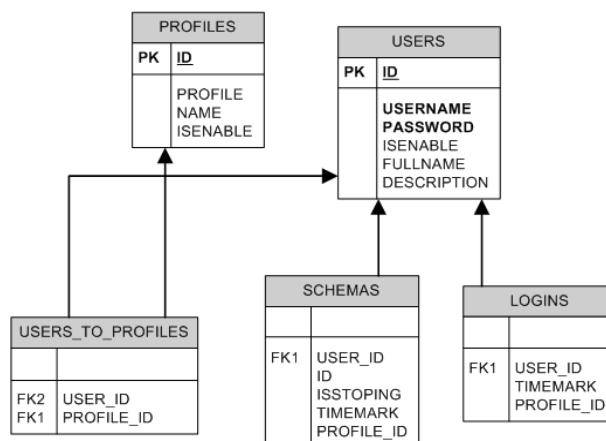


Рис. 8. ER-диаграмма системной БД РПК

тически удаленный объект серверного модуля принимает только этот параметр и выполняет необходимые действия.

Все экземпляры клиентского приложения работают автономно друг от друга и контролируются индивидуально средствами административного модуля. Вход в систему происходит путем комбинации логин/пароль, которая хранится в таблице **USERS**. Реализована схема «один пользователь – много профилей». После успешной идентификации пользователя запускается механизм настройки GUI клиентского модуля согласно профилю. Тем самым устанавливается жесткая связь между функциональностью клиентского приложения и XML-содержимым профиля.

Количество одновременно работающих клиентов ограничивается лишь аппаратными ресурсами ЭВМ, на которой размещен серверный модуль.

4. Пример использования РПК в задачах анализа данных

На Рис. 9 приведен фазовый портрет системы проб, демонстрирующий ранжирование по агрегированным характеристикам участков угледобычи Кузнецкого угольного бассейна.

В моделях фазовых абсциссы и ординаты как позитивные показатели для выбора были учтены высшая (Q_DAF_S) и низшая (Q_R_I) теплота сгорания угля, толщина угольных пачек (PACK_THICK) и толщина пластического слоя (Y). Как негативные показатели рассмат-

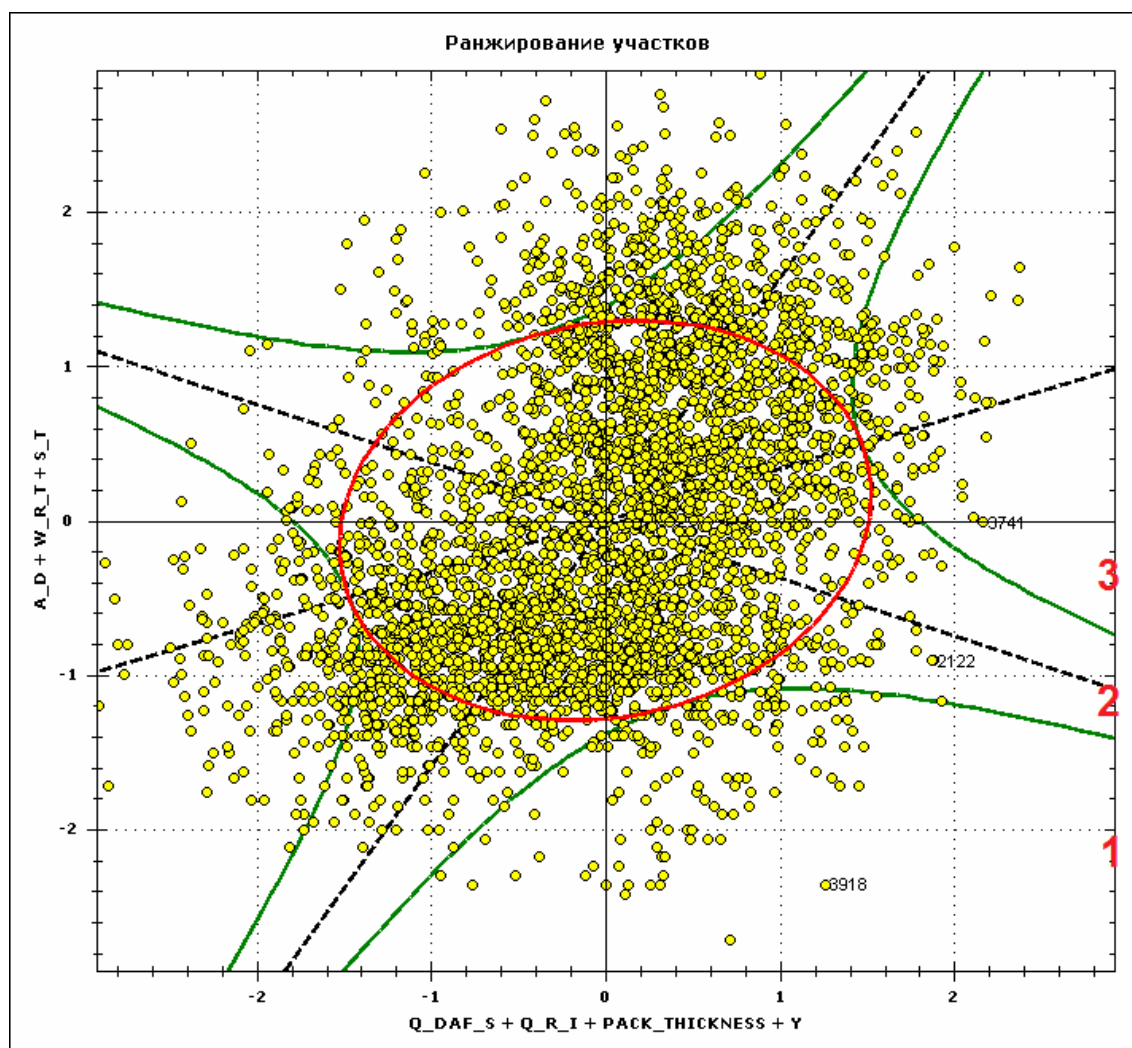


Рис. 9. Фазовый портрет системы проб угля с участков Кузнецкого угольного бассейна в координатах агрегированных характеристик положительных и негативных показателей

1,2,3 – номера зон

ривались зольность (A_D), содержание серы (S_T) и влажность (W_R_T). Систему составили более 4000 проб с активно разрабатываемых и новых, неосвоенных участков.

Предложенные агрегированные характеристики позволяют решить задачу выбора участка с гарантированным сбытом угля (энергетика и/или металлургия) при низких затратах на добычу и обогащение. Для упрощения интерпретации фазового портрета положительные и негативные показатели разнесены по фазовым осям – по абсциссе и ординате соответственно. Очевидно, интерес для выбора представляет 4-ый квадрант плоскости (Рис. 9), где сосредоточены изображающие

точки проб с высоким уровнем положительных показателей и низким – негативных.

Эллипс отсекает ординарные пробы (внутренняя область) от проб выдающихся. Гиперболические границы задают дополнительную дифференциацию плоскости:

- пробы с очень низким уровнем одного или нескольких негативных показателей и не-плохими положительными показателями – зона 1 между ординатой и ветвью нижней гиперболы;
- пробы с высоким уровнем одного или нескольких положительных показателей и достаточно низкими негативными показателями – зона 3 между абсциссой и ветвью правой гиперболы;

- пробы с перемежающимися по силе показателями – зона 2 между ветвями нижней и правой гипербол вдоль асимптоты.

Чтобы исследовать соотношения показателей проб-лидеров, рассмотрим модели информативности, упорядоченные по индексу объекта. Для отмеченных на Рис. 9 проб на Рис. 10, а-в приведены распределения информативности их свойств.

Из Рис. 10, а видно, что в группу лидеров участок с пробой №3918 попал за счет аномально малой зольности, в то время как позитивные показатели ничем выдающимся не отличаются. Отличительной чертой участка с пробой №2122 (Рис. 10, б) является высокий уровень толщины пластического слоя и мощности пласта, но при этом совершенно провально выглядят теплоты сгорания, и все это при достаточно низких негативных показателях. Участок с пробой №3741 (Рис. 10, в) показал свои лидирующие свойства только за счет аномально высокой высшей теплоты сгорания, на фоне

которой остальные позитивные и негативные показатели выглядят умеренно низкими.

Заключение

Распределенный программный комплекс "ЭнтропияПлюс" рассматривается как инструмент для получения новых знаний о закономерностях функционирования и развития систем горного производства. Предлагаемая разработка является реализацией новой технологии обработки информации средствами современных информационных технологий, что позволяет ставить и решать для горного производства принципиально новые задачи по анализу данных.

Универсальность энтропийного метода анализа по отношению к предметным областям, фундаментальность критериев оценки состояния систем и их элементов дают основания для переноса, в дальнейшем, расчетных схем и экспертных заключений на другие технические системы и отрасли с аналогичными по смыслу задачами и показателями.

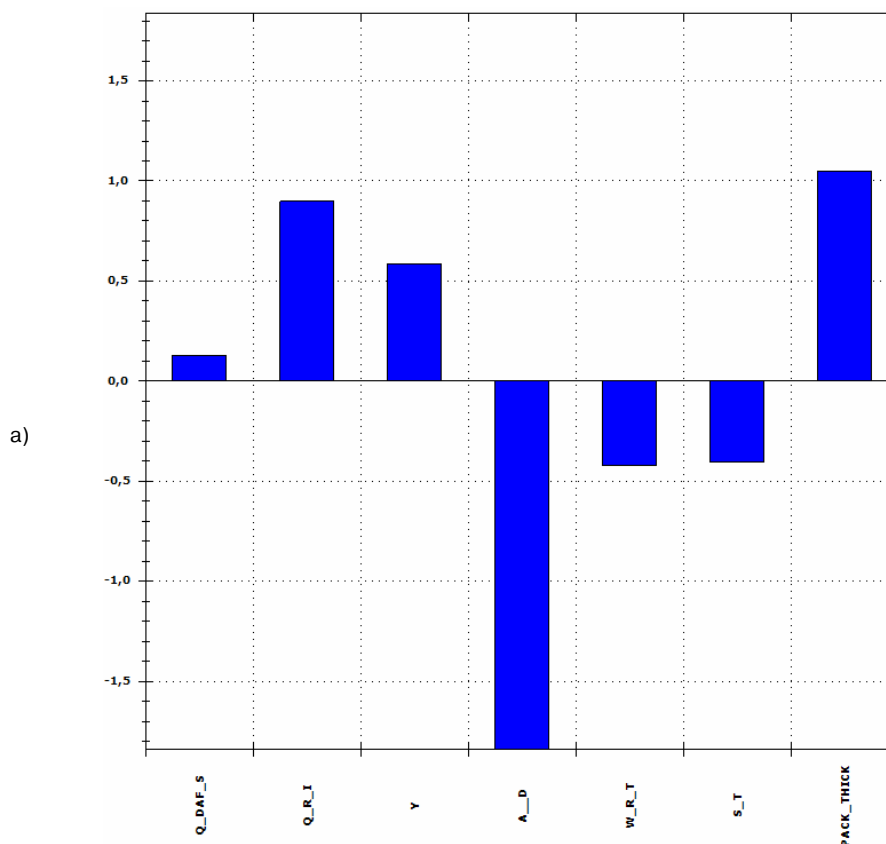


Рис. 10. Распределения информативности по показателям проб-лидеров по агрегированным характеристикам
а – проба №3918

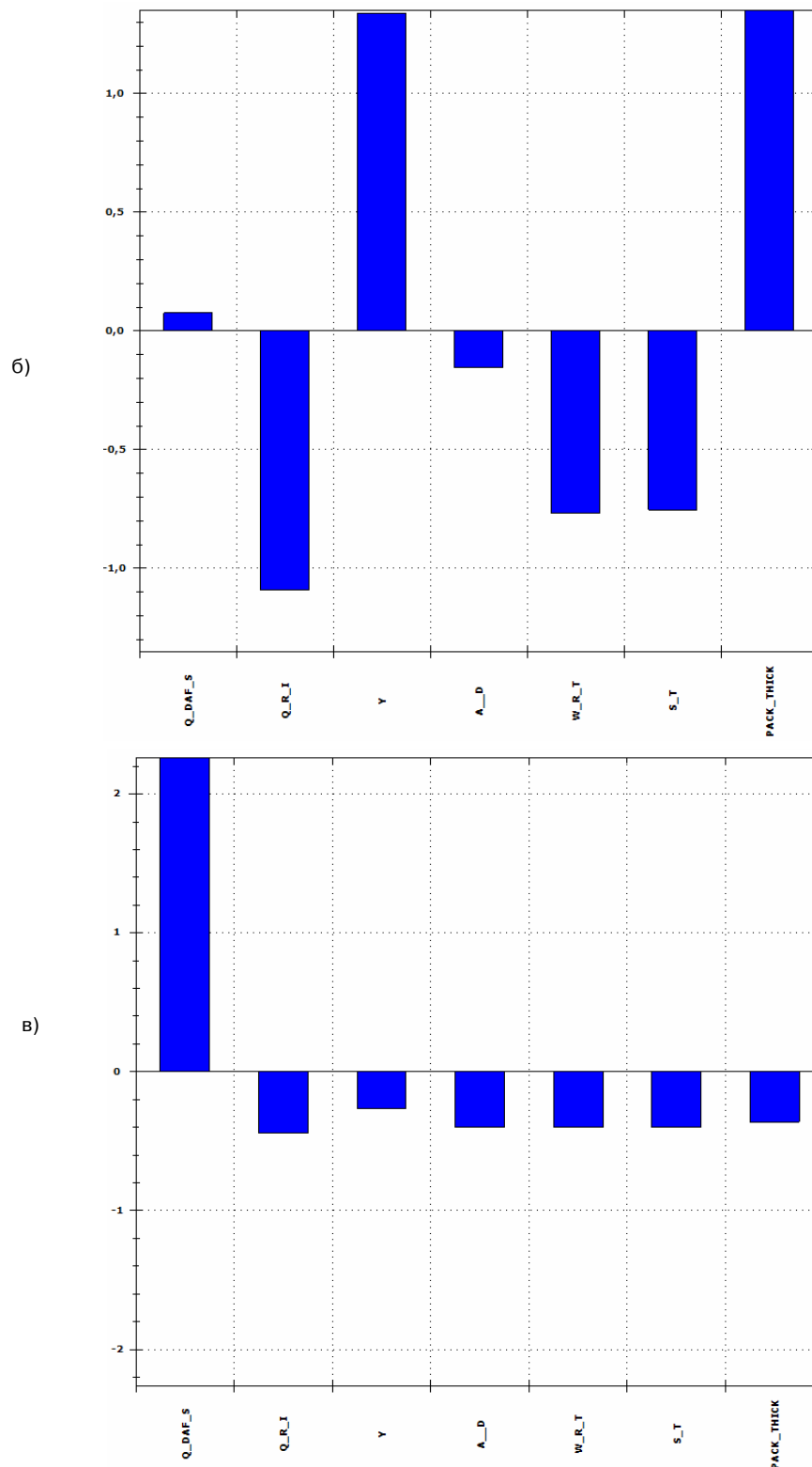


Рис. 10. Распределения информативности по показателям проб-лидеров по агрегированным характеристикам
б – №2122; в – №3741

Литература

1. Логов А.Б., Замараев Р.Ю., Логов А.А. Анализ состояния систем уникальных объектов / Вычислительные технологии, том 10, №5, 2005, – С.49-53.
2. Логов А.Б., Замараев Р.Ю., Логов А.А. Моделирование тенденций поведения элементов систем уникальных объектов / Вычислительные технологии, том 10, №5, 2005, – С.54-56.
3. Логов А.Б., Замараев Р.Ю., Логов А.А. Алгоритмы энтропийного метода анализа для отображения свойств объекта в фазовом пространстве / Вычислительные технологии, том 10, №6, 2005, – С.75-81.
4. Логов А.Б., Замараев Р.Ю., Логов А.А. Анализ функционального состояния промышленных объектов в фазовом пространстве. Институт угля и углехимии СО РАН, Кемерово: 2004 – 168 с. (8,9 уч. изд. л.)
5. Замараев Р. Ю., Опарин В.Н., Попов С.Е., Потапов В.П., Пястуневич О.Л. Использование ГИС-технологий для комплексной оценки участков отработки запасов полезных ископаемых / Физико-технические проблемы разработки полезных ископаемых. – 2008. – Т. 1, № 3. – С. 91-101.
6. Маклин С., Нафтел Дж., Уильямс К. Microsoft .NET Remoting / Пер. с англ. – М.: Торгово-издательский дом «Русская редакция», 2003. 384 с.

Замараев Роман Юрьевич. Старший научный сотрудник Института угля и углехимии Сибирского отделения РАН. Окончил Кузбасский государственный технический университет 1996 году. Кандидат технических наук. Автор 32 печатных работ, в том числе 3-х монографий. Область научных интересов: математическое моделирование технологических систем и процессов, анализ данных. E-mail: r.zamaraev@icc.kemsc.ru.

Попов Семен Евгеньевич. Старший научный сотрудник Института угля и углехимии Сибирского отделения РАН. Окончил Кемеровский государственный университет в 2002 году. Кандидат технических наук. Автор 12 печатных работ. Область научных интересов: математическое моделирование технологических систем и процессов, распределенные системы и GRID-технологии. E-mail: s.popov@icc.kemsc.ru.