

## ТЕХНОЛОГИЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЯ В ДИАГНОСТИКЕ ПРОФЕССИОНАЛЬНЫХ НЕЙРОТОКСИКОЗОВ<sup>1</sup>

А.Г. Иванов, М.П. Дьякович, С.В.Бахвалов

А.Г. Иванов

e-mail: iag2009@yandex.ru

М.П. Дьякович

e-mail: marik914@rambler.ru

*Ангарский филиал УРАМН ВСНЦ экологии человека СО РАМН – НИИ медицины труда и  
экологии человека*

С.В.Бахвалов

e-mail: bsv.istu@edu

*Национальный исследовательский Иркутский государственный технический университет*

### Аннотация

Рассмотрены основы информационной технологии поддержки принятия решений в диагностике профессиональных нейротоксикозов. Показана концепция, методический состав, модельная структура и порядок применения экспертной системы, реализующей такую технологию.

**Введение.** Хронические профессиональные нейротоксикозы (ПН) развиваются у работающих лиц вследствие производственных контактов с нейротропными веществами, депонирующимися в органах и системах организма, и вызывающими нарушения здоровья. У работников современных химических производств, клинические формы ПН проявляются при стаже работы с токсикантом свыше 15 лет, и завершаются дегенеративными явлениями в обменных процессах, сердечно-сосудистой и пищеварительной системах через 6-7 лет болезни. К этому времени патологические изменения нервной системы приобретают стойкий характер [1,2] с формированием психоорганического синдрома [3] со значительными личностными нарушениями, свидетельствующими о социально-психической дезадаптации [4], что обуславливают трудность реабилитации таких больных. Хронические нейротоксикозы, составляющие небольшую долю в структуре профессиональной заболеваемости РФ, имеют существенную социальную значимость, так как приводят к тяжелой инвалидизации лиц трудоспособного возраста. Ранняя идентификация признаков ПН и прогноз развития заболевания являются актуальными задачами профпатологической практики. Решение этих задач затруднено в связи с отсутствием патогномичных признаков в клинической картине заболевания, слабоспецифичной для различных нейротоксикантов и мало отличающейся от проявлений распространенных стрессовых, дезадаптационных, невротических нарушений и астенического синдрома, сопровождающего многие соматические заболевания и органические поражения мозга различного генеза [2,5]. На причинность заболевания может влиять ряд социально-психологических факторов, а также употребление работником

---

<sup>1</sup> Работа выполнена при поддержке гранта РГНФ № 10-06-12121в

бытовых нейротоксикантов, таких, как алкоголь, что существенно осложняет возможность установления профессиональной этиологии болезни. Клиницисту необходимо не только определить нозологическую форму, при которой имеют место непротиворечивые отношения между множеством наблюдаемых патологических признаков и интегрирующим их понятием диагноза, но и учесть уровни профессиональных и непрофессиональных факторов, с которыми каузально связаны эти признаки. С целью преодоления указанных трудностей, предлагается применение в диагностической практике проблемно-ориентированной информационно-аналитической системы (ИАС), реализующей элементы искусственного интеллекта для извлечения, накопления, применения и развития предметных знаний. Такие качественные инновационные улучшения в системно-информационном обеспечении медицины труда будут способствовать сохранению трудового потенциала предприятий, где существует риск заболевания ПН.

**Системная концепция.** Дифференциальная диагностика ПН рассматривается как задача многофакторной классификации наблюдения относительно ряда категорий, таких, как группа клинического риска и нозологическая форма. Степень соответствия данного наблюдения образцу каждой из категорий выясняется путем совместной оценки разнородных показателей, совокупность которых включает индексы функционального, соматического, психоневрологического, иммунологического, биохимического, анамнестического и гигиенического блоков. В каждой из постановок рассматриваемой задачи, структура оцениваемого объекта предопределена для однородной тестовой серии известной мощности, что обуславливает требование применения классифицирующих моделей, сформированных на основе обучения выборкой клинических случаев. Отчетливый эффект разделения целевых классов был получен в результате применения метода дискриминантного анализа обучающей выборки наблюдений, для идентификации и прогнозирования токсической энцефалопатии от воздействия комплекса химических веществ [6]. Эффективность применения полученных указанным методом информационно-математических моделей для решения задач медицинской диагностики широко показана практикой [7,8], чему, однако, в слабой степени сопутствует возможность формирования новых знаний о структуре анализируемых данных. Развитие возможностей наблюдения и анализа состояния здоровья работающих в динамике, увеличение прозрачности в соответствиях модельных процессов и артефактов данным наблюдений и врачебным заключениям, потребовала расширения и реорганизации методической базы сопровождения диагностики. Согласно обозначенным требованиям, была выработана концепция гибридной экспертной системы (ЭС) дифференциальной диагностики ПН «ONTIS» («Occupational NeuroinToxications Identification System»), в которой используется регенерируемая решающая модель на основе иерархической сети доверия Байеса. Указанная ЭС предназначена для использования в качестве интеллектуального сервиса ИАС, интегрируемой в метасистему, включающую медицинскую информационную систему (МИС) клинко-диагностического осмотра врачами - профпатологами и систему регистрации результатов санитарно-гигиенического мониторинга условий труда. Модель интеграции ИАС с метасистемой предполагает возможность конфигурирования глубины структурно-функционального совмещения ИАС и метасистемы при сохранении их способности к взаимодействию в условиях автономности программных платформ и средств обеспечения. Минимальный объем такого совмещения охватывает средства и протоколы

коммуникации ЭС с информационными ресурсами внешних систем, набор аналитических функций ИАС, сокращенный вариант ее базы данных (БД). Включение последней в конфигурацию дополнительно к БД метасистемы обуславливается необходимостью поддержки специальных сущностей, относящихся к функционированию ЭС, и файлового репозитория, предназначенного для сохранения ее модельных артефактов. В случае ограничений метасистемы или ее отсутствия, целесообразен выбор конфигурации ИАС, в соответствии с которой она реализует минимальный проблемно-ограниченный вариант функций регистрации, учета и хранения данных [9], а также предоставляет расширенные средства аутентификации пользователей и защиты чувствительной информации [10]. В любой из конфигураций ИАС, пользователям предоставляется доступ к набору программных средств, в соответствии с групповыми правами, относящимися к основным (гигиенист, клиницист, диагност-профпатолог) и вспомогательным (когнитолог, разработчик и администратор) системным ролям (рис.1).

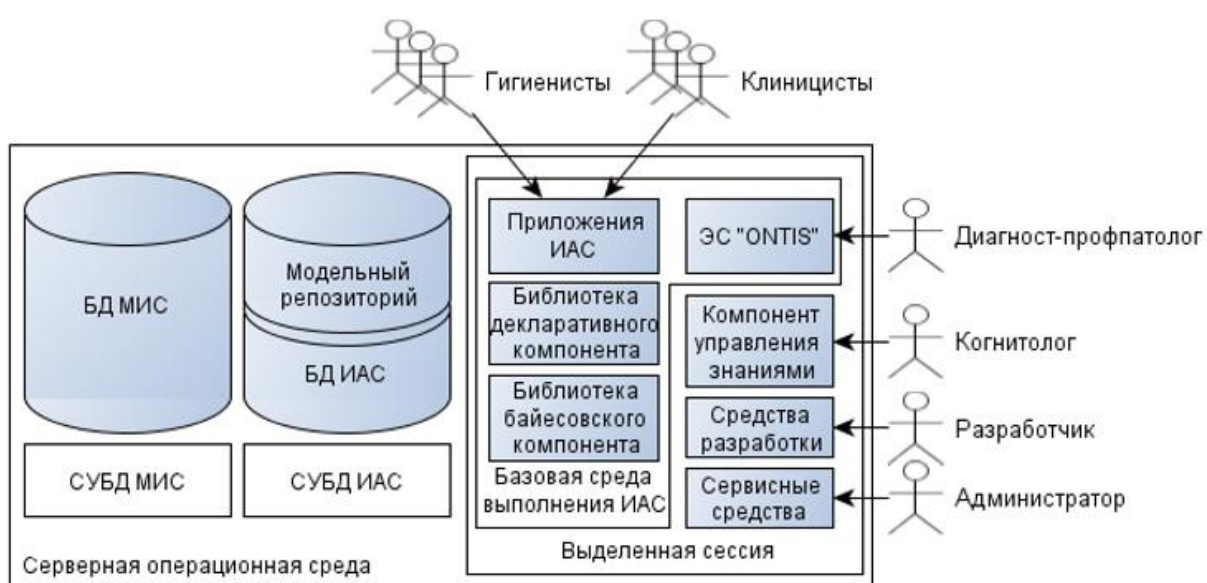


Рис. 1. Концепция системно-информационного обеспечения поддержки медицинской диагностики

**Модельно-технологическая концепция.** Обсуждаемая ИАС является человеко-машинной системой, предназначенной для сбора и хранения медико-социальных и гигиенических данных, осуществления их обработки в рамках дифференциально-диагностического цикла, в качестве одного из результатов предполагающей накопление и совершенствование формализованного проблемно-ориентированного опыта. Процесс развития знаний системы начинается с экспорта записей существующей диагностической истории, в результате чего в БД ИАС формируется первичная база наблюдений,  $S_{ПБН}$ . Описание каждого наблюдения включает набор гетерогенных индексных значений и соответствующую ему диагностическую формулировку, присвоенную в результате врачебного решения. Анализ выборки  $S^a_{ПБН}$ , где  $a$  - класс диагностируемой патологии, относительно критерия Шеннона или оценки расстояния между векторами данных «признак - диагноз», позволяет сжать признаковое пространство путем отбора наиболее информативных индексов [11]. Векторы значений полученной индексной группы

представляют ключевой когнитивный ресурс для идентификации базы классифицирующих правил (БКП), фиксирующих эмпирические закономерности, присутствующих в выборке. Поскольку число информативных наблюдаемых индексов, в основном, оказывается сравнительно большим (9-30), отдельные базы правил формируются для  $k$  выборочных перспектив  $\{P_k\}_1^K \subset S^a_{ПБН}$ , выделяемых тематически (по видам и блокам обследований), путем корреляционного анализа признаков индексов или экспертно. Для каждой  $P_k$ , определяется латентная переменная - общий результирующий фактор  $R(P_k)$ , так, что  $P_k \supset (\{I_m\}_1^M, R)$ , где  $\{I_m\}_1^M$  - набор индексов-наблюдений в  $P_k$ . Далее в  $S^a_{ПБН}$  определяется корневая перспектива,  $P_{root}$ , в качестве наблюдений включающая общие факторы  $\{P_1..P_k\}$ , а в качестве цели - показатель искомого диагноза. Если требуется рассмотреть симптомокомплекс в нескольких аспектах, то в итоговой структуре могут быть совмещены несколько перспектив  $P_{root}$ , целевой признак каждой из которых соответствует отдельному диагностическому аспекту. Совмещаемые экземпляры (слои)  $P_{root}$  разделяют обязательный набор признаков-свидетельств, определяемый общими факторами в  $\{P_1..P_k\}$ . В отдельные слои  $P_{root}$  при необходимости могут быть включены дополнительные признаки-наблюдения из  $\{\{I_m\}_1^M, R\}$  для  $\{P_k\}_1^K$ . В отдельных случаях, например, для обеспечения возможности совместной оценки аспектов диагностической картины, может потребоваться построение «субкорневой» перспективы. Область значений  $R(P_k)$  дискретна и, кроме  $P_{root}$ , определяется путем кластерного анализа  $P_k$  или с помощью аналогичного метода классификации без обучения, с последующим экспертным наименованием значений результата для выявленных групп. Следующим необходимым этапом идентификации БКП является дискретизация количественных показателей  $D(I_m)$ , выполняемая тривиально в случае заданных границ диапазонов нормы и патологии. Определение неизвестных значений порогов градации признака может быть выполнено на основе информационного критерия [12]. К массиву дискретных объектов  $\{p_i \supset (b_i, r_i)\}_1^{N(P_k)}, b_i \in D(I_m), r_i \in R(P_k)$ , применяется серия итераций дисперсионного анализа, в результате чего формируется вектор показателей силы влияния полиморфных качественных признаков-наблюдений и их сочетаний на результирующий индекс,  $\{\eta_k^2\}_1^{M+M'}(1)$ , где  $M > 0$  - количество проанализированных признаков,  $M' \geq 0$  - двухпризнаковых сочетаний. Если  $M' > 0$ , тогда формируется дополнительный уровень структуры  $S^a_{ПБН}$  с  $M'$  внутренних перспектив  $P_k^{*m}$ , каждая из которых включает пару столбцов наблюдений и искусственный общий фактор. Классифицирующие правила для  $P_k^{*m}$  формируются в соответствии с рассматриваемым способом. Представлением  $P_k^{*m}$  в  $P_k$ , принимается общий фактор. Наряду с вектором (1), базовая модель классифицирующего правила включает матрицу вкладов  $C = (c_{i,j})_{i=1, j=1}^{n, M-M'} = \{0..n\}$ , где  $n$  - наибольшее возможное число градаций, которое может варьироваться для различных признаков  $P_k$ . Значения  $C$  кодируют задаваемые экспертом минимальные начальные оценки вклада в частный диагностический результат  $j$ -го

показателя в  $i$ -й градации. Для шаблонов «норма-отклонение» и «снижение-норма-превышение»  $c_{i,j}$  может принимать значение 0 для нормы и 1 для девиации. При построении для частного заключения рангового домена типа «норма – подострое состояние – острое состояние», кодирование выполняется, начиная от 0 (для нормы) с единичным шагом. В качестве главного продукта такого анализа  $P_k$ , как набор преобразованных столбцов  $C$ , принимается итоговая оценочная матрица признаков вкладов  $C' \supset (c'_{i,j})_{i=1,j=1}^{n,m=M-M'}$ ;  $c'_{i,j} = c_{i,j} \cdot \eta_j^2$ ,  $n$  определяется по аналогии с матрицей  $C$ . Сумма оценок  $C'$  для выборочного объекта вида  $p_i \supset (b_i) \in P$  определяется как  $\zeta_i = \sum_{j=1}^{M-M'} c'(subseq(b_i, j), j)$ , где  $subseq(b_i)$  - функция адресации в  $C'$  к оценке градации  $b_{i,j}$ . Вычисление  $\zeta$  объектов требуется для предварительной маркировки порожденных объектов указанного вида, по классам внутри обучающей совокупности  $S_L^a$ , изоморфной  $S^a_{ПБН}$  и предназначенной для генерирования фрагментов байесовской решающей модели на основе перспектив  $\{P'_k\}_1^K \subset S_L^a$ . Каждую такую перспективу формирует набор объектов - уникальных комбинаций из  $M - M'$  индексных значений, при этом  $N(P'_k) = \tilde{A}_g^{M-M'}$  в случае, когда количество градаций  $g = const$ , иначе  $N(P) = \prod_{j=1}^{M-M'} q_j$ . Для завершения классификации обучающих объектов выполняется построение шкалы сумм, интервалы которой могут быть неравномерными, определяются и обозначаются в соответствии с информационным критерием [12] или экспертно. Окончательная маркировка объектов  $P'_k$  выполняется путем сопоставления  $\zeta$  на полученной шкале. Применимость обсуждаемой техники наилучшая для случаев однородного рангового домена  $R(P'_k)$  (например, степень выраженности симптома). Для перспектив  $S_L^a$  с номинальным доменом результата (совместные и несовместные симптомы и нозологические формы), возрастает значимость продукционного подхода к решению задачи классификации обучающих объектов. Существенные ограничения на применение описанного способа предобработки  $S^a_{ПБН}$  и  $S_L^a$  накладывает малый объем, разреженность, слабая репрезентативность  $S_{ПБН}$ . В подобном случае, для синтеза решающей модели ЭС, когнитологу необходимо вербализовать врачебный опыт, самостоятельно подготовить наборы необходимых экспертных свидетельств, и, при необходимости, базы продукционных классифицирующих правил для каждой перспективы в  $S_L^a$ . В ряде других случаев, когда все индексы перспективы определены на одном и том же ранговом домене, при одинаковой значимости вкладов наблюдений, способ маркировки объектов может быть тривиален и сводиться к нахождению предельного значения в группе индексов. Такая техника применяется, например, при определении группы соматического здоровья по наихудшему баллу в индексах оценки состояния систем организма. Путем интерпретации обучающих совокупностей в перспективах  $S_L^a$  при помощи «жадного» алгоритма с применением равномерной эквивалентной метрики Байеса-Дирихле (BDeu) [13], выполняется построение и параметрическое обучение фрагментов решающей модели

$\{BN_f\}_1^{K+1}$ , реализующих структурный шаблон классификатора с конвергентными присоединениями узлов-свидетельств к узлу-цели. Предопределенная конфигурация фрагмента с главной таблицей условных вероятностей (conditional probabilities table, CPT) в узле общего фактора  $T_i^1, i \in [1..k]$ , способствует сквозной модельной прозрачности в ЭС и обеспечивает возможность контроля обратной связи с обучающей совокупностью. Синтез иерархической сети доверия  $HBN^a$  (ИСД) осуществляется на уровне файловых описаний, путем интеграции фрагментов по сценарию, вырабатываемому на основе структурной метамодели  $S_L^a$ , пример которой приведен на рис. 2.

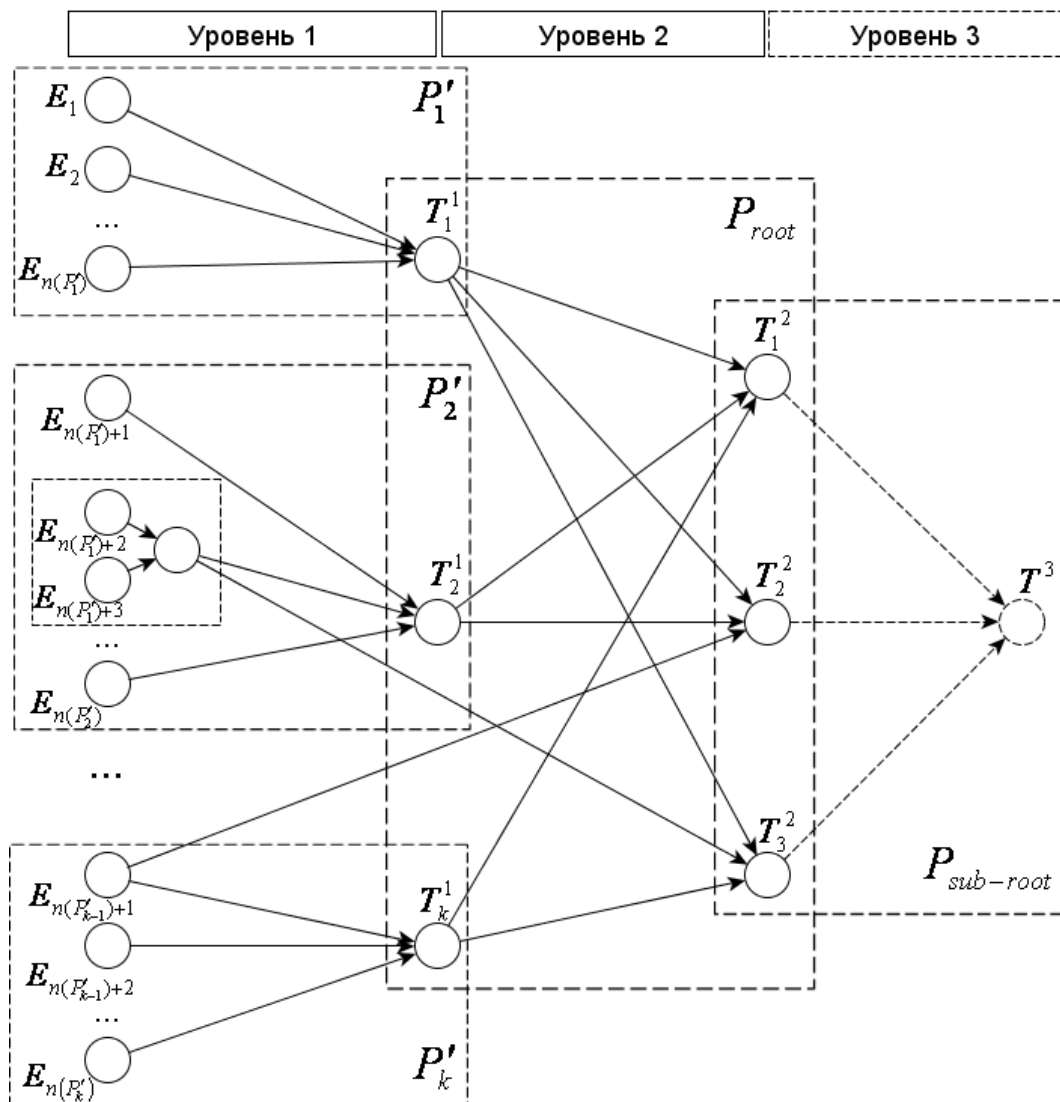


Рис. 2. Пример структурной метамодели для иерархической сети доверия

При условии методического единства в построении фрагментов сети, для регулирования степени точности значений CPT целевого узла каждого из них, объем сегмента обучающей совокупности в  $P'_k$  может быть изменен по принципу репликации выборки:  $N_i = N_{i-1} + N_1$ . При этом изменение точности условных вероятностей  $\Delta acc$  происходит по закону, близкому к убывающей экспоненциальной зависимости, и

одинаково для всех элементов данной СРТ. Приобретение опыта в ЭС происходит по мере ее врачебного применения, путем модификации сегментов  $S_L^a$  внесением конкретных примеров и контрпримеров. Скорость нормируемого перераспределения условных вероятностей в столбцах подчиняется той же закономерности, что и скорость регулирования точности СРТ. Указанное обуславливает нарастание разницы в степени уверенности суждений ЭС о случаях различных классов в процессе ее использования. В соответствии со структурной метамоделью  $HBN^a$ , при исследовании в ЭС клинического случая, на основании исходных свидетельств, в узлах-целях 1-го уровня формируется апостериорная вероятность частных оценок, значения которых передаются на входы узлов-целей 2-го уровня. Вектор апостериорных вероятностей этих узлов может быть интерпретирован диагнозом, а значения приняты в качестве итога консультации, и, при необходимости, сопоставлены в «субкорневой» перспективе, содержащей единственный целевой узел  $T^3$ . Для ЭС существует определяемая механизмом обработки нечетких фактов возможность задания в узлах-наблюдениях нормированных «мягких свидетельств», с чем связано модельное ограничение, указывающее на необходимость разделения домена состояний свидетельства по признаку совместности. Исключение данного признака на этапе определения значений свидетельств, приведет к некорректным результатам интерпретации фактов. Для поддержания целостности диагностической истории ЭС, каждый экземпляр решающей модели, принятый к использованию в практике, фиксируется после первого диагностического акта, сохраняется в модельном репозитории, и не может быть удален или модифицирован, а только использован или реплицирован для создания отдельного уточненного образца  $HBN^a$ . Жизненный цикл (ЖЦ) модельных артефактов ЭС представлена на рис. 3.

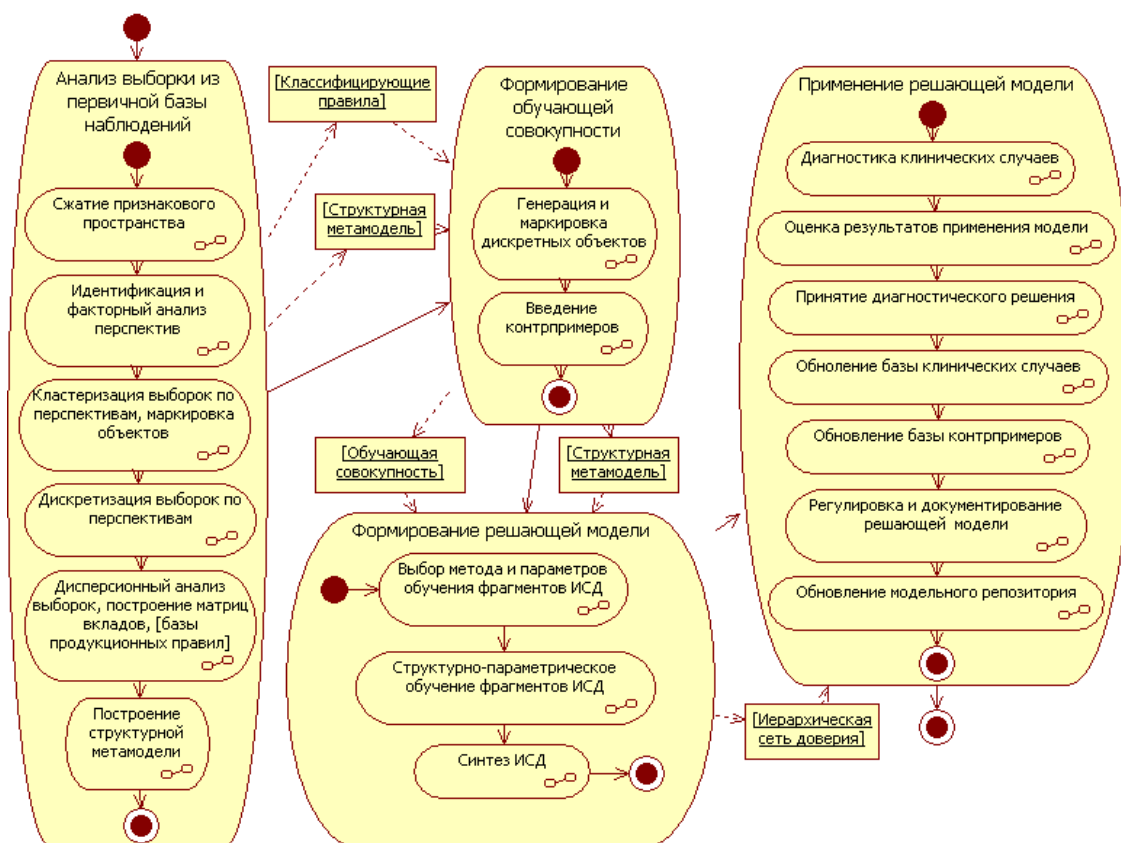


Рис. 3 . Последовательность этапов ЖЦ модельных артефактов ЭС «ONTIS»

**Реализационный подход.** Построение конфигураций ИАС регламентируется требованиями масштабируемости служб уровня данных, возможности построения комплекса приложений на основе единой программной технологии, обеспечения сетевого взаимодействия с компонентами метасистемы. Уровень данных обеспечивается СУБД Firebird<sup>2</sup>. Разработка ИАС ведется на платформе Java2 и сервера приложений jBoss<sup>3</sup>. В качестве основы ядра ЭС применяется библиотека построения и моделирования байесовских сетей «jSMILE»[14]. Для реализации в ЭС необходимых статистических техник выбрана библиотека исследования данных «WEKA»[15]. Собственная функциональность ЭС включает обмен с БД ИАС, дисциплины управления конструированием, настройкой, обменом модельных артефактов и их использованием для диагностического вывода. Исследование свойств и жизненного цикла (ЖЦ) решающих моделей может быть выполнено в графической среде «GeNIe 2.0» (рис. 4) [14].

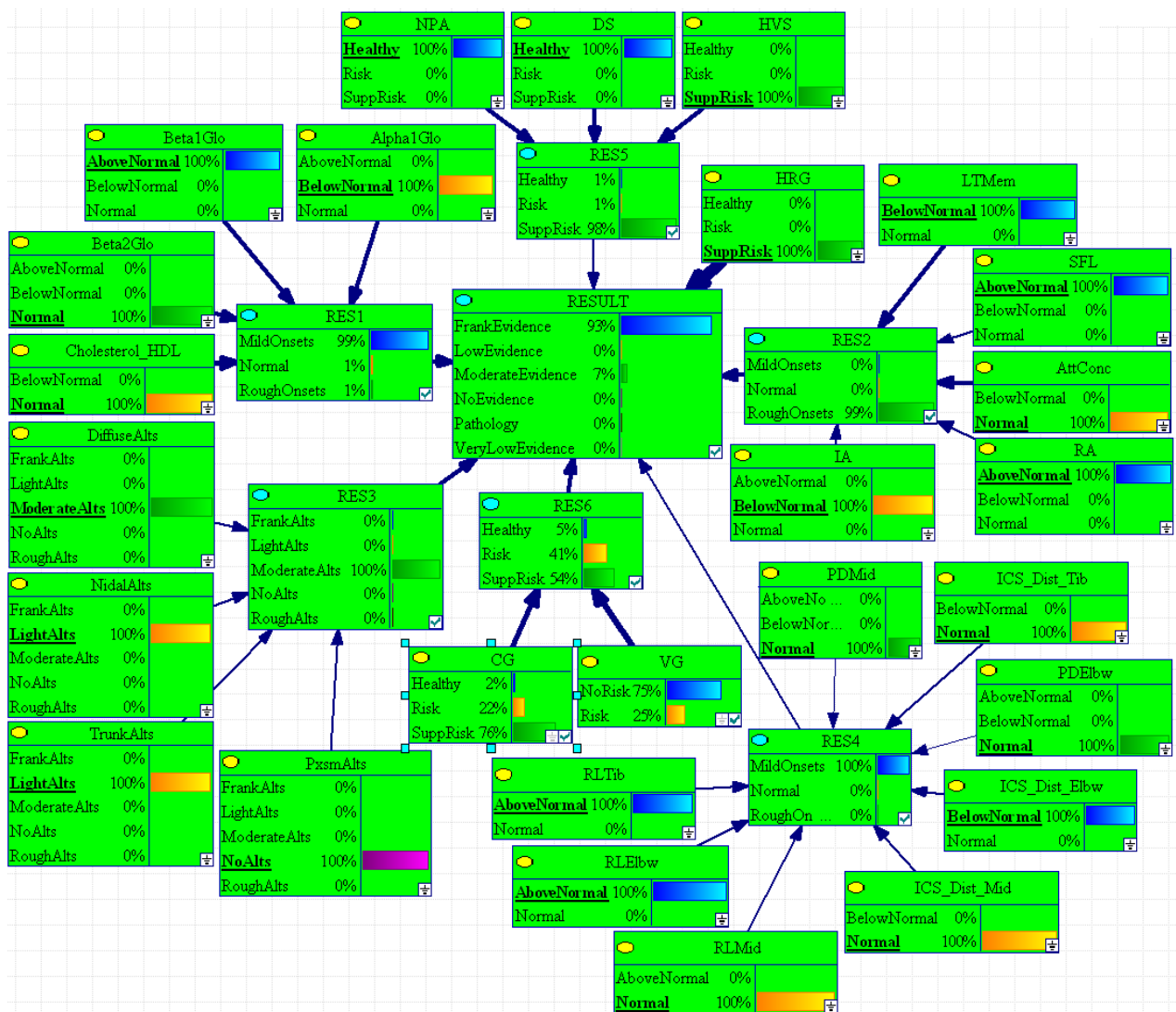


Рис. 4. Пример варианта решающей модели для диагностики хронической ртутной интоксикации (ХРИ)

<sup>2</sup> <http://www.firebirdsql.org/>

<sup>3</sup> <http://www.jboss.org>



Создание и сопровождение программной инфраструктуры может осуществляться с помощью интегрированных сред разработки для платформы Java2. Для взаимодействия с метасистемой служит разработка обменных форматов на основе XML, и утилит объектно-реляционного преобразования и транспорта между смежными информационными базами. Расширенный вариант граничных приложений ИАС ориентирован на использование легких каркасов веб-приложений и архитектурный шаблон MVC<sup>4</sup>. Общей средой исполнения компонентов ИАС является серверная операционная система, предоставляющая средства автоматизации таких сервисных задач, как резервирование данных, и обеспечивающая первичный механизм авторизации пользователей. Предлагаемый вариант реализации обсуждаемой технологии отвечает требованиям кросс-платформенности, гибкости и масштабируемости ИАС, и создает предпосылки к созданию методик и средств технического мониторинга системы, и планирования ее развития.

**Заключение.** Для обеспечения поддержки принятия решений в дифференциальной диагностике ПН, был предложен подход к созданию информационной технологии, основывающейся на предоставлении врачу-профпатологу проблемно-ориентированного экспертного инструмента, как сервиса информационно-аналитической системы в среде обработки клинико-гигиенических данных. Особенности подхода являются конфигурируемость интеграции ИАС в метасистему, гибридная организация ЭС, воплощающая комплекс статистических и декларативных методов, наличие механизма порождения начальной обучающей совокупности, итерационная модель ЖЦ системных компонентов. К достоинствам подхода относятся комплексная информационная поддержка идентификации ПН, сквозная модельная прозрачность ЭС, наглядность вывода, интерпретируемость и комментируемость вырабатываемых рекомендаций, способность к приобретению нового опыта, управление банком решающих моделей, применимость подхода к решению различных задач медицинской диагностики, отсутствие расходов на приобретение лицензий базовых программных средств и открытый исходных код существенной их части. К ограничениям подхода следует отнести существенную зависимость качества исходной решающей модели и трудоемкости процесса ее выработки от информационных свойств первичной базы наблюдений, и сравнительно большой объем вычислений при генерации и использовании решающих моделей. В качестве основных направлений дальнейшей работы рассматриваются формулирование и реализация баз классификационных правил и оценка применимости альтернативных интерфейсов декларативного языка PROLOG на платформе Java2, создание дополнительных приложений анализа и добычи данных для извлечения новых знаний об особенностях патогенеза ПН, выработка подходов к автоматизированному планированию развития системы.

## ЛИТЕРАТУРА

1. Трошин В.В. Профессиональные нейротоксикозы//Медицинский альманах. 2010.№2. С.52-61.

---

<sup>4</sup> <http://java.sun.com/blueprints/patterns/MVC-detailed.html>

2. Пивень Б.Н. Экзогенно-органические заболевания головного мозга. М.: Медицина; 1998. 144с.
3. Колесов В.Г., Мещерягин В.А., Лахман О.Л., Шевченко О.И. Психопатологические проявления в отдаленном периоде профессиональных нейротоксикозов. Журнал неврологии и психиатрии. 2005. № 1. С. 25-29.
4. Дьякович М.П., Казакова П.В. Ценностно-мотивационная сфера и качество жизни больных с хронической ртутной интоксикацией профессионального генеза. Известия Самарского научного центра Российской академии наук. 2008. № 2. С. 32-35.
5. Лахман О.Л. Течение энцефалопатии в отдаленном периоде профессиональной хронической ртутной интоксикации / О.Л. Лахман, В.Г. Колесов, О.К. Андреева // Медицина труда и про-мышленная экология. – 2003. - № 3. – С. 46 - 48.
6. Пат. 2279091 РФ. Способ диагностики токсической энцефалопатии от воздействия комплекса токсических веществ / Е.В.Катаманова, В.Г.Колесов, О.Л.Лахман, В.П.Ильин // 2006, Бюл. 18
7. Ульянычев, Н. В., Ульянычева, В. Ф., Колосов, В. П., Перельман, Ю. М. Использование дискриминантного анализа при разработке диагностических (прогностических) решающих правил / Н. В. Ульянычев и др. // Информатика и системы управления. - 2009. - N 4 (22). - С. 13-15
8. J.R. Bourne, V. Hamel, J.W. Ward, A discriminant-analysis-based scoring system for the electroencephalogram, in: Conference Proceedings IEEE Southeastcon 80, IEEE Press, New York, 1980.
9. Иванов А.Г. Бахвалов С.В. Организация яруса данных информационно-аналитической системы идентификации профессиональных нейроинтоксикаций / Вестник ИрГТУ. 2011. №8. С.81-86.
10. Иванов А.Г. Защита персональных данных в информационно-аналитической системе идентификации профессиональных нейроинтоксикаций / А.Г. Иванов, С.В. Бахвалов // Труды IV Всероссийской конференции с международным участием "Винеровские чтения", г. Иркутск, 2011 г., Том 2, с. 85-91.
11. Корлякова М.О., Твердохлеб Н.С. Анализ подходов к определению информативности признаков. // Научная сессия МИФИ-2006. Сборник научных трудов. В 16 томах. Т.3. Интеллектуальные системы и технологии. М.: МИФИ, 2006. 256 с. С. 146-147
12. Куренков, Н. И. Информационный критерий и его использование для решения задач обработки многомерных данных / Н. И. Куренков, С. Н. Ананьев // Информационные технологии. - 2007. - N 9. - С. . 59-64.
13. H. Steck (2008). Learning the Bayesian network structure: Dirichlet Prior versus Data. In D.A. McAllester and P.Myllymaki (eds.), Proc. 24th Conf. Uncertainty in Artificial Intelligence:511-518.
14. Marek J. Druzdel. SMILE: Structural Modeling, Inference, and Learning Engine and GeNie: A Development Environment for Graphical Decision-Theoretic Models. In Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99), pp. 342-343, Orlando, Florida, July 18-22, 1999

15. Remco R. Bouckaert, Eibe Frank, Mark A. Hall, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. WEKA-experiences with a java open-source project. *Journal of Machine Learning Research*, 11:2533-2541, 2010.