

О ПОСТРОЕНИИ БАЗЫ ЗНАНИЙ ПРОГНОЗИРУЮЩИХ ИНФОРМАЦИОННЫХ СИСТЕМ

ЕРОХИН А.Л., ПОДУФАЛОВ С.Ю.

Излагаются результаты исследований возможности построения базы знаний прогнозирующих систем на базе геоинформационных систем. Рассматриваются различные технологии прогнозирующих систем. Указываются основные пути создания прогнозирующих геоинформационных систем для управления энергосетями.

В настоящее время среди информационных систем выделилось отдельное направление - геоинформационные системы (ГИС). Они играют все более значительную роль в изучении среды обитания и организации повседневной деятельности человека. Эти системы можно разделить на две группы — констатирующие и прогнозирующие. Констатирующие системы включают в себя всю массу информации об имевших место ранее и протекающих в настоящее время явлениях. С увеличением объема и точности ценность этой информации непрерывно возрастает. Возрастает также и техническая сложность получения достоверной информации, что требует все более высокой точности и более значительных материальных расходов.

Приведем пример. Представляя собой некоторую систему, Земля является частью окружающего её пространства и неразрывно с ним связана, физическое состояние окружающего пространства непосредственно влияет на протекающие на Земле процессы. Непрерывное изменение состояния окружающего пространства вызывает соответственные непрерывные изменения физического состояния Земли. Взаимосвязанность процессов приводит к необходимости получения взаимосвязанной информации, которая должна служить аналитической основой выявления реальных причинно-следственных связей между явлениями и служить базой для создания ГИС второй группы - прогнозирующих.

Прогнозирующие ГИС качественно отличаются от систем констатирующих. Первые должны дать возможность выявить условия возникновения конкретных физических явлений, определить момент и место их проявления и охарактеризовать протекание их во времени. Поэтому прогнозирующие системы требуют не только высокой точности исходной информации, но также доказательной адекватности математической модели и высокой точности технологий обработки информации.

Актуальной задачей является создание прогнозирующих систем для управления энергетическими сетями. В таких инженерных сетях должны решаться задачи управления потокораспределением [2], прогнозирования нагрузки, накопления “образов” нештатных ситуаций, предсказания аварийных

ситуаций, поддержки принятия решений при авариях.

При создании прогнозирующих ГИС можно выделить такие основные технические проблемы. Это описание физического состояния исследуемого объекта (например, Земли или инженерной сети) и геометрических его особенностей, получение исходных данных необходимого качества, наличие аналитического аппарата, способного своевременно и с требуемой точностью получить и обработать исходные данные и определить реакцию объекта как единой системы на воздействие на нее окружающего пространства.

Проблема аналитического описания физических и геометрических особенностей окружающего пространства решается современной наукой с достаточно высокой точностью. Уровень современных знаний в этой области дает возможность получить необходимую для анализа состояния информацию. В первую очередь это относится к возможности описания силовых полей окружающего пространства в требуемый для анализа момент времени.

Успехи в решении проблемы описания физического состояния Земли и её геометрической формы значительно скромнее. С помощью современных спутниковых технологий может быть получено геометрическое описание земной поверхности с весьма высокой степенью точности. Однако вопросы исследования физического состояния Земли не только как единого физического тела, но, в частности, и физического состояния земной коры, атмосферы и гидросферы требуют значительных усилий. Изучение физического состояния земной коры, атмосферы и гидросферы, как непосредственной среды обитания человека, имеет в настоящее время первостепенное значение.

Довольно сложной и трудоемкой, требующей значительных материальных затрат, является проблема получения исходных данных требуемого качества. Сложность этой проблемы заключается не только в необходимости высокой точности выполняемых измерений и в том, что зачастую исходные данные невозможно получить с помощью прямых измерений. Сложность её заключается также в том, что исследуется непрерывно изменяющаяся система и необходимо получать обширные взаимосвязанные данные, соответствующие определенным известным её изменениям. В то же время, существующие технические возможности позволяют получить значительный объем исходной информации.

При анализе существующего физического состояния Земли и его прогнозировании возникает необходимость быстрой обработки большого количества различных исходных данных, которые должны соответствовать одному и тому же моменту времени. Существующие средства связи и вычислительной техники обладают высокой производительностью и способны обеспечить как своевременное получение необходимой исходной информации, так и требуемую скорость и точность её обработки.

Сложной с разных точек зрения является проблема используемого для обработки исходной информации аналитического аппарата. Применяемые в настоящее время аналитические технологии либо приводят к неопределенности решения исходной задачи и для её раскрытия требуют применения методов, изменяющих саму анализируемую задачу, либо используют численные математические методы, которые не являются аналитическими и, удовлетворяя априорно принятым критериям, не дают возможности получить результаты требуемой точности.

Рассмотрим различные технологии прогнозирующих систем.

1. Физическая геометрия. Использование геометрии, объективно связывающей параметры протекающих в реальных телах физических процессов с их геометрической формой, позволяет решить эту проблему. Аналитическая технология определения физических параметров, протекающих в реальных телах процессов, основанная на физической геометрии, дает возможность получить результаты обработки исходной информации требуемой точности и создать адекватную математическую модель процесса [1].

2. Распознавание образов и использование эталонов. Если программа находит в эталонном наборе признаков похожий, то событие в системе идентифицировано и ему присвоено соответствующее значение кода.

3. Анализ признаков. Признаки, которые нужно идентифицировать, измеряются и сравниваются с геометрическими параметрами.

4. Обучение. Если пользователь (диспетчер, эксперт) исправит неправильно распознанные аварии и запишет эти исправления в виде “обученных” данных, то программа сможет использовать эти данные в будущих процессах распознавания.

5. Нечеткая логика. Это процесс принятия решения, при котором условия и правила не должны выполняться стопроцентно. При идентификации учитывается даже относительно нечеткая выраженность признаков. С помощью этого способа распознавания образов можно комбинировать различные результаты нескольких способов контроля. Это повышает вероятность распознавания в критических случаях, например, при нечетко выраженных причинах аварии в энергосистеме.

Поэтому, начиная с середины 80-х годов, разработчики стали экспериментировать с такими способами распознавания, которые были разработаны на основе аналогии с человеческим мозгом.

В естественных и искусственных нейронных сетях обработка информации происходит в соответствии с законами так называемой нечеткой, или fuzzy-логики. Алгоритмические способы решения проблемы, такие как распознавание образов, работают по раз и навсегда установленным четким правилам: в обучающихся системах создаются дополнительные

временные правила, которые тем не менее формулируются совершенно однозначно. Нечеткая логика учитывает тот факт, что мозг может прийти к приемлемому решению проблемы даже без однозначных параметров.

При этом естественно в размышлениях используются такие вербальные понятия как “довольно”, “почти” или “значительно”. Fuzzi-логические экспертные системы “думают” аналогично человеку, их принцип работы не аналогичен цифровой системе. Они знают не только “истина” и “ложь”, но также “много” и “мало”. Однако чтобы понятия “мало”, “больше” и “много” можно было использовать, они должны быть по крайней мере четко разграничены между собой.

Для этого результаты исследования технологий низшего уровня (например, анализ признаков) сначала преобразуются в нечеткие, постепенно изменяющиеся параметры, так называемые лингвистические (вербальные) переменные. В системах с нечеткой логикой эти переменные объединяют несколько точных анализируемых параметров. Такой комплекс параметров называется также нечетким множеством. Каждому такому множеству присваивается собственное имя переменной. Обычно для этого используют следующие сокращения:

- NB (negative big) — для больших отклонений;
- NM (negative medium) — для средних отклонений;
- NS (negative small) — для малых отклонений;
- ZE (zero) — при отсутствии отклонений.

В системах с нечеткой логикой для обозначения дополнительных отклонений от идеального значения можно использовать также и другие переменные (PS - positive small, PM - positive medium и PB - positive big).

Насколько точно математико-геометрическая величина признака, найденная по способу анализа признаков, указывает, например, на аварию типа “обрыв линии вследствие роста дерева”, можно представить в виде функции принадлежности. В этом случае степень принадлежности ко множеству всех возможных признаков выражается числом от 0 до 1.

Нечеткая логика может выступать как посредник между рассмотренными технологиями. Если несколько таких нечетких величин, которые могут быть получены из различных экспертных проверок, связать между собой операторами И, ИЛИ и ЕСЛИ-ТО, то вероятность правильного заключения повышается.

Для программ распознавания аварий это означает следующее: ЕСЛИ два признака аварии 1 выражены недостаточно четко, И однозначно существуют три других И, контроль показал, что возможна комбинация с предшествующей аварией (признаком) 2 и последующей 3, ТО такая экспертная система решает в пользу 1.

Результаты операции ЕСЛИ-ТО снова передаются на соответствующие переменные и окончательно преобразуются в термины нечеткой логики. Точным выходным параметром будет признак 1. Системы, построенные на нечеткой логике, могут выдавать окончательные результаты исследования, которые явно противоречат некоторым используемым правилам.

Экспертная система, основанная исключительно на правилах, ошибется уже на первом невыполненном правиле и для перестраховки предоставит выбор пользователю.

Функционирование любой прогнозирующей системы может быть сведено к решению совокупности соответствующих задач. В работах по данной теме обычно отсутствует общепринятый термин “задача”. При этом речь, как правило, идет не только о строгой словесной формулировке этого термина, но и об отсутствии необходимых знаний о структуре задачи, ее основных компонентах, видах ее конкретных представлений и т.д., что, несомненно, тормозит решение ряда важных проблем, связанных с созданием прогнозирующих ГИС.

При теоретическом анализе задач и процессов их решения обычно ограничиваются характеристикой понятия задачи в рамках какой-либо из сложившихся наук. Между тем для проектирования и создания прогнозирующих систем важно нахождение того общего, что есть в подходах различных наук к описанию задач, и определение на этой основе общенаучного понятия задачи [3].

В [3] было предложено определять задачу T в виде четверки:

$$Nm : T = \langle X, Q, F, Y \rangle, \quad (1)$$

где Nm - имя задачи; X - входные данные и их спецификация; Q - описание того, что необходимо найти в задаче (описание цели), и описание условий, при которых цель может быть достигнута; F - конечная последовательность правил (= алгоритм), преобразующая X в Y в соответствии с Q ; Y - выходные данные.

Уточним и детализируем компоненты, входящие в определение понятия “задача”. Будем считать, что имя задачи $Nm(T)$ и имя алгоритма ее решения $Nm(F)$ совпадают, т.е. $Nm(T) = Nm(F)$. Пользуясь введенным соглашением, в дальнейшем преимущественно будем ссылаться на $Nm(F)$.

Будем различать входные данные $Dt(X)$ и их спецификацию $Sp(X)$. Очевидно, что одному и тому же $Dt(X)$ могут соответствовать в разных задачах различные спецификации, и наоборот. Например, $Dt(X) =$ (рельеф) можно специфицировать и как “строку символов”, и как “слово русского языка, обозначающее вид поверхности участка Земли”. Аналогично будем различать вы-

ходные данные $Dt(Y)$ и соответствующую им спецификацию $Sp(Y)$.

Каждому F ставится в соответствие описание его функционального назначения, обозначаемое через $As(F)$. Оно семантически связано с $Sp(Y)$. Пусть $As(F) =$ (решить проблему, в каком районе города отключить свет при перерасходе лимита электроэнергии), а $Sp(Y) =$ (отключение света в определенном районе города). Эти два понятия связаны между собой семантическими отношениями синонимии: “Решить проблему отключения света при перерасходе лимита электроэнергии – это значит найти определенный район города, в котором можно отключить свет при наименьшем ущербе”. Кроме функционального назначения, для каждого F определена его спецификация $Sp(F)$, включающая в себя язык описания F , метод решения, требуемые вычислительные ресурсы для его реализации и т.п. И, наконец, каждому F ставится в соответствие условие его выполнения - Cnd .

С учетом сделанных уточнений и добавлений введем понятие “решенная задача”:

$$T = \langle Dt(X), Sp(X), Nm(F), As(F), Sp(F), Cnd(P), Pr(F) \nabla (Dc(F) \& \& Pr(F)), Dt(Y), Sp(Y) \rangle, \quad (2)$$

где символ ∇ – знак логической операции, “разделительное ИЛИ”; $\&$ – конъюнкция.

Приведенное описание характеризует конкретную задачу. Учитывая свойство массовости алгоритма F , опишем задачу в общем виде:

$$T = \langle Sp(X), Nm(F), As(F), Sp(F), Cnd(F), Dc(F) \vee Pr(F), Sp(Y) \rangle. \quad (3)$$

Организация хранения знаний играет решающую роль в функционировании информационных систем. Рассмотрим принципы организации хранения знаний, основу которых составляют описания задач в виде (2) и (3).

Пусть задано конечное множество описаний задач S . Каким требованиям должно удовлетворять это множество, чтобы база знаний, сформированная на его основании, обеспечивала возможность “интеллектуального поведения” информационной системы?

Требования, предъявляемые к множеству задач S . Процесс решения любой задачи сводится к реализации правил замены:

$$Dt(X)(Sp(X)) \Rightarrow Dt(Y)(Sp(Y)), \quad (4)$$

где символ \Rightarrow читается как “...заменяется на...” и служит для обозначения соответствующего отношения, которое (для одного и того же F) является

антирефлексивным (поскольку $\forall T(Sp_i(X) \neq Sp_i(Y))$) и антисимметрическим, так как

$$\forall T_i \neg \exists! F_i((Sp_i(X) \Rightarrow Sp_i(Y) \& \& Sp_i(X) \Rightarrow Sp_i(X))) . \quad (5)$$

Свойство транзитивности вообще не имеет смысла, поскольку

$$\forall F_i \exists! Sp_i(X) \exists! (Sp_i(Y)(Sp_i(X) \Rightarrow Sp_i(Y)) ,$$

где $\exists!$ обозначает квантор единственности.

Такое правило будем называть правилом F-замены.

Таким образом, сопоставление различных задач из S достаточно проводить на уровне $Sp(X)$ и $Sp(Y)$ этих задач.

Требование 1:

$$\neg \exists T_j \in S(Sp_i(X) = Sp_i(Y) \& \& (Sp_i(Y) \subset Sp_i(X))) . \quad (6)$$

Требование 2:

$$\forall T_i \in S \exists T_j \in S(T_i \cap T_j \neq 0) . \quad (7)$$

Требование 2 означает, что множество S не содержит изолированных, или независимых задач.

Требование 3:

$$\forall T_i \in S \neg \exists T_j \in S(T_i = T_j) . \quad (8)$$

Данное требование подчеркивает уникальность каждой задачи из S с учетом всех ее атрибутов. Будем различать начальные, промежуточные и конечные задачи.

Определение 1. Задача $T_i \in S$ называется начальной, если

$$\neg T_j \in S(Sp_i(X) \cap Sp_i(Y) \neq 0) . \quad (9)$$

В практическом плане важную роль играют частично начальные задачи.

Определение 2. Задача $T_i \in S$ называется частично начальной, если

$$\begin{aligned} & \exists Sp_i(X_\mu) \subset Sp_i(X) \exists Sp_i(X_\nu) \subset \\ & \subset Sp_i(X) Sp_i(X_\mu) \cup Sp_i(X_\nu) = Sp_i(X) , \\ & \neg \exists T_j \in S(Sp_i(X_\mu) \cap Sp_i(Y) \neq 0) \exists T_k \in S \forall , \\ & Sp_i(X_{\nu_s}) \in Sp_i(X_\nu) , \\ & Sp_i(X_{\nu_s}) \in Sp_k(Y) . \end{aligned} \quad (10)$$

Определение 3. Задача $T_i \in S$ называется конечной, если

$$\neg \exists T_j \in S(Sp_i(Y) \cap Sp_j(X) \neq 0) . \quad (11)$$

Определение 4. Задача $T_i \in S$ называется частично конечной, если

$$\begin{aligned} & \exists Sp_i(Y_\mu) \subset Sp_i(Y) \exists Sp_i(Y_\nu) \subset \\ & \subset Sp_i(Y) Sp_i(Y_\mu) \cup Sp_i(Y_\nu) = Sp_i(Y) , \\ & \neg \exists T_j \in S(Sp_i(Y_\mu) \cap Sp_i(X) \neq \\ & \neq 0) \exists T_k \in S \forall Sp_i(Y_{\nu_s}) \in Sp_i(Y_\nu) , \\ & Sp_i(Y_{\nu_s}) \in Sp_k(X) . \end{aligned} \quad (12)$$

Определение 5. Задача $T_i \in S$ называется промежуточной, если

$$\begin{aligned} & \forall Sp_i(X_\mu) \in Sp_i(X) \forall Sp_i(Y_\nu) \in \\ & \in Sp_i(Y) \exists T_j \exists T_k ((Sp_i(X_\mu) \in \\ & \in Sp_i(Y) \& (Sp_i(Y_\nu) \in Sp_k(X))) . \end{aligned} \quad (13)$$

Требование 4:

$$\begin{aligned} & \forall T_i \in S \exists T_j \in S(Sp_i(X) \cap \\ & \cap Sp_j(X) = 0) \forall (Sp_i(X) = \\ & = Sp_j(X) \forall (Sp_i(X) \subset \\ & \subset Sp_j(X) \forall (Sp_i(X) \cap Sp_j(X) \neq 0) . \end{aligned} \quad (14)$$

Требование 5:

$$\begin{aligned} & \forall T_i \in S \exists T_j \in S(Sp_i(Y) \cap \\ & \cap Sp_j(X) = 0) \forall (Sp_i(Y) = \\ & = Sp_j(Y) \forall (Sp_i(Y) \subset \\ & \subset Sp_j(Y) \forall (Sp_i(Y) \cap Sp_j(Y) \neq 0) . \end{aligned} \quad (15)$$

Требования 4, 5, являясь избыточными, призваны путем явного перечисления подчеркнуть, что между $Sp(X)$ и $Sp(Y)$ различных задач могут быть любые связи или они могут отсутствовать.

Требование 6:

$$\begin{aligned} & \exists T_i \exists T_j ((Sp_i(X) = Sp_j(X)) \& \\ & \& Sp_j(Y) = Sp_j(Y)) \rightarrow (Nm(F_i) \neq \\ & \neq (Nm(F_j)) \& (Sp(F_i) \neq Sp(F_j)) , \end{aligned} \quad (16)$$

где $A \rightarrow B$ обозначает, что “A приводит к B” или “B логически следует из A”.

Данное требование уточняет требование 3 и определяет возможность объединения требований 4 и 5 применительно к двум задачам T_i и T_j для случая, когда $Sp_i(X) = Sp_j(X)$ и $Sp_i(Y) = Sp_j(Y)$.

Выполнение требования 4 (в целях выделения конкретного вида связи) по отношению к начальным задачам можно проверить (при работе с S большой размерности) только при условии, что $Sp(X)$ обладает свойством расслоения.

Определение 6. Будем говорить, что $Sp(X_\mu)$ расслаивается, если

$$\exists T_i \in S \exists T_j \in S ((Sp(X_\mu) \in Sp_i(X) \& (Sp(X_\mu) \in Sp_j(X))), \quad (17)$$

где T_j и T_i – начальные задачи.

Следующим определением введем аналогичное понятие для $Sp(Y)$.

Определение 7. Будем говорить, что $Sp_k(Y_\mu) \in Sp_k(Y)$ расслаивается, если

$$\exists T_i \exists T_j \exists T_k ((Sp_k(Y_\mu) \in Sp_i(X)) \& (Sp_k(Y_\mu) \in Sp_j(X))).$$

Проверка выполнения требования 4 (в целях выделения конкретного вида связи) по отношению к конечным задачам будет возможной (также при работе с S большой размерности) только при условии, что $Sp(Y)$ обладают свойством слияния.

Определение 8. Будем говорить, что $Sp_i(Y_\mu) \in Sp_j(Y)$ конечной задачи T_i сливается с $Sp_i(Y_\nu) \in Sp_j(Y)$ конечной задачи T_j , если $Sp_i(Y_\mu) = Sp_j(Y_\nu)$.

Для промежуточных задач понятие “слияние” вводится следующим определением.

Определение 9. Будем говорить, что $Sp_i(Y_\mu)$ и $Sp_i(Y_\nu)$ сливается, если

$$\exists T_i \exists T_j \exists T_k (((Sp_i(Y_\mu) \in Sp_i(Y) = (Sp_j(Y_\nu) \in Sp_j(Y)) \& (Sp_i(Y_\mu) \in Sp_k(X) \& (Sp_i(Y_\nu) \in Sp_k(X))).$$

В данной работе не рассматривается вопрос связности элементов $Sp(X_i) \in Sp(X)$ и $Sp(Y_i) \in Sp(Y)$ между собой в пределах соответствующих множеств $Sp(X)$ и $Sp(Y)$ и между этими множествами.

Введенные определения и сформулированные требования, предъявляемые к множеству решенных задач S , подчеркивают фактор связности описаний задач, который приводит в результате сопоставления этих описаний и выделения сходного в различном и различного в сходном к соответствующей структуре.

Таким образом, существует реальная возможность решения задачи создания прогнозирующих ГИС. Известные методы нечеткого условного вывода нуждаются в улучшении. База знаний и сама система поддержки принятия решений должна адаптироваться к прогнозируемым ситуациям и быть способна осуществлять функции управления не хуже, чем лицо, принимающее решение в комфортной обстановке.

Реализация механизма адаптации системы поддержки принятия решений позволит исключить человека-оператора из временного баланса процесса управления, устранить неблагоприятное влияние человеческого фактора в условиях дефицита времени на эффективность функционирования энергосистемы.

Кроме существующих технических проблем, имеется также проблема организационная. Создание прогнозирующих информационных систем (таких, как исследование физического состояния Земли или система поддержки принятия решений при авариях в энергосистеме) представляет собой многофакторную задачу и требует применения обширных знаний и информации в различных научных дисциплинах. Это исключает возможность решения данной задачи научным коллективом, специализирующимся в одной научной дисциплине. Проблема заключается в практической необходимости создания работоспособного координированного научного сообщества, обладающего большими потенциальными возможностями. Если удастся объединить усилия для совместного решения поставленной глобальной задачи, то уже сейчас можно считать реально возможным создание прогнозирующих геоинформационных систем, насущно необходимых человеку в его повседневной жизни и деятельности.

Литература: 1. *Тези доповідей третьої Всеукраїнської конференції з геоінформаційних технологій “Теорія, технологія, впровадження ГІС” / ГІС-Асоціація України. К. 1997. С. 45-47* 2. *Приобретение знаний / Под ред. С. Осуги, Ю. Саэки. М.: Мир, 1990. С. 175-179* 3. *Тевяшев А.Д.* Оперативное управление потокораспределением в инженерных сетях в условиях неопределенности / Автореф. дис. на соиск. уч. степени д-ра техн. наук. Харьков: ХИРЭ, 1994. 58 с. 4. *Варсак М.И., Ловицкий В.А.* Вопросы анализа и синтеза задачно-решающей системы оперативного управления основным производством // АСУ и приборы автоматизации. 1975. Вып. 34. С. 37-40 5. *Бакаев А.А.* Методы организации и обработки знаний. АН Украины. Институт кибернетики им. В.М.Глушкова. К.: Наук. думка. 1993. 149 с.

Поступила в редколлегию 04.10.2000

Рецензент: д-р техн. наук Зацеркляный Н.М.

Ерохин Андрей Леонидович, канд. техн. наук, доцент кафедры информатики Университета внутренних дел Украины, доцент кафедры программного обеспечения ХТУРЭ. Научные интересы: системы поддержки принятия решений, современные компьютерные информационные технологии. Увлечения и хобби: компьютер, радиолюбительство, музыка. Адрес: Украина, 61166, Харьков, пр. 50-летия СССР, 27, тел. 50-30-80, 40-94-46, e-mail: eal@adm.univd.kharkov.ua

Подуфалов Сергей Юрьевич, аспирант кафедры программного обеспечения ЭВМ ХТУРЭ. Научные интересы: современные компьютерные информационные технологии. Увлечения и хобби: компьютер, музыка. Адрес: Украина, Харьковская обл., г. Чугуев, ул. Кожедуба, 9, кв.63, тел. (05746) 2-32-60, (0572) 40-94-46.