

**ОБ ОДНОМ СПОСОБЕ ОБРАБОТКИ ПРИЧИННО-СЛЕДСТВЕННЫХ СВЯЗЕЙ В
ЭКСПЕРТНОЙ СИСТЕМЕ ПРОГРАММНОГО КОМПЛЕКСА ПРИМА**

Червинский С.М.
НИЦ «НиР БСМ» УрО РАН, Екатеринбург

АННОТАЦИЯ

Базы знаний экспертных систем нередко базируются на байесовских сетях, образуемых диагнозами, один из которых должна поставить система, и связанными с ними признаками (фактами), по своей сути являющимися следствием того или иного диагноза. Однако, как показано в статье, если признаки коррелированы, например, сами образуют причинно-следственные связи, то поведение системы может стать нелогичным и в этом смысле неадекватным. Предлагаемый способ обработки подобных признаков позволяет не только уменьшить время реакции системы, но и придать ее диалогу с пользователем более осмысленный характер.

ВВЕДЕНИЕ

Для анализа состояния трубопровода, оптимизации его эксплуатации и автоматизации этапов его расчета в НИЦ «НиР БСМ» УрО РАН разработан программный комплекс (ПК) ПРИМА. Одним из модулей ПК является экспертная система (ЭС), машина вывода которой основана на вероятностном анализе.

База знаний ЭС представляет собой байесовскую сеть, узлы которой суть — некоторые переменные-события, а связи между узлами — вероятностные отношения между событиями. Определенная группа попарно несовместных событий выделена в отдельный класс — это гипотезы (диагнозы) $h_i, i=1,2,\dots,n$, вероятность которых требуется определить. Остальные события относятся к фактам $e_j, j=1,2,\dots,m$, также называемым свидетельствами. С каждым узлом сети связан определенный хранимый в базе знаний (БЗ) текст: для гипотезы это — подробное описание диагноза, для свидетельства — вопрос, ответом на который оператор системы подтверждает или опровергает достоверность события e_j .

Пусть свидетельство e_j связано с k гипотезами $h_i, i=1,2,\dots,k$ таким образом, что из наступления события h_i следует наступление события e_j , другими словами:

$$\bigcup_{i=1}^k h_i = h_1 + h_2 + \dots + h_k \Rightarrow e_j.$$

Значит, свидетельство e_j может быть представлено в виде суммы попарно несовместных событий: $e_j = e_j h_1 + e_j h_2 + \dots + e_j h_k$, а его вероятность определяется по формуле полной вероятности:

$$P(e_j) = \sum_{i=1}^k P(h_i)P(e_j|h_i)$$

Таким образом, в БЗ кроме упомянутых текстов необходимо хранить априорные вероятности всех гипотез $P(h_i)$, характеризующие узлы h_i , и вероятности $P(e_j|h_i)$, определяющие значения связей между узлами e_j и h_i (Рис.1.). Вероятности $P(h_i)$ и $P(e_j|h_i)$ задаются экспертами при формировании базы знаний, тогда как вероятности свидетельств $P(e_j)$ вычисляются машиной вывода ЭС на каждом шаге итерации. Основную трудность для экспертов составляет определение $P(e_j|h_i)$ — вероятность появления признака e_j при условии, что подтвержден диагноз h_i . В то же время все гипотезы первоначально можно считать равновероятными, приняв $P(h_i)=1/n, i=1,2,\dots,n$. Это позволит избежать погрешностей при распознавании маловероятных (редких) диагнозов. После обработки любого свидетельства e_j априорные вероятности $P(h_i)$ всех связанных с ним гипотез заменяются апостериорными

$P(h_i|e_j)$, и дальнейшие вычисления строятся уже на них. $P(h_i|e_j)$ определяется формулой Байеса:

$$P(h_i|e_j) = P(h_i) P(e_j|h_i) / P(e_j)$$

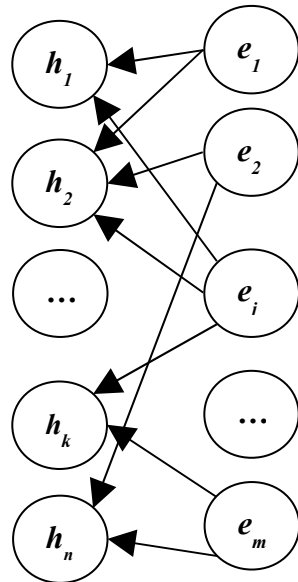


Рис. 1. Пример сети.

Байесовская сеть экспертной системы — классическая модель причинно-следственных связей, в которой на основании свидетельств (следствий) определяется соответствующая гипотеза — причина, их вызывающая. Машина вывода ЭС программного комплекса ПРИМА может работать как в автоматическом, так и в диалоговом режиме. В первом вероятностный вывод производится путем построения и анализа так называемой диагностической таблицы, в которой вычисляются апостериорные вероятности гипотез при переборе всех сочетаний свидетельств со значениями *подтверждено/опровергнуто*. В диалоговом режиме сначала определяется наиболее значимое свидетельство, и задается связанный с ним вопрос, на который (в отличие от автоматического режима с двумя вариантами ответа) может быть введено неопределенное значение. Пользователь устанавливает мышкой бегунок на экранной шкале в диапазоне $[-1, +1]$. Так, ответу 'нет' соответствует -1 , 'да' — $+1$, 'не знаю' — 0 . Число -0.8 , например, можно оценить как 'скорее всего, нет'. После получения ответа для свидетельства e_j следует расчет апостериорных вероятностей $P(h_i|e_j)$ всех связанных с этим свидетельством гипотез и замена ими априорных вероятностей $P(h_i)$. Затем машина вывода пересчитывает веса оставшихся свидетельств и задается вопрос о свидетельстве с наибольшим весом. Процесс повторяется до тех пор, пока какая-либо гипотеза не будет определена системой как наиболее вероятная. (Механизмы учета неопределенных ответов пользователя, определения наиболее вероятной гипотезы и нахождения веса свидетельства в данной статье не рассматриваются.) Исходя из начального состояния БЗ, очевидно, что первый вопрос системы будет одинаков для любого сеанса работы ЭС. Чего уже не скажешь о последующих вопросах, так как ответ на первый заранее неизвестен.

В силу того, что порядок обработки свидетельств (а значит и порядок следования вопросов) заранее не определен, важно, чтобы они не были противоречивыми, неопределенными или взаимоисключающими. Это может быть достигнуто при независимых свидетельствах. На практике свидетельства часто считают независимыми (особенно, если их число велико) даже при наличии существенной корреляции между ними[1]. Допущение о

независимости упрощает обработку байесовской сети и вполне подходит для автоматического режима функционирования экспертной системы. Однако в диалоговом режиме могут возникнуть логические неувязки, вызывающие у пользователя ощущение получения повторных или противоречивых вопросов.

ТРАДИЦИОННЫЙ ВЕРОЯТНОСТНЫЙ ВЫВОД

Для лучшего понимания необходимости обработки связанных свидетельств в экспертной системе ПК ПРИМА имеет смысл рассмотреть традиционный вероятностный вывод на примере базы знаний геологоразведки. Пусть среди ста диагнозов в базе знаний у нас имеется три гипотезы

- h_1 - здесь угольное месторождение;
- h_2 - здесь железорудное месторождение;
- h_3 - похоже, здесь ничего нет,

связанные с двумя свидетельствами, утверждающими, что в данной точке имеется:

- e_1 - аномалия вызванной поляризации;
- e_2 - выход на поверхность углистых сланцев.

Пусть все гипотезы равновероятны: $P(h_1) = P(h_2) = P(h_3) = 0.01$.

Зададим условные вероятности $P(e_j|h_i)$. Допустим, на основании статистических данных известно, что аномалия вызванной поляризации вблизи угольного месторождения наблюдается в 98% случаев, вблизи железорудного месторождения — в 97% , а при отсутствии этих месторождений — лишь в 15% случаев. Аналогично, пусть выход на поверхность углистых сланцев в 85% случаев говорит о присутствии угольного месторождения, в 10% — железорудного месторождения и маловероятен (5%), если никаких месторождений нет.

Определим апостериорные вероятности гипотез при подтверждении обоих свидетельств.

$$P(h_1|e_1e_2) = \frac{P(h_1)P(e_1|h_1)P(e_2|h_1)}{P(h_1)P(e_1|h_1)P(e_2|h_1) + P(h_2)P(e_1|h_2)P(e_2|h_2) + P(h_3)P(e_1|h_3)P(e_2|h_3)} =$$

$$= \frac{0,01 \cdot 0,98 \cdot 0,85}{0,01 \cdot 0,98 \cdot 0,85 + 0,01 \cdot 0,97 \cdot 0,1 + 0,01 \cdot 0,15 \cdot 0,05} = 0,888.$$

Аналогично получим $P(h_2|e_1e_2) = 0,104$ и $P(h_3|e_1e_2) = 0,008$.

Если выход на поверхность углистых сланцев не подтверждается, необходимо вычислить апостериорные вероятности гипотез по формуле

$$P(h_i|e_1\bar{e}_2) = \frac{P(h_i)P(e_1|h_i)P(\bar{e}_2|h_i)}{P(h_1)P(e_1|h_1)P(\bar{e}_2|h_1) + P(h_2)P(e_1|h_2)P(\bar{e}_2|h_2) + P(h_3)P(e_1|h_3)P(\bar{e}_2|h_3)}$$

Учитывая, что отсутствие события e_j есть событие \bar{e}_j , а $P(\bar{e}_j|h_i) = 1 - P(e_j|h_i)$, получим

$$P(h_1|e_1\bar{e}_2) = \frac{0,01 \cdot 0,98 \cdot 0,15}{0,01 \cdot 0,98 \cdot 0,15 + 0,01 \cdot 0,97 \cdot 0,9 + 0,01 \cdot 0,15 \cdot 0,95} = 0,126$$

$$P(h_2|e_1\bar{e}_2) = 0,751; \quad P(h_3|e_1\bar{e}_2) = 0,123.$$

Вычислим подобным образом $P(h_i|\bar{e}_1e_2)$, $P(h_i|\bar{e}_1\bar{e}_2)$ и сведем все данные в диагностическую таблицу (табл. 1). (Полученный результат, как это видно из формул, не зависит от значений апостериорных вероятностей гипотез, важно лишь, чтобы было справедливо наше допущение о том, что они равновероятны.) Из расчетов следует, что при наличии обоих свидетельств существенно увеличивается вероятность угольного

месторождения. При наличии только аномалии вызванной поляризацией повышается вероятность железорудного месторождения, а при отсутствии свидетельств вероятности месторождений резко уменьшаются. При отсутствии e_1 и наличии e_2 диагноз системы остается неопределенным, и, как будет показано ниже, этот столбец из диагностической таблицы вообще может быть удален.

h_i	$P(e_1 h_i)$	$P(e_2 h_i)$	$P(h_i e_1e_2)$	$P(h_i e_1\bar{e}_2)$	$P(h_i \bar{e}_1e_2)$	$P(h_i \bar{e}_1\bar{e}_2)$
h_1	0.98	0,85	0,888	0,126	0,272	0,004
h_2	0.97	0,10	0,104	0,751	0,048	0,032
h_3	0.15	0,05	0,008	0,123	0,680	0,964

Табл. 1. Вероятности свидетельств и апостериорные вероятности гипотез

При автоматическом применении этих свидетельств (среди множества других геофизических характеристик) к множеству точек исследуемой территории можно построить изолинии распределения предполагаемых месторождений на данной территории. Скорость современных компьютеров зачастую позволяет не заботиться о качестве алгоритмов, если кроме быстрого действия алгоритма не затронуты другие его характеристики. Вот и в нашем случае изъян алгоритма проявляется только при последовательной обработке *связанных* свидетельств в диалоговом режиме. Как же проявляется этот изъян?

СВЯЗАННЫЕ СВИДЕТЕЛЬСТВА

Дело в том, что давая ответ на одно из связанных свидетельств, мы тем самым уже подтверждаем или опровергаем (в зависимости от типа связи) коррелированный с ним факт. Или несколько фактов. Например, для медицинской ЭС подтверждение нормального давления у пациента автоматически должно приводить к опровержению свидетельств, касающихся повышенного и пониженного артериального давлений. А это предполагает совершение над базой знаний некоторых действий, отличных от вероятностного вывода. Другими словами, в случае связанных свидетельств необходимо сначала обработать все их взаимосвязи и лишь затем перейти к вычислениям по Байесу. В противном случае поведение системы вряд ли можно будет назвать разумным.

Производимые машиной вывода действия также зависят от типа связей между фактами. Так, можно выделить свидетельства, находящиеся между собой во временных отношениях: *раньше — позже, старше — младше* и т.п. Сюда же можно добавить свидетельства, входящие в *причинно-следственные* отношения или состоящие в связях типа *обобщение — конкретизация, агрегация — детализация* и пр.

Возвращаясь к примеру, заметим, что одной из причин аномалии вызванной поляризацией (e_1) может служить выход на поверхность углистых сланцев (e_2), т.е. наличие причинно-следственная связь $e_2 \Rightarrow e_1$. Описанный алгоритм вероятностного вывода экспертной системы не учитывает связи между свидетельствами, поэтому утвердительный ответ на свидетельство e_2 не оградит систему от задания вопроса по факту e_1 . Совершенно не обязательно, что это произойдет на следующей итерации, но в любом случае вызовет недоумение пользователя, ведь ответ на этот вопрос (утвердительный!) он, по сути, уже дал. Еще большее недоумение и нервозность может вызвать вопрос системы относительно углистых сланцев при недавно полученном отрицательном ответе на свидетельство e_1 : если нет аномалии вызванной поляризацией, значит, нет никаких причин ее вызывающих, в том числе и выхода на поверхность углистых сланцев.

Чтобы этого не происходило, принято решение принудительно назначить всем свидетельствам-причинам нулевой вес, и система в этом случае никогда не задаст вопрос о причине до того, как обработает следствие. Причинно-следственную связь можно

запрограммировать как импликацию *посылка* \Rightarrow *следствие*, тогда связь $e_2 \Rightarrow e_1$ следует рассматривать как совокупность четырех групп совместных и несовместных событий:

1. $e_1 e_2$ - совместные события, эквивалентны e_2 ;
2. $\overline{e_1} e_2$ - совместные события, требуется анализ;
3. $\overline{e_1} \overline{e_2}$ - противоречие, несовместные события;
4. $e_1 \overline{e_2}$ - совместные события, эквивалентны $\overline{e_1}$.

Таким образом, если ответ на свидетельство-следствие e_1 будет отрицательным (группы 3 и 4), свидетельство-причина e_2 будет просто забыто, а вероятности всех гипотез пересчитаются в соответствии с вероятностными характеристиками своих связей со свидетельством e_1 . И наоборот, если факт e_1 будет подтвержден, в базе знаний не будет произведено никаких изменений за исключением присвоения свидетельству e_2 веса свидетельства e_1 . Машине вывода ничего не остается, как тут же выбрать событие e_2 и задать пользователю соответствующий вопрос (группы 1 и 2).

Возможно, что одно и то же свидетельство-следствие порождается несколькими свидетельствами-причинами. Так, например, аномалия вызванной поляризации e_1 может также порождаться фактом прохождения вблизи исследуемой точки какого-либо трубопровода. Поведение ЭС ничем не будет отличаться от только что описанного: если не подтверждается аномалия вызванной поляризации, то ни одна из упомянутых причин не рассматривается. В противном случае веса обеих причин резко возрастают, что говорит о том, что они непременно будут обработаны, причем порядок их обработки уже не имеет значения. Также возможно, что один и тот же факт может служить причиной появления нескольких свидетельств-следствий, и только подтверждение их всех приведет к обработке причинного для них свидетельства. В общем случае, когда любое свидетельство-следствие

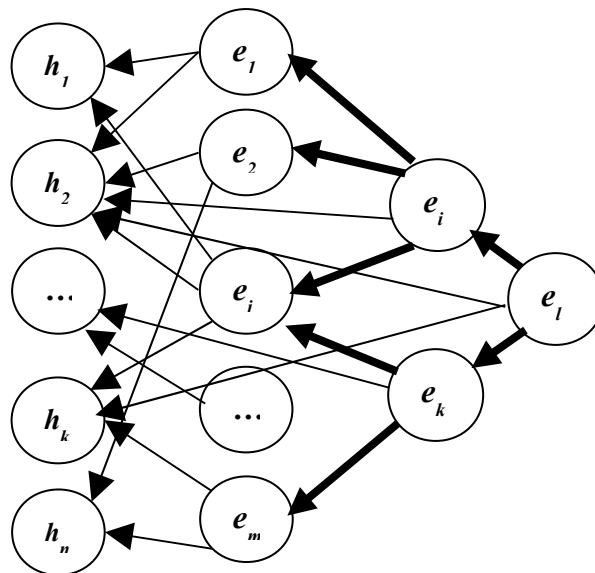


Рис. 2. Пример сети с учетом взаимосвязей между фактами.

может являться причиной других свидетельств, Байесовская сеть может выглядеть так, как показано на Рис.2. Жирными стрелками обозначены причинно-следственные связи между узлами-свидетельствами. Причем в отличие от связей факт-гипотеза никакие вероятностные характеристики этих связей в базе знаний не хранятся: достаточно в описании свидетельства-следствия иметь список ссылок на свидетельства-причины, в описании

которых по этому поводу установлен соответствующий флаг.

Несмотря на усложнение Байесовской сети, время реакции системы на очередное свидетельство не увеличилось, так как обработка в среднем половины связанных свидетельств происходит без привлечения формулы Байеса (см. выше), и пересчет апостериорной вероятности гипотез и новых весов свидетельств не производится. Для выяснения степени влияния свидетельств, составляющих причинно-следственные связи, на конечный результат работы системы необходимо провести дополнительный анализ. Если выяснится, что основное воздействие на вероятности гипотез оказывают свидетельства-причины, то возможно еще более упростить обработку свидетельств-следствий, искусственно сведя к минимуму их влияние на гипотезы и оставив им роль переключателей с состояниями *да/нет*, а неопределенные ответы разрешить только для свидетельств-причин.

Во всяком случае, к числу положительных моментов в работе ЭС при введении причинно-следственных связей между свидетельствами следует отнести изменение поведения системы, ставшего более «разумным» и не ставящим в тупик оператора ЭС. Что же касается автоматического режима работы ЭС, то формирование колонок, представляющих несовместные события (группа 3), в диагностической таблице не производится, что также ускоряет работу системы.

Литература

1. Биргер И.А. Техническая диагностика. — М.: «Машиностроение», 1978. — 240 с.