УДК 004.8

А. Л. Тулупьев

ОЦЕНКА ЧУВСТВИТЕЛЬНОСТИ РЕЗУЛЬТАТА ЛОКАЛЬНОГО АПРИОРНОГО ЛОГИКО-ВЕРОЯТНОСТНОГО ВЫВОДА В ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Рассматривается способ оценки зависимости вариации результатов (чувствительности) априорного вывода от допустимой вариации исходных данных во фрагменте знаний алгебраической байесовской сети. Предлагаемый способ основан на решении совокупности задач линейного программирования. Получена верхняя оценка характеристики чувствительности, линейно зависящая от радиуса вариации исходных данных.

Ключевые слова: *оценка чувствительности, логико-вероятностный вывод, алгебраические байесовские сети.*

В базах знаний интеллектуальных систем вследствие определенных причин [1, 2] накапливаются не только достоверные знания, но также и знания, отличающиеся неопределенностью, неточностью, нечеткостью. Эти знания нельзя исключить, поскольку перечисленные „отрицательные“ факторы не означают абсолютного отсутствия сведений о предметной области, а лишь характеризуют неполноту доступных знаний о ней [3]. Зачастую даже в условиях информационного дефицита [4] (т.е. неполноты знаний и/или их неопределенности) можно принять обоснованное решение, учтя все доступные сведения.

Представить неопределенность знаний можно различными способами — например, приписать утверждению меру его истинности, в качестве которой будем рассматривать вероятностную меру. Небольшие совокупности утверждений с вероятностными оценками их истинности могут быть организованы во фрагменты знаний (ФЗ). Набор фрагментов знаний, в свою очередь, образует базу фрагментов знаний с неопределенностью. Представлением этой базы для разработки как структур данных в программном коде, так и самого кода являются, в частности, алгебраические байесовские сети (АБС) [2, 5, 6]. Предложенные в работах [7, 8] алгоритмы позволяют проверять непротиворечивость АБС (поддержание непротиворечивости), выводить новые оценки истинности утверждений на основе известных (априорный вывод), учитывать влияние поступивших свидетельств на вероятность истинности элементов АБС (апостериорный вывод).

Цель настоящей статьи — исследовать чувствительность априорного вывода локально, т.е. когда он осуществляется в отдельном фрагменте знаний. Под чувствительностью понимается зависимость результата вывода от точности указания исходных данных. Исследовать чувствительность — значит оценить характеристики зависимости вариации результата от вариации исходных данных. Для каждой предметной области известны конкретные определения указанных вариаций и зависимости между ними, а также сформированы наборы характеристик последних.

Обозначения. Согласно работам [2, 5—8] введем следующие обозначения:

$\mathbf{P}$ — вектор вероятностей истинности элементов фрагмента знаний, упорядоченных специальным образом [5];

$\mathbf{P}_\circ$ — вектор скалярных (точечных) оценок этих вероятностей, здесь индекс „$\circ$“ указывает на использование исходных данных;

$\hat{\mathbf{P}}$ — вектор проварьированных вероятностей истинности элементов фрагмента знаний, а $\mathbf{\Delta} = \hat{\mathbf{P}} - \mathbf{P}$ — собственно вариация исходных данных, которая предполагается допустимой, т.е. проварьированные вероятности непротиворечивы: $\mathbf{I}\hat{\mathbf{P}} \geq \mathbf{0}$;

f — утверждение, составленное из атомарных утверждений, вошедших во фрагмент знаний; в силу этого вероятность f выражается в виде скалярного произведения через вероятности истинности элементов ФЗ: $p = p(f) = \mathbf{L} \cdot \mathbf{P}$; с учетом вышеописанной допустимой вариации изменившаяся оценка вероятности f будет иметь вид $\hat{p} = \hat{p}(f) = \mathbf{L} \cdot \hat{\mathbf{P}}$.

Для соблюдения требования аксиоматики вероятностной логики должна выполняться система неравенств, записанная в векторном виде: $\mathbf{IP} \geq \mathbf{0}$, матрица $\mathbf{I} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{[n]}$ — степень Кронекера [5], здесь n — число атомарных утверждений, вошедших во фрагмент знаний. Предполагается, что набор исходных данных непротиворечив: $\mathbf{IP}_\circ \geq \mathbf{0}$.

Для измерения вариаций исходных данных и результата необходимы две метрики: метрика r, аргументы которой есть векторы вероятностей истинности элементов ФЗ: $r(\hat{\mathbf{P}}, \mathbf{P})$; метрика d, аргументы которой — оценки вероятности истинности утверждения $f : d(\hat{p}, p)$.

Постановка задачи. Пусть известно положительное число — радиус вариации $\delta > 0$, которое ограничивает вариацию исходных данных: $r(\hat{\mathbf{P}}, \mathbf{P}) \leq \delta$. Требуется с учетом данных и ограничений, перечисленных выше, исследовать величину $d(\hat{p}, p)$: либо определить ее функциональную зависимость $d(\hat{p}, p) = V(\hat{\mathbf{P}}, \mathbf{P}, \delta, \ldots)$, либо найти ее верхнюю границу (точную верхнюю границу, максимум) ε. При этом процесс вычислений должен опираться на хорошо изученные и реализованные алгоритмы.

При анализе чувствительности особую значимость имеет выбор метрик: именно им определяется сложность, разрешимость и „алгоритмизуемость“ возникающих в процессе исследования экстремальных задач. В рассматриваемом случае предлагается выбрать следующие метрики: $r(\hat{\mathbf{P}}, \mathbf{P}) = \max\limits_{i=0,\, \ldots,\, 2^n-1} \left| \hat{p}_i - p_i \right|$ и $d(\hat{p}, p) = \left| \hat{p} - p \right|$. Использование таких метрик, как будет видно в дальнейшем, приведет к решению задач линейного программирования (ЗЛП); евклидова же метрика, например, потребовала бы решения задач с более сложным множеством квадратичных ограничений.

Предлагаемый подход к решению задачи. Рассмотрим экстремальные задачи, которые позволяют вычислить характеристики чувствительности результата априорного вывода p к допустимой вариации исходных данных. Вычислим величины

$$\varepsilon = \sup_{\substack{\mathbf{I}\hat{\mathbf{P}} \geq \mathbf{0}, \mathbf{IP} \geq \mathbf{0}, \\ r(\hat{\mathbf{P}}, \mathbf{P}) \leq \delta}} \left\{ d(\hat{p}, p) \right\} \text{ и } \varepsilon(\mathbf{P}_\circ) = \sup_{\substack{\mathbf{I}\hat{\mathbf{P}} \geq \mathbf{0},\, \mathbf{IP} \geq \mathbf{0}, \\ r(\hat{\mathbf{P}}, \mathbf{P}) \leq \delta,\, \mathbf{P} = \mathbf{P}_\circ}} \left\{ d(\hat{p}, p) \right\},$$

при этом величина $\varepsilon(\mathbf{P}_\circ)$ необходима для исследования ситуации в случае с конкретным набором исходных оценок $\mathbf{P} = \mathbf{P}_\circ$.

В первую очередь, сведем ограничения экстремальных задач к набору линейных равенств и неравенств.

Очевидно, что ограничения $\mathbf{I}\hat{\mathbf{P}} \geq \mathbf{0}$, $\mathbf{IP} \geq \mathbf{0}$ и $\mathbf{P} = \mathbf{P}_\circ$ изначально линейны, их изменять не требуется. Неравенство $r(\hat{\mathbf{P}}, \mathbf{P}) \leq \delta$ раскрывается как $\max\limits_{i=0,\,...,\,2^n-1} \left|\hat{p}_i - p_i\right| \leq \delta$, что эквивалентно системе линейных неравенств $\left\{-\delta \leq \hat{p}_i - p_i \leq \delta\right\}_{i=0}^{i=2^n-1}$. Таким образом, все ограничения в рассмотренных экстремальных задачах линейные.

По определению целевая функция $d(\hat{p}, p) = \left|\hat{p} - p\right|$. Чтобы свести экстремальную задачу к двум ЗЛП, достаточно решить последние отдельно для разности $\hat{p} - p$ и разности $p - \hat{p}$.

Таким образом, процесс оценивания характеристик чувствительности сводится к решению совокупности задач линейного программирования:

$$\varepsilon = \max_{\substack{\mathbf{I}\hat{\mathbf{P}}\geq\mathbf{0},\mathbf{IP}\geq\mathbf{0},\\ r(\hat{\mathbf{P}},\mathbf{P})\leq\delta}} \left\{\hat{p} - p, p - \hat{p}\right\}, \quad \varepsilon(\mathbf{P}_\circ) = \max_{\substack{\mathbf{I}\hat{\mathbf{P}}\geq\mathbf{0},\ \mathbf{IP}\geq\mathbf{0},\\ r(\hat{\mathbf{P}},\mathbf{P})\leq\delta,\ \mathbf{P}=\mathbf{P}_\circ}} \left\{\hat{p} - p, p - \hat{p}\right\}.$$

Алгоритмы решения ЗЛП известны, хорошо изучены, а также реализованы в библиотеках, рассчитанных на ряд систем программирования.

Верхняя оценка. Можно оценить характеристики чувствительности и не прибегая к решению ЗЛП, пусть и несколько потеряв в точности оценки. В соответствии с полученной оценкой можно предъявить обоснованные требования к точности измерения, определения или расчета исходных данных.

Учитывая, что $\hat{p}_0 = p_0 = 1$ [2, 5], оценим

$$d(\hat{p}, p) = \left|\hat{p} - p\right| = \left|\mathbf{L}\cdot\hat{\mathbf{P}} - \mathbf{L}\cdot\mathbf{P}\right| = \left|\mathbf{L}\cdot(\hat{\mathbf{P}} - \mathbf{P})\right| =$$

$$= \left|\sum_{i=0}^{2^n-1} l_i\left(\hat{p}_i - p_i\right)\right| \leq \sum_{i=1}^{2^n-1}\left|l_i\right|\cdot\underbrace{\left|\left(\hat{p}_i - p_i\right)\right|}_{\leq\delta} \leq \sum_{i=1}^{2^n-1}\left|l_i\right|\delta = \delta\sum_{i=1}^{2^n-1}\left|l_i\right|.$$

Таким образом, установлена легко вычисляемая верхняя граница для оценки чувствительности — $\varepsilon \leq \delta\sum\limits_{i=1}^{2^n-1}\left|l_i\right|$. Вопрос о том, точна ли эта верхняя граница, остается открытым.

Некоторые результаты, представленные в данной статье, получены в рамках проекта, поддержанного Российским фондом фундаментальных исследований (грант № 09-01-00861-а).

СПИСОК ЛИТЕРАТУРЫ

1. *Крейнович В. Я., Нгуен Т. Х., Городецкий В. И.* и др. Применение интервальных степеней доверия: аналитический обзор // Информационные технологии и интеллектуальные методы: Сб. трудов СПИИРАН. СПб.: СПИИРАН, 1999. Вып. 3. С. 6—61.

2. *Тулупьев А. Л., Николенко С. И., Сироткин А. В.* Байесовские сети: логико-вероятностный подход. СПб.: Наука, 2006. 607 с.

3. *Нариньяни А. С.* НЕ-факторы: неточность и недоопределенность — различие и взаимосвязь // Изв. РАН. Сер. Теория и системы управления. 2000. № 5. С. 44—56.

4. *Хованов Н. В.* Анализ и синтез показателей при информационном дефиците. СПб.: Изд-во С.-Петербург . ун-та, 1996. 196 с.

5. *Тулупьев А. Л.* Алгебраические байесовские сети: локальный логико-вероятностный вывод: Учеб. пособие. СПб.: СПбГУ — ООО Изд-во „Анатолия“, 2007. 80 с.

6. *Тулупьев А. Л.* Алгебраические байесовские сети: глобальный логико-вероятностный вывод в деревьях смежности: Учеб. пособие. СПб.: СПбГУ — ООО Изд-во „Анатолия“, 2007. 40 с.

7. *Тулупьев А. Л., Сироткин А. В., Николенко С. И.* Синтез согласованных оценок истинности утверждений в интеллектуальных информационных системах // Изв. вузов. Приборостроение. 2006. Т. 49, № 7. С. 20—26.

8. *Тулупьев А. Л., Николенко С. И., Никитин Д. А., Сироткин А. В.* Синтез апостериорных оценок истинности суждений в интегрированных базах знаний: детерминированный вариант // Там же. 2006. Т. 49, № 11. С. 35—39.

Сведения об авторе

Александр Львович Тулупьев — канд. физ.-мат. наук, доцент; Санкт-Петербургский институт информатики и автоматизации РАН, лаборатория прикладной информатики; E-mail: alt@iias.spb.su

Рекомендована кафедрой
технологий программирования
СПбГУ ИТМО

Поступила в редакцию
15.02.08 г.