УСТОЙЧИВОСТЬ И МНОЖЕСТВЕННАЯ УСТОЙЧИВОСТЬ ГЛОБАЛЬНОЙ НЕПРОТИВОРЕЧИВОСТИ АЛГЕБРАИЧЕСКИХ БАЙЕСОВСКИХ СЕТЕЙ

А.В. Сироткин[1], А.Л. Тулупьев[2], С.И. Николенко[3]

[1–3]Санкт-Петербургский государственный университет,
198904, Санкт-Петербург, Библиотечная площадь, д. 2

[1]<sirotkin@hotbox.ru>; [3]<sergey@logic.pdmi.ras.ru>

[2]Санкт-Петербургский институт информатики и автоматизации РАН
199178, Санкт-Петербург, 14-я линия ВО, д. 39

[2]<alt@iias.spb.su>

УДК 681.3
А.В. Сироткин, А.Л. Тулупьев, С.И. Николенко. **Устойчивость и множественная устойчивость глобальной непротиворечивости алгебраических байесовских сетей** // Труды СПИИРАН. Вып. 2, т. 2. — СПб.: Наука, 2005.
Аннотация. *Вводятся понятия устойчивости глобальной непротиворечивости и множественной устойчивости глобальной непротиворечивости алгебраической байесовской сети, и характеризуется взаимная зависимость их количественных характеристик.* — Библ. 5 назв.

UDC 681.3
A.V. Sirotkin, A.L. Tulupyev, S.I. Nikolenko. **Stability and Multi-stability of Algebraic Bayesian Network's Global Consistency** // SPIIRAS Proceedings. Issue 2, vol. 2. — SPb.: Nauka, 2005.
Abstract. *We state definitions for global consistency stability and global consistency multi-stability in an algebraic Bayesian network and consider the relationships between the stabilities quantitative measures.* — Bibl. 5 items.

1. Введение

Алгебраические байесовские сети (АБС) [1, 2], непротиворечивость которых рассматривается в данной работе, являются одной из моделей баз знаний (БЗ) с вероятностной степенью неопределенности. Они состоят из фрагментов знаний (ФЗ). Математической моделью ФЗ АБС является идеал цепочек конъюнкций, построенный над заданным множеством атомарных пропозициональных формул. Каждый элемент идеала имеет точечную или интервальную оценку вероятности своей истинности. Такая конструкция позволяет представить неопределенность знаний, используя уже хорошо развитый аппарат теории вероятностей.

С одной стороны, аксиоматика теории вероятностей накладывает ряд ограничений на возможные значения оценок истинности, поэтому после того, как элементам идеала цепочек конъюнкций назначены какие-то начальные оценки вероятности истинности, необходимо проверить непротиворечивость (совместность) полученного набора оценок. С другой стороны, даже если такой набор окажется непротиворечивым, исходные оценки вероятности истинности могут быть уточнены за счет взаимного влияния, а также ограничений, наложенных аксиоматикой теории вероятностей ([1, 3, 4]). Процесс проверки непротиворечивости и уточнения исходных оценок называется *поддержанием непротиворечивости* [распределения оценок вероятности истинности элементов] идеала конъюнкций, а процесс уточнения оценок вероятности истинности для заданной пропозициональной формулы – *априорным выводом.*

В настоящей работе мы обратимся к вопросам *устойчивости* состояния глобальной непротиворечивости, т.е. к тому, как изменяются результаты при заданном колебании начальных данных. Знания об устойчивости того или иного вычислительного процесса имеют прямое практическое значение. Если колебания результата велики по сравнению с вариацией исходных данных, то для получения результата с приемлемой точностью приходится требовать исходных данных с относительно более высокой степенью точности. В случае же алгебраических байесовских сетей, где исходные данные задаются экспертным путём, уточнение их может потребовать больших затрат, а может быть и просто невозможно.

2.Основные обозначения

Рассмотрим множество атомарных пропозициональных формул $A=\{x_1,\ldots,x_n\}$. Введем обозначение аргументного места (литерала) $\tilde{x}$. Аргументное место может принимать одно из двух значений $\tilde{x}\in\{x,\overline{x}\}$, где $x\in A$, а $\overline{x}$ — его логическое отрицание. Набор всех возможных пропозициональных формул над A обозначим $F=F(A)$, а набор *квантов* — цепочек конъюнкций аргументных мест максимальной длины со всеми возможными означиваниями — обозначим как $Q=Q(A)=\{\tilde{x}_1\tilde{x}_2\ldots\tilde{x}_n\}$. Введём функцию $S:F\to 2^Q$, $S(f)=S_f$, где S_f — множество квантов, участвующих в СДНФ пропозициональной формулы f. *Идеалом* цепочек конъюнкций над заданным множеством атомарных формул $A'\subseteq A$ назовем все непустые положительно означенные цепочки конъюнкций атомарных пропозициональных формул:

$$C=C(A')=\{x_{i_1}x_{i_2}\ldots x_{i_k}\mid 1\le i_1,i_2,\ldots,i_k\le m\}$$

Рассмотрим Q как множество элементарных событий. Зададим на нем исходное распределение вероятностей $p^\circ:Q\to[0;1]$, такое, что $\forall q\in Q\ p^\circ(q)\ge 0$, $\sum_{q\in Q}p^\circ(q)=1$. Введем вероятность $p:2^Q\to[0;1]$ следующим образом: $\forall(S\subseteq Q)\ \ p(S)=\sum_{q\in S}p^\circ(q)$.

На этом шаге мы получили вероятностное пространство $\left\langle Q,2^Q,p\right\rangle$. Определим вероятностную меру на множестве F следующим образом: $\forall(f\in F)\ p(f)=p(S(f))$. Взяв за основу такое распространение вероятности на множество пропозициональных формул, мы получим новое вероятностное пространство: $\langle Q,F,p\rangle$. Эта структура вероятностного пространства по Н. Нильссону задает вероятность над всеми пропозициональными формулами, построенными над множеством атомарных формул A.

Пусть над элементами идеала положительно означенных цепочек конъюнкций C задано точечное распределение вероятности. Назовем это распределение непротиворечивым, если на соответствующем наборе квантов Q можно подобрать такое распределение вероятностей, которое отвечает аксиоматике вероятностей, а также задает те же вероятности элементов идеала цепочек конъюнкций. То есть, точечное распределение вероятностей на идеале цепочек конъюнкций непротиворечиво тогда и только тогда, когда на квантах можно по-

Труды СПИИРАН. 2005. Вып. 2. Т. 2. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2005. Issue 2. V. 2. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

добрать такое распределение вероятностей, которое индуцирует исходное на элементах идеала.

Множество ограничений, налагаемых на эти вероятности, состоит из двух частей: ограничений, вытекающих из аксиоматики теории вероятностей, и ограничений, налагаемых собственно предметной областью.

В дальнейшем множество ограничений, вытекающих из требований аксиоматики теории вероятностей относительно элементов идеала над n-элементным набором атомарных пропозиций, будем обозначать E^n (например, E^2 содержит ограничения $p(x_1x_2) \ge 0$, $p(x_1) - p(x_1x_2) \ge 0$, $p(x_2) - p(x_1x_2) \ge 0$, $1 - p(x_1) - p(x_2) + p(x_1x_2) \ge 0$). Соответствующий набор ограничений из предметной области будем обозначать D^n (например, D^2 содержит $p_0^-(x_1) \le p(x_1) \le p_0^+(x_1)$, $p_0^-(x_2) \le p(x_2) \le p_0^+(x_2)$, $p_0^-(x_1x_2) \le p(x_1x_2) \le p_0^+(x_1x_2)$, а в случае расширенного фрагмента знаний, возможно, содержит и другие линейные ограничения). Объединение множеств ограничений обозначается $R^n = E^n \cup D^n$.

Мы говорим, что АБС непротиворечива, если существует точечное распределение вероятностей над пропозициональными формулами $p : C \to [0;1]$, удовлетворяющее множеству ограничений R^n. Вопрос о непротиворечивости АБС сводится к решению задачи линейного программирования [4].

3. Устойчивость

Термин «устойчивость», вообще говоря, означает чувствительность результата к колебаниям начальных данных. В случае алгебраических байесовских сетей можно выделить несколько различных понятий, каждое из которых заслуживает названия «устойчивость»:

- устойчивость непротиворечивости — устойчивость бинарного ответа «да/нет» относительно колебаний начальных данных;
- устойчивость априорного вывода — устойчивость априорных оценок вероятностей одних конъюнктов алгебраической байесовской сети относительно колебаний оценок других конъюнктов;
- устойчивость апостериорного вывода — устойчивость результата процесса учёта поступающих свидетельств относительно колебаний исходных оценок.

В настоящей работе мы ограничимся рассмотрением первого аспекта устойчивости. Устойчивость априорного вывода частично рассмотрена в [5]**.** Рассмотрение устойчивости апостериорного вывода остаётся интересной открытой проблемой.

4. Устойчивость непротиворечивости

Дадим формализацию понятия устойчивости поддержания непротиворечивости. Можно рассмотреть два естественных подхода к определению этого понятия.

Назовём алгебраическую байесовскую сеть N на множестве переменных $A = \{x_1, \ldots, x_n\}$ *устойчиво непротиворечивой* с *показателем устойчивости* $\varepsilon > 0$, если она непротиворечива, и для всякого конъюнкта

$$x_{i_1} x_{i_2} \ldots x_{i_k} \in N$$

и всяких

$$-\varepsilon \le \delta^l, \delta^u \le \varepsilon$$

байесовская сеть N', множество ограничений которой вместо ограничения

$$p_0^-(x_{i_1} x_{i_2} \ldots x_{i_k}) \le p(x_{i_1} x_{i_2} \ldots x_{i_k}) \le p_0^+(x_{i_1} x_{i_2} \ldots x_{i_k})$$

содержит ограничение

$$p_0^-(x_{i_1} x_{i_2} \ldots x_{i_k}) + \delta^l \le p(x_{i_1} x_{i_2} \ldots x_{i_k}) \le p_0^+(x_{i_1} x_{i_2} \ldots x_{i_k}) + \delta^u,$$

также непротиворечива.

Иными словами, алгебраическая байесовская сеть устойчиво непротиворечива, если она остается непротиворечивой при небольших колебаниях одного из начальных данных. Другим естественным подходом в данном случае будет следующее определение:

Назовём алгебраическую байесовскую сеть N на множестве переменных $A = \{x_1, \ldots, x_n\}$ *множественно устойчиво непротиворечивой* с *показателем устойчивости* $\varepsilon > 0$, если она непротиворечива, и для всякого набора конъюнктов

$$x_{i_{11}} x_{i_{12}} \ldots x_{i_{1k_1}}, x_{i_{21}} x_{i_{22}} \ldots x_{i_{2k_2}}, \ldots, x_{i_{m1}} x_{i_{m2}} \ldots x_{i_{mk_m}} \in N$$

и всяких

$$-\varepsilon \le \delta_1^l, \delta_1^u, \delta_2^l, \delta_2^u, \ldots, \delta_m^l, \delta_m^u \le \varepsilon$$

байесовская сеть N', множество ограничений которой вместо ограничений вида

$$p_0^-(x_{i_{l1}} x_{i_{l2}} \ldots x_{i_{lk_l}}) \le p(x_{i_{l1}} x_{i_{l2}} \ldots x_{i_{lk_l}}) \le p_0^+(x_{i_{l1}} x_{i_{l2}} \ldots x_{i_{lk_l}})$$

содержит ограничения

$$p_0^-(x_{i_{l1}} x_{i_{l2}} \ldots x_{i_{lk_l}}) + \delta_l^l \le p(x_{i_{l1}} x_{i_{l2}} \ldots x_{i_{lk_l}}) \le p_0^+(x_{i_{l1}} x_{i_{l2}} \ldots x_{i_{lk_l}}) + \delta_l^u,$$

также непротиворечива.

Иными словами, алгебраическая байесовская сеть множественно устойчиво непротиворечива, если она остается непротиворечивой при одновременных небольших колебаний любой совокупности начальных данных.

Следующий пример показывает, что эти устойчивости могут иметь на одной и той же АБС разные показатели.

Пример. Рассмотрим АБС на фрагменте знания второго порядка со следующими ограничениями: $0.2 \le p(x_1 x_2) \le 0.6$, $0.6 \le p(x_1) \le 1.0$, $0.6 \le p(x_2) \le 1.0$. В этом случае сеть будет устойчиво непротиворечива с показателем 0.2, но не будет множественно устойчиво непротиворечива с показателем 0.2, так как в этом случае оценки $0.4 \le p(x_1 x_2) \le 0.4$, $0.8 \le p(x_1) \le 0.8$, $0.8 \le p(x_2) \le 0.8$, соответствуют сдвигам границ отрезков на 0.2, но данная сеть противоречива.

Вопрос же о соотношении этих понятий для расширенных фрагментов знаний пока остаётся открытым.

Следующие две теоремы показывают соотношения между показателями двух типов устойчивости.

Труды СПИИРАН. 2005. Вып. 2. Т. 2. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2005. Issue 2. V. 2. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

Теорема. Устойчиво непротиворечивая алгебраическая байесовская сеть с показателем устойчивости ε является множественно устойчиво непротиворечивой с показателем устойчивости $\frac{\varepsilon}{2^n-1}$.

Доказательство. Если АБС устойчиво непротиворечиво, то существует такое распределение, что вероятность заданного конъюнкта $X = x_{i_1}x_{i_2}\ldots x_{i_k}$ лежит в интервале $[p_0^- + \varepsilon, p_0^+ - \varepsilon]$, а остальные — в исходных интервалах. Построим распределения $p_X : C \to [0,1]$ для каждого конъюнкта. Теперь задача сводится к тому, чтобы построить новое распределение p, для которого уже все вероятности можно отделить от границ исходных интервалов одновременно.

Определим $p(X) = \sum_{i \in C} p_i(X)/(2^n - 1)$. Очевидно, что

$$p(X) \in [p_0^-(X) + \varepsilon/(2^n-1), p_0^+(X) - \varepsilon/(2^n-1)].$$

Покажем, что полученное распределение непротиворечиво. Выберем произвольное ограничение из E^n. Выпишем теперь это ограничение, для каждого распределения p_X. Просуммируем выписанные неравенства и разделим на их количество $2^n - 1$. Мы получим верное неравенство, в точности говорящее, что построенное нами распределение p удовлетворяет выбранному ограничению. А так как ограничения выбрано произвольно, то p удовлетворяет всем ограничениям. Это и значит, что распределение p непротиворечиво. □

В общем случае доказать лучший результат невозможно, так как существует следующий пример.

Пример. Рассмотрим фрагмент знаний порядка n со следующими ограничениями:

$\varepsilon^k \le p(x_{i_1}x_{i_2}\ldots x_{i_k}) \le \varepsilon^k + 3(1-\varepsilon)^n$, для нечетного k

$\varepsilon^k - 3(1-\varepsilon)^n \le p(x_{i_1}x_{i_2}\ldots x_{i_k}) \le \varepsilon^k$, для четного k

где ε выбран так, что $\varepsilon^{k-1} - \varepsilon^k \ge (1-\varepsilon)^n$, $\varepsilon^k - 3(1-\varepsilon)^n \ge 0$ и $\varepsilon^k + 3(1-\varepsilon)^n \le 1$ для любого k.

Указанный ФЗ является устойчиво непротиворечивым с показателем $(1-\varepsilon)^n$, и при этом максимальный показатель множественной устойчивой непротиворечивости в точности $\frac{(1-\varepsilon)^n}{2^n-1}$.

Теорема. Пусть АБС множественно устойчиво непротиворечива с показателем ε. Тогда, если все интервалы оценок истинности конъюнктов этой АБС имеют длину не менее 4ε, то данная АБС устойчиво непротиворечива с показателем 2ε.

Доказательство. Рассмотрим произвольный конъюнкт $X = x_{i_1}x_{i_2}\ldots x_{i_k}$ и покажем, что существует распределение p такое, что

$$p_0^-(X)+2\varepsilon \le p(X) \le p_0^+(X)-2\varepsilon \text{ и } p_0^-(Y) \le p(Y) \le p_0^+(Y)$$

при $Y \ne X$. Так как наша АБС множественно устойчиво непротиворечива, то существует распределение p_ε такое, что $p_0^-(Y)+\varepsilon \le p_\varepsilon(Y) \le p_0^+(Y)-\varepsilon$ для всех $Y \in C$. Возможны три случая:

(1). $p_0^-(X)+\varepsilon \le p_\varepsilon(X) \le p_0^-(X)+2\varepsilon$;

(2). $p_0^-(X)+2\varepsilon \le p_\varepsilon(X) \le p_0^+(X)-2\varepsilon$;

(3). $p_0^+(X)-2\varepsilon \le p_\varepsilon(X) \le p_0^+(X)-\varepsilon$.

Если имеет место случай (2), то необходимое распределение уже построено. Положим, что имеет место случай (1). Покажем, что существует распределение такое, что $p(X)=p_0^-(X)+2\varepsilon$. Для удобства записи введем обозначение $\delta = p_0^-(X)+2\varepsilon - p_\varepsilon(X)$ и заметим, что $\delta \le \varepsilon$. Возьмем распределение p_ε и, меняя его значение на квантах, построим новое распределение p. Мы увеличим значение на квантах входящих в СДНФ конъюнкта X и уменьшим на квантах не входящих так, чтобы

$$\sum\nolimits_{q \in S(X)} (p(q)-p_\varepsilon(q)) = \delta \text{ и } \sum\nolimits_{q \notin S(X)} (p(q)-p_\varepsilon(q)) = -\delta .$$

Это можно сделать т.к.

$$\sum\nolimits_{q \notin S(X)} p_\varepsilon(q) = 1 - p_\varepsilon(X) \ge \varepsilon \ge \delta .$$

Значение распределения на любом конъюнкте определяется суммой значений распределения на некотором наборе квантов, а, следовательно, в результате только что описанного изменения значения на каждом конъюнкте изменятся не более чем на δ. Значит $p_0^-(Y) \le p(Y) \le p_0^+(Y)$, для любого конъюнкта Y, и по построению $p_0^-(X)+2\varepsilon \le p(X) \le p_0^+(X)-2\varepsilon$. Нам удалось построить необходимое распределение. Для случая (3) построение аналогично, но мы будем уменьшать значения на квантах входящих в СДНФ конъюнкта X и увеличивать на не входящих. □

Требование на минимальную длину интервала естественно, так как показатель устойчивости не может превышать половины длины наименьшего интервала. Но если это условие не выполняется, то показатель устойчивой непротиворечивости будет в точности половина наименьшего интервала. Чтобы это доказать, просто в качестве ε, возьмем четверть длины наименьшего интервала.

Следующий пример показывает, что невозможно доказать большего увеличения показателя при переходе от множественной устойчивой непротиворечивости к устойчивой непротиворечивости.

Пример. Рассмотрим АБС на фрагменте знания второго порядка со следующими ограничениями: $0.3 \le p(x_1 x_2) \le 0.6$, $0 \le p(x_1) \le 0.4$, $0 \le p(x_2) \le 0.4$. В этом случае сеть будет множественно устойчиво непротиворечива с максимальным показателем 0.05, при этом максимальный показатель устойчивой непротиворечивости будет 0.1.

Отметим, что, разумеется, не всякая непротиворечивая сеть является устойчиво непротиворечивой. В качестве тривиального примера можно рассмот-

Труды СПИИРАН. 2005. Вып. 2. Т. 2. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2005. Issue 2. V. 2. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

реть сеть, представляющую собой фрагмент знаний второго порядка с ограничением $p(x_1)=0$. Тогда обязательно $p(x_1x_2)=0$, и, разумеется, всякое сколь угодно малое колебание оценок вероятности $p(x_1x_2)$, отделяющее эту оценку от нуля, сделает данную алгебраическую байесовскую сеть противоречивой.

Релевантным в рассмотренном примере было то, что оценка одного или нескольких конъюнктов сети была точечной, а не интервальной. Покажем, что это существенно для неустойчивости.

Теорема. Рассмотрим алгебраическую байесовскую сеть N на множестве атомарных пропозициональных формул $A=\{x_1,\ldots,x_n\}$. Пусть после процесса поддержания непротиворечивости каждый из отрезков

$$[p_0^-(x_{i_1}x_{i_2}\ldots x_{i_k}), p_0^+(x_{i_1}x_{i_2}\ldots x_{i_k})]$$

имеет непустую внутренность (ненулевую длину). Тогда эта сеть после процесса поддержания непротиворечивости устойчиво непротиворечива. Более того, в этом случае она множественно устойчиво непротиворечива. В случае же расширенной байесовской сети это утверждение неверно.

Доказательство. После процесса подержания непротиворечивости мы гарантированно можем построить серию распределений p_X такую, что для $\forall Y \in C$

$$p_X(X)=\frac{p_0^-(X)+p_0^+(X)}{2} \text{ и } p_0^-(Y)\le p_X(Y)\le p_0^+(Y).$$

Положим

$$\varepsilon=\min_{Y\in C}\left(\frac{p_0^+(Y)-p_0^-(Y)}{2}\right).$$

Тогда эта серия распределений показывает, что сеть является устойчиво непротиворечивой с показателем ε. А значит и множественно устойчиво непротиворечива. □

5. Выводы

В работе было показано, что из наличия устойчивости глобальной непротиворечивости следует наличие множественной устойчивости, показатель которой связан некоторым коэффициентом с показателем устойчивости. Очевидно, что из наличия множественной устойчивости глобальной непротиворечивости следует устойчивость глобальной непротиворечивости; однако эту связь можно охарактеризовать более точно: показатель устойчивости, как правило, будет больше, чем показатель множественной устойчивости. Достаточные условия и коэффициент увеличения указаны в основном тексте работы.

Стоит отметить, что приведённые примеры в точности реализуют указанные коэффициенты, а, следовательно, лучший результат в общем случае получить невозможно. Тем более интересным представляется поиск условий, которые позволят «улучшить» найденные соотношения для отдельных классов ситуаций.

Труды СПИИРАН. 2005. Вып. 2. Т. 2. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2005. Issue 2. V. 2. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

Литература

[1] *Городецкий В.И.* Алгебраические байесовские сети — новая парадигма экспертных систем // Юбилейный сборник трудов институтов Отделения информатики, вычислительной техники и автоматизации Российской академии наук, т. 2. М.: РАН, 1993.
[2] *Городецкий В.И.* Адаптация в экспертных системах // Известия РАН. Серия «Техническая кибернетика», №5. 1993.
[3] *Тулупьев А.Л.* Алгебраические байесовские сети: теоретические основы и непротиворечивость. СПб.: СПИИРАН, 1995.
[4] *Тулупьев А.Л.* Алгебраические байесовские сети: логико-вероятностный подход к моделированию баз знаний с неопределенностью. СПб.: СПИИРАН, 2000.
[5] *Никитин Д.А., Тулупьев А.Л., Ромашова М.Н.* Вычисление согласованных оценок истинности в вероятностных и нечетких фрагментах знаний // Труды СПИИРАН. Вып. 1, т. 2. СПб.: СПИИРАН, 2002.