УДК 004.8

А.А. Фильченков, А.Л. Тулупьев

АЛГОРИТМ ВЫЯВЛЕНИЯ АЦИКЛИЧНОСТИ ПЕРВИЧНОЙ СТРУКТУРЫ АЛГЕБРАИЧЕСКОЙ БАЙЕСОВСКОЙ СЕТИ НА ОСНОВЕ ОЦЕНКИ ЧИСЛА РЕБЕР В МИНИМАЛЬНОМ ГРАФЕ СМЕЖНОСТИ

Фильченков А.А., Тулупьев А.Л. **Алгоритм выявления ацикличности первичной структуры алгебраической байесовской сети на основе оценки числа ребер в минимальном графе смежности.**
Аннотация. Условием работы алгоритмов глобального логико-вероятностного вывода в алгебраической байесовской сети (АБС) является отсутствие циклов в ее вторичной структуре. Первичная структура, над которой можно построить ациклическую вторичную, называется ациклической. *Цель работы* — предложить алгоритм выявления ацикличности первичной структуры на основе оценки числа ребер в ее вторичной структуре без непосредственного построения вторичной структуры, а также оценка сложности этого алгоритма. В работе сформулирован алгоритм выявления ацикличности первичной структуры на основе оценки числа ребер в минимальном графе смежности полным перебором, доказана его корректность, оценена его сложность, предложено улучшение скорости работы этого алгоритма, доказана корректность и оценено время работы улучшенного алгоритма. Также рассмотрены возможности улучшения скорости работы этого алгоритма за счет использования алгоритмов построения элементов третичной полиструктуры АБС.
Ключевые слова: алгебраические байесовские сети, вероятностные графические модели систем знаний, глобальная структура, графы смежности ацикличность первичной структуры.

Filchenkov A.A., Tulupyev A.L. **Algorithm for Detection Algebraic Bayesian Network Primary Structure Acyclicity Based on Number of Minimal Join Graph Edges Estimating.**
Abstract. The condition for algebraic Bayesian networks (ABN) global logical-probabilistic inference algorithms performance is the absence of cycles in its secondary structure. The primary structure, on which an acyclic secondary structure can be synthesized is called an acyclic one. The goal of the work is to propose an algorithm to detect primary structure acyclicity based on estimates of the number of edges in its secondary structure without the direct construction of the secondary structure, and estimation of the algorithm complexity. The algorithm for detection ABN primary structure acyclicity based on number of minimal join graph edges estimating via brute force is formulated, its correctness is proven, its complexity is estimated, improvement in the speed of this algorithm is proposed, the improved algorithm correctness if proven and its performance time are estimated . Also the possibility for improving the algorithm performance speed through the use of algorithms for ABN tertiary polystructure elements synthesis is discussed.
Keywords: algebraic Bayesian networks, quaternary structure, machine learning, probabilistic graphical knowledge models, global structure, join graphs, primary structure acyclicity.

1. Введение. Алгебраическая байесовская сеть (АБС) — представитель класса логико-вероятностных графических моделей баз фрагментов знаний с неопределенностью [10, 12, 13, 18, 19], которая может быть применима [30] как одна из математических моделей анализа информационной безопасности [2, 3]. В частности, АБС могут использоваться [6] для оценки защищённости системы от социо-инженерных атак [1, 17]. Обработка интервальных оценок также оказывается востребованной в анализе гранулярных данных о рекордных интервалах между последними эпизодами [7, 8].

Работа алгоритмов глобального логико-вероятностного вывода (ГЛВВ) в АБС опирается на ее вторичную структуру [9, 11, 14, 15, 19, 32, 36], которая представляется графом смежности [4, 5, 19, 20, 21, 32–42] — особым графом, формальное определение которого будет приведено в тексте статьи. В настоящее время алгоритмы ГЛВВ могут быть проведены лишь для АБС, вторичная структура которой ациклична [9, 11, 14, 16]. Однако не над любой первичной структурой АБС можно построить ацикличную вторичную структуру. Те первичные структуры, над которыми возможно построение ациклической вторичной структуры, называются *ациклическими*, тогда как все прочие — циклическими.

Построения первичной и вторичной структур АБС являются подзадачами глобального обучения АБС [18, 19, 21]. Требования к методам построения данных структур имеют не только теоретическую природу, но также исходят из особенностей реализации соответствующих алгоритмов. Так, в рамках функциональности визуальной инструментальной платформы для работы с АБС предполагается [29], что ее первичная структура может быть как синтезирована на основе каких-либо «сырых» данных (статистических выборок, мнений экспертов, иных источников), так и явно задана пользователем. В последнем случае требуется обеспечить проверку в режиме реального времени, является ли вводимая пользователем первичная структура ациклической.

Прямой ответ на этот вопрос через построение, перебор и анализ всех возможных вторичных структур или даже через построение и анализ единственного представителя множества минимальных графов смежности (которое совпадает с множеством ациклических вторичных структур, если последнее не пусто) требует значительных временных затрат и в общем случае не подходит для работы в режиме, удобном пользователю. В связи с этим возникает вопрос относительно выработки методов выявления ацикличности первичной структуры по косвенным признакам.

Так, было предложено два подобных подхода [26, 31, 32]. Первый из них опирается на анализ еще одной — четвертичной структуры АБС и предполагает выявление особых циклов в ней [32, 36]. При этом построение четвертичной структуры в общем случае предшествует построению вторичной структуры.

Второй метод, который будет использован в этой статье, опирается на подсчет числа ребер в минимальных графах смежности без непосредственного их построения. Метод основывается на подсчете числа компонент связности в особых подграфах максимального графа смежности — сильных сужений.

Цель работы — предложить алгоритм выявления ацикличности первичной структуры на основе оценки числа ребер ее вторичной структуры без непосредственного построения этой структуры, а также оценить сложность этого алгоритма.

2. Основные определения и обозначения. Будем следовать обозначениям, введенным в работах [18, 19, 32, 37, 40].

Определение 1. *Граф* — пара $\langle V, E\rangle$, где V — множество вершин графа, а E — множество ребер, каждое из которых является неупорядоченной парой (v_i, v_j), $i \neq j, v_i, v_j \in V$. Для удобства будем через V и E обозначать функции от графа, возвращающие множество его вершин и множество его ребер соответственно: $V(G') = V'$; $E(G') = E'$, где $G' = \langle V', E'\rangle$. Также V от множества ребер будет возвращать множество концов этих ребер: $V(E) = \{v | \exists e \in E, \exists u: e = (v, u)\}$.

Определение 2. *Алфавит* — множество атомарных пропозициональных формул $A = \{x_1, \dots, x_n\}$.

Определение 3. *Слово* V — подмножество алфавита: $V \subseteq A$.

Определение 4. *Математическая модель фрагмента знаний в рамках графово-смежностного подхода* (для краткости будем говорить о *фрагментах знаний* (ФЗ)) — слово над заданным алфавитом.

Определение 5. *Набор максимальных фрагментов знаний* (МФЗ) — это набор фрагментов знаний, таких, что ни один фрагмент знаний не содержится ни в каком другом фрагменте знаний. Набор МФЗ — это *первичная структура* АБС.

Определение 6. *Вес* $W(K)$ *максимального фрагмента знаний* K — слово, которое состоит из атомов, вошедших в K.

Определение 7. *Сепаратор* $W(K_i, K_j)$ двух максимальными фрагментами знаний K_i и K_j — пересечение весов соответствующих МФЗ:

$$W(K_i, K_j) = W(K_i) \cap W(K_j).$$

Определение 8. Два максимальных фрагмента знаний называются *сепарированными*, если их сепаратор не пуст.

Определение 9. *Граф максимальных фрагментов знаний* — граф, построенный над набором МФЗ, такой, что ребра в нем возможны только между сепарированными вершинами. Вес ребра определяется как соответствующий сепаратор.

Определение 10. *Магистральный путь* $M(K_i, K_j)$ между двумя сепарированными максимальными фрагментами знаний K_i и K_j в графе МФЗ — это такой путь от K_i до вершины K_j, что вес любой принадлежащей этому пути вершины содержит сепаратор K_i и K_j.

Определение 11. Граф *магистрально связен,* если между каждой парой сепарированных вершин существует магистральный путь.

Определение 12. *Граф смежности* — магистрально связный граф МФЗ. Граф смежности — это *вторичная структура* АБС. Над одной и той же первичной структурой можно построить множество графов смежности.

Определение 13. *Минимальный граф смежности* — граф смежности, число ребер которого минимально. Множество минимальных графов смежности совпадает с множеством нередуцируемых графов смежности [5].

Определение 14. *Максимальный граф смежности* $G_{\max}$ — наибольший по числу ребер граф смежности. Так как в графе МФЗ возможны не все ребра, а только те, которые соединяют вершины, пересечение весов которых не пусто, то максимальный граф смежности вовсе необязательно совпадает с полным подграфом. Для заданного множества вершин существует и при этом единственный максимальный граф смежности [35].

Дополнительно предполагается, что первичная структура *связна* — что максимальный граф смежности, построенный над ней, будет связен. Это предположение обусловлено тем, что в обратном случае наборы вершин из каждого компонента связности имело бы смысл рассматривать как отдельные АБС.

Определение 15. Множество связных первичных структур будем обозначать как **CPS.**

Определение 16. Сужение на вес U — подграф максимального графа смежности, веса всех вершин и ребер которого содержат вес U.

Определение 17. *Значимая клика веса* U — сужение на вес U, совпадающий с каким-либо сепаратором максимального графа смежности. Множество всех значимых клик будем обозначать как Clique. Каждая значимая клика является полным графом.

Определение 18. *Сильное сужение веса* U — значимая клика веса U, из которой удалили все ребра веса U. Сильное сужение на вес U будем обозначать как $G_{\max} \nleftarrow U$.

Сильное сужение представляет собой компоненты связности, на которые разбивается соответствующая значимая клика.

Теорема 1 [34]. Число ребер любого минимального графа смежности над заданной АБС равно $\sum_{u \in U} \mathrm{Conn}(G_{\max} \downarrow U) - |U|$, где $\mathrm{Conn}(G)$ — число компонент связности графа G, а U — множество значимых весов.

Теорема 2 (2-я теорема о циклах в минимальных графах смежности). Над связной первичной структурой можно построить ациклический минимальный граф смежности тогда и только тогда, когда

$$\sum_{u \in U} \mathrm{Conn}(G_{\max} \downarrow U) - |U| = |V| - 1,$$

где $\mathrm{Conn}(G)$ — число компонент связности графа G, U — множество значимых весов, а V — множество вершин первичной структуры.

Будем считать, что на вход алгоритмов поступает неупорядоченное множество МЗФ $\mathrm{Weights}$, элементы которого представляются битовыми последовательностями фиксированной длины. Последнее гарантирует нам, что объединение и пересечение двух весов, а также проверка включения одного веса в другой, выполняются за $o(1)$.

В записи алгоритмов

- запись «$a \leftarrow b$» обозначает присвоение значения элемента b элементу a;
- запись «$S \rightarrow e\colon \mathrm{Cond}(e)$» обозначает извлечение произвольного элемента, удовлетворяющего условиям Cond, из множества S в переменную e. $\mathrm{Cond}(e)$ может отсутствовать, тогда речь идет об извлечении произвольного элемента множества S.

NUMBER OF COMPONENTS — функция, которая по заданным множеству вершин V и множеству ребер E графа возвращает число его компонент связности. NUMBER OF COMPONENTS работает за $O(|V| + |E|)$.

NUMBER OF REVERSED COMPONENTS — функция, которая по заданным множеству вершин V и множеству ребер E графа возвращает число компонент связности графа, обратного данному (т.е. графа, в котором две вершины смежны тогда и только тогда, когда они не смежны в исходном). NUMBER OF REVERSED COMPONENTS работает за$O(|V| + |V|^2 - |E|)$.

Множество UsefulWeighs будем считать упорядоченным, а множества Vertices, Edges и все Edges$[u]$ — неупорядоченными.

3. Алгоритм выявления ацикличности первичной структуры на основе оценки числа ребер в минимальном графе смежности полным перебором. Данный алгоритм (листинг 1) для заданного набора МФЗ определяет, можно ли над этой первичной структурой построить ациклический граф смежности, и, если можно, возвращает значение FALSE, иначе возвращает значение TRUE.

```
Require: Weights
Ensure: ∄w1, w2 ∈ Weights: w1 ⊆ w2, Weights ∈ CPS
UsefulWeights ← ∅
e ← 0
for all w1, w2 ∈ Weights, w1 ≠ w2 do
   w ← w1 ∩ w2
   if w ≠ ∅ do
      UsefulWeighs ← UsefulWeighs ∪ {w}
   end if
end for
for all u ∈ UsefulWeights
   Vertices ← ∅
   Edges ← ∅
   for all w ∈ Weights do
      if u ⊂ w do
         for all v ∈ Vertices
            if w ∩ v ≠ u do
               Edges ← Edges ∪ {(w, v)}
            end if
         end for
         Vertices ← Vertices ∪ {w}
      end if
   end for
   e ← e + NumberOfComponents(Vertices, Edges)
end for
hasNoCycles ← (e = |Weights| − 1 + |UsefulWeighs|)
return hasNoCycles
```

Листинг 1. Алгоритм выявления ацикличности первичной структуры на основе оценки числа ребер в минимальном графе смежности полным перебором.

В цикле (3–8) перебираются все пары вершин, и в том случае, если они сепарированы (5–7), их сепаратор добавляется к множеству значимых весов (6).

В цикле (9–23) перебираются все значимые веса, и для каждого значимого веса вычисляется число компонент связности в соответствующем сильном сужении.

В цикле (12–21) перебираются все вершины. Если выбранный значимый вес содержится в рассматриваемой вершине (13–20), то вершина добавляется к множеству вершин соответствующего значимого сужения, и к множеству ребер данного сужения добавляются все те, один из концов которых совпадает с данной вершиной. Так, в цикле (14–18) перебираются все вершины, уже добавленные к вершинам сильного сужения, и, если сепаратор двух выбранных вершин не совпадает с рассматриваемым весом (15–17), то он добавляется к множеству ребер, входящих в соответствующее сильное сужение (16).

На шаге (22) переменная e увеличивает свое значение на число компонент связности в сильном сужении на рассматриваемый вес.

На шаге (24) переменной $hasNoCycles$ присваивается значение TRUE в том случае, если сумма чисел компонент связности всех сильных сужений равна сумме числа вершин и числа значимых весов, уменьшенных на единицу, и FALSE в обратном случае.

Утверждение 1. Алгоритм выявления ацикличности первичной структуры на основе оценки числа ребер в минимальном графе смежности полным перебором для ациклической первичной структуры вернет значение TRUE , а для циклической — значение FALSE.

Доказательство. По завершению цикла (3–8) будет построено множество значимых весов UsefulWeights. Поскольку для каждого значимого веса в цикле (9–23) перебираются все вершины в цикле (12–21), то условие (13) будет рассмотрено для каждой соответствующей пары, таким образом, на шаге (19) в множество Vertices будут добавлены только те вершины, которые входят в сильное сужение на соответствующий вес. Таким образом, для каждого u из (9) при достижении шага (22) множество Vertices будет совпадать с множеством вершин, входящих в значимое сужение на вес u.

В цикле (14–18) будут рассмотрены все возможные пары вершин, входящих в значимое сужение на вес u. Сепаратор каждой пары содержит вес u, следовательно, те пары, сепаратор которых не совпадает с u, являются концами ребер в сильном сужении на вес u. Следовательно, на шаге (16) в множество Edges будут добавляться ребра сильного сужения, и поскольку будут рассмотрены все возможные пары вершин этого сильного сужения, то при достижении шага (22) множество Edges будет совпадать с множеством ребер, входящих в значимое сужение на вес u.

Таким образом, для каждого значимого веса u на шаге (22) переменная e увеличится на число компонент связности в соответствующем сильном сужении, поэтому при достижении шага (24) e будет равно сумме числа компонент связности каждого сильного сужения.

Из теоремы 2 следует, что первичная структура ациклична тогда и только тогда, когда равенство, проверяемое на шаге (24), имеет место. Следовательно, алгоритм корректен.

Утверждение 2. Сложность алгоритма выявления ацикличности первичной структуры на основе оценки числа ребер в минимальном графе смежности полным перебором равна

$$O\left(|W|^2 + \sum\nolimits_{u \in U} (|V_u| + |V_u|^2 - E_u + \log_2 |U|\, E_u + |W|)\right),$$

где W — множество вершин исходного графа (первичной структуры), U — множество значимых весов, E_u — множество ребер максимального графа смежности веса u; V_u — множество вершин, входящих в значимое сужение на вес u.

Доказательство. На выполнение шагов (3) и (4) потребуется в сумме

$$2|W|^2, \tag{1}$$

поскольку две операции будут осуществляться для каждой пары вершин.

На добавление значимого веса на шаге (5) всего потребуется не более

$$2 \log_2 |U| \cdot \sum\nolimits_{u \in U} E_u, \tag{2}$$

т. к. оно будет осуществляться ровно $\sum_{u \in U} E_u$ раз (для всех пар, пересечение весов которых непусто, т. е. образует значимый вес), требуя $2 \log_2 |U|$ операций для добавления элемента в упорядоченное множество UsefulWeights.

На выполнение проверки на шаге (13) потребуется

$$|W||U|, \tag{3}$$

поскольку он будет выполнен для каждой вершины и для каждого значимого веса.

На выполнение проверки на шаге (15) потребуется

$$\sum\nolimits_{u \in U} \frac{|V_u|^2}{2}, \tag{4}$$

поскольку она будет осуществлена для каждого значимого сужения для всех пар вершин, в него входящих.

На добавление значимого веса на шаге (16) всего потребуется

$$\sum\nolimits_{u \in U} (|V_u|^2 - E_u), \tag{5}$$

т. к. оно будет осуществляться для каждого значимого веса u ровно $|V_u|^2 - E_u$ раз, поскольку каждое ребро каждого сильного сужения будет добавлено ровно один раз, требуя одну операцию для добавления элемента в неупорядоченное множество Edges.

На добавление вершины на шаге (19) всего потребуется не более

$$\sum\nolimits_{u \in U} |V_u|, \tag{6}$$

т. к. оно будет осуществляться для каждого значимого веса u ровно $|V_u|$ раз, поскольку каждая вершина сильного сужения будет добавлена ровно один раз, требуя одну операцию для добавления элемента в неупорядоченное множество Vertices.

На выполнение шага (22) потребуется

$$\sum\nolimits_{u \in U} O(|V_u| + |V_u|^2 - E_u). \tag{7}$$

Сложив формулы (1)–(7) и сгруппировав их, получим искомое выражение.

4. Улучшенный алгоритм выявления ацикличности первичной структуры на основе оценки числа ребер в минимальном графе смежности полным перебором. Данный алгоритм (листинг 2) для заданного набора МФЗ определяет, можно ли над этой первичной структурой построить ациклический граф смежности, и если можно, то возвращает значение FALSE, а в противном случае - значение TRUE.

```
Require: Weights
Ensure: ∄w₁, w₂ ∈ Weights: w₁ ⊆ w₂, Weights ∈ CPS
UsefulWeights ← ∅
e ← 0
for all w₁, w₂ ∈ Weights, w₁ ≠ w₂ do
   w ← w₁ ∩ w₂
   if w ≠ ∅ do
      UsefulWeighs ← UsefulWeighs ∪ {w}
      Edges[w] ← Edges[w] ∪ {(w₁, w₂)}
   end if
end for
for all u ∈ UsefulWeights
```

```
11:     Vertices ← ∅
12:     for all w ∈ Weights do
13:        if u ⊂ w do
14:           Vertices ← Vertices ∪ {w}
15:        end if
16:     end for
17:     e ← e + NumberOfReseversedComponents(Vertices, Edges[u])
18:  end for
19:  hasCycles ← (e = |Weights| − 1 + |UsefulWeighs|)
20:  return hasCycles
```

Листинг 2. Улучшенный алгоритм выявления ацикличности первичной структуры на основе оценки числа ребер в минимальном графе смежности полным перебором.

В цикле (3–9) перебираются все пары вершин и в том случае, если они сепарированы (5–8), их сепаратор добавляется к множеству значимых весов (6), а пара добавляется к множеству ребер соответствующего веса (7).

В цикле (10–18) перебираются все значимые веса и для каждого значимого веса вычисляется число компонент связности в соответствующем сильном сужении.

В цикле (12–16) перебираются все вершины. Если выбранный значимый вес содержится в рассматриваемой вершине (13–15), то вершина добавляется к множеству вершин соответствующего значимого сужения (14).

На шаге (17) переменная e увеличивает свое значение на число компонент связности в сильном сужении на рассматриваемый вес.

На шаге (19) переменной $hasNoCycles$ присваивается значение TRUE в том случае, если сумма чисел компонент связности всех сильных сужений равна сумме числа вершин и числа значимых весов, уменьшенных на единицу, и FALSE в обратном случае.

Утверждение 3. Алгоритм выявления ацикличности первичной структуры на основе оценки числа ребер в минимальном графе смежности полным перебором для ациклической первичной структуры возвратит значение TRUE , а для циклической — значение FALSE.

Доказательство. По завершению цикла (3–9) будет построено множество значимых весов UsefulWeights и семейство множеств ребер каждого значимого веса.

Поскольку для каждого значимого веса в цикле (10–18) перебираются все вершины в цикле (12–16), то условие (13) будет рассмотрено для каждой соответствующей пары, таким образом, на шаге (17) в множество Vertices будут добавлены только те вершины, которые входят в сильное сужение на соответствующий вес. Таким образом, для каждого u из (9) при достижении шага (17) множество Vertices будет совпадать с множеством вершин, входящих в значимое сужение на вес u.

Для каждого значимого веса u на шаге (22) переменная e увеличится на число компонент связности в соответствующем сильном сужении, потому что сильное сужение представляет собой значимую клику (являющуюся кликой — полным графом), из которой удалили все ребра веса u.

Следовательно, при достижении шага (19) e будет равно сумме числа компонент связности каждого сильного сужения.

Из теоремы 2 следует, что первичная структура ациклична тогда и только тогда, когда равенство, проверяемое на шаге (19), имеет место. Следовательно, алгоритм корректен.

Утверждение 4. Сложность улучшенного алгоритма выявления ацикличности первичной структуры на основе оценки числа ребер в минимальном графе смежности полным перебором равна

$$O\left(|W|^2 + \sum\nolimits_{u \in U} (|V_u| + |V_u|^2 - E_u + \log_2 |U|\, E_u + |W|)\right),$$

где W — множество вершин исходного графа (первичной структуры), U — множество значимых весов, E_u — множество ребер максимального графа смежности веса u; V_u — множество вершин, входящих в значимое сужение на вес u.

Доказательство. Как уже было показано (утверждение 2), на выполнение шагов (4) и (5) потребуется в сумме

$$2|W|^2. \tag{8}$$

На добавление значимого веса на шаге (6) всего потребуется не более

$$2\log_2 |U| \cdot \sum\nolimits_{u \in U} E_u. \tag{9}$$

На добавление ребра на шаге (7) всего потребуется

$$\sum\nolimits_{u \in U} (|V_u|^2 - E_u), \tag{10}$$

т. к. оно будет осуществляться для каждого значимого веса u ровно $|V_u|^2 - E_u$ раз, поскольку каждое ребро каждого сильного сужения будет добавлено ровно один раз, требуя одну операцию для добавления элемента в неупорядоченное множество Edges.

Как было показано (утверждение 2), на выполнение проверки на шаге (13) потребуется

$$|W||U|, \tag{11}$$

а на добавление вершины на шаге (14) — не более

$$\sum\nolimits_{u \in U} |V_u|. \tag{12}$$

Наконец, на выполнение шага (17) потребуется

$$\sum\nolimits_{u \in U} O(|V_u| + |V_u|^2 - E_u). \tag{13}$$

Сложив формулы (8)–(13) и сгруппировав их, получим искомое выражение.

Хотя оценка сложности алгоритма не изменилась, тем не менее, улучшенный алгоритм будет всегда выполняться быстрее, поскольку он не делает лишних проверок на вес ребер внутри сильного сужения (листинг 1, шаг (15)), вместо этого добавляя ребра в соответствующее множество при построении множества значимых весов (листинг 2, шаг(7)). Соответственно, время работы алгоритма улучшено на

$$\sum\nolimits_{u \in U} \frac{|V_u|^2}{2}.$$

Следует отметить, что улучшение скорости частично компенсируется затратами памяти — вместо одного множества Edges, улучшенный алгоритм требует хранения семейства множеств $\mathrm{Edges}[u], u \in U$.

5. Возможность использования алгоритмов синтеза элементов третичной полиструктуры для ускорения работы алгоритма выявления ацикличности. Построение семейства сильных сужений тесно связано с построением элементов третичной структуры, в которых синтезируются значимые клики. Так, в частности, существуют алгоритмы (алгоритм построения родительского графа при помощи потомков [27], алгоритм построения родительского графа снизу–вверх [27], алгоритм построения родительского графа над множеством стереоклик перебором собственных ребер [28], Алгоритм построения родительского графа над множеством стереоклик полным перебором [28]), которые основаны на построении компонент связности значимого сужения. Однако этот алгоритм предполагает построение графовой структуры (родительского графа) над множеством значимых клик, что потребует дополнительного времени, поэтому нельзя утверждать, что подобный алгоритм будет работать заведомо быстрее, чем предложенный в настоящей статье.

Однако построение родительского графа требуется для некоторых алгоритмов синтеза множества минимальных графов смежности [23, 25], (следует отметить, что также существуют методы, не требующие синтеза этого графа [22, 24]). Следовательно, построенный элемент третичной полиструктуры может быть использован далее для синтеза вторичной структуры.

6. Заключение. В работе предложен алгоритм выявления ацикличности первичной структуры на основе оценки числа ребер в минимальном графе смежности полным перебором. Этот алгоритм позволяет отвечать на вопрос «Возможно ли построить над данной первичной структурой ациклический граф смежности?» без непосредственного построения самих графов смежности.

Сложность работы алгоритма зависит от ряда численных характеристик глобальной структуры АБС, а именно от числа МФЗ, числа значимых весов, числа ребер каждого значимого веса и числа вершин в каждой значимой клики. Поскольку пока не выявлено никаких фундаментальных соотношений между указанными величинам, оценку сложности графа нельзя упростить. Это на текущем этапе делает невозможным сравнение скорости работ предложенного алгоритма и алгоритма, основанного на анализе четвертичной структуры [31]. В работе также предложено ускорение алгоритма, которое основывается на том, что ребра следует группировать по их весам уже при переборе всех сепараторов набора вершин.

Сильные сужения структурно близки к значимым кликам, поэтому возможно создание алгоритмов выявления ацикличности первичной структуры, использующих алгоритмы синтеза того или иного элемента третичной структуры.

Литература

1. *Ванюшичева О.Ю., Тулупьева Т.В., Пащенко А.Е., Тулупьев А.Л., Азаров А.А.* Количественные измерения поведенческих проявлений уязвимостей пользователя, ассоциированных с социоинженерными атаками. // Труды СПИИРАН. 2011. Вып. 19. С. 34–47.
2. *Котенко И.В., Степашкин М.В., Юсупов Р.М.* Модели и методы информационной безопасности математические модели, методы и архитектуры для защиты компьютерных сетей: аналитический обзор перспективных направлений исследований по результатам международного семинара MMM-ACNS-2005 // Труды СПИИРАН. 2006. Т. 2. № 3. С. 11-29.
3. *Котенко И.В., Юсупов Р.М.* Перспективные направления исследований в области компьютерной безопасности // Защита информации. Инсайд. 2006. № 2. С. 46.
4. *Опарин В.В., Тулупьев А.Л.* Синтез графа смежности с минимальным числом ребер: формализация алгоритма и анализ его корректности // Труды СПИИРАН. СПб.: Наука, 2009. Вып. 11. С. 142–157.

5. *Опарин В.В., Фильченков А.А., Тулупьев А.Л., Сироткин А.В.* Матроидное представление семейства графов смежности над набором фрагментов знаний // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. 2010. Вып. 4. С. 73–76.
6. *Сироткин А.В., Тулупьев А.Л., Фильченков А.А., Пащенко А.Е., Тулупьева Т.В., Мусина В.Ф.* Особенности вероятностных графических моделей комплекса «информационная система – персонал» для оценки его защищенности от социоинженерных атак // Научная сессия НИЯУ МИФИ-2012 (30 января–4 февраля 2012 г., Москва). Аннотации докладов. В 3 т. Т. 3: Стратегические информационные технологии в атомной энергетике и промышленности. Проблемы информационной безопасности в системе высшей школы. Экономические и правовые проблемы инновационного развития атомной отрасли. Образование в национальном исследовательском ядерном университете. М.: НИЯУ МИФИ, 2012. С. 80.
7. *Суворова А.В., Пащенко А.Е., Тулупьева Т.В.* Оценка характеристик сверхкороткого временного ряда по гранулярным данным о рекордных интервалах между событиями // Труды СПИИРАН. 2010. Вып. 12. С. 170–181.
8. *Суворова А.В., Тулупьев А.Л., Пащенко А.Е., Тулупьева Т.В., Красносельских Т.В.* Анализ гранулярных данных и знаний в задачах исследования социально значимых видов поведения // Компьютерные инструменты в образовании. №4. 2010. С. 30–38.
9. *Тулупьев А.Л.* Алгебраические байесовские сети: глобальный логико-вероятностный вывод в деревьях смежности: Учеб. пособие. СПб.: СПбГУ; ООО Издательство «Анатолия», 2007. 40 с. (Сер. Элементы мягких вычислений).
10. *Тулупьев А.Л.* Алгебраические байесовские сети. Логико-вероятностный подход к моделированию баз знаний с неопределенностью. СПб.: СПИИРАН, 2000. 292 с.
11. *Тулупьев А.Л.* Алгебраические байесовские сети: система операций глобального логико-вероятностного вывода // Информационно-измерительные и управляющие системы. 2010. №11. С. 65–72.
12. *Тулупьев А.Л.* Алгебраические байесовские сети. Теоретические основы и непротиворечивость. СПб.: СПИИРАН, 1995. 76 с.
13. *Тулупьев А.Л.* Апостериорные оценки вероятностей в алгебраических байесовских сетях // Вестн. С.-Петерб. ун-та. Сер. 10. 2012. Вып. 2. С. 51–59.
14. *Тулупьев А.Л.* Байесовские сети: логико-вероятностный вывод в циклах. СПб.: Изд-во С.-Петербургского ун-та, 2008. 140 с. (Элементы мягких вычислений).
15. *Тулупьев А.Л.* Согласованность данных и оценка вероятности альтернатив в цикле стохастических предпочтений // Известия высших учебных заведений: Приборостроение. 2009. № 7. С. 3–8.
16. *Тулупьев А.Л., Абрамян А.К.* Логико-вероятностный вывод в направленном БСД-цикле // Труды СПИИРАН. 2007. Вып. 4. С. 87–118.
17. *Тулупьев А.Л., Азаров А.А., Пащенко А.Е.* Информационные модели компонент комплекса «Информационная система – персонал», находящегося под угрозой социоинженерных атак // Труды СПИИРАН. 2010. Вып. 3 (14). С. 50–57.
18. *Тулупьев А.Л., Николенко С.И., Сироткин А.В.* Байесовские сети: логико-вероятностный подход. СПб.: Наука, 2006. 607 с.
19. *Тулупьев А.Л., Сироткин А.В., Николенко С.И.* Байесовские сети доверия: логико-вероятностный вывод в ациклических направленных графах. СПб.: Изд-во С.-Петерб. ун-та, 2009, 400 с.
20. *Тулупьев А.Л., Столяров Д.М., Ментюков М.В.* Представление локальной и глобальной структуры алгебраической байесовской сети в Java-приложениях // Труды СПИИРАН. 2007. Вып. 5. СПб.: Наука, 2007. С. 71–99.

21. *Тулупьев А.Л., Фильченков А.А., Вальтман Н.А.* Алгебраические байесовские сети: задачи автоматического обучения // Информационно-измерительные и управляющие системы. 2011. № 11, т. 9. С. 57-61.
22. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи самоуправляемых клик // Труды СПИИРАН. 2010. Вып. 1 (12). С. 119–133.
23. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи самоуправляемых клик-собственников // Труды СПИИРАН. 2010. Вып. 3 (14) С. 150–169.
24. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи клик владений // Труды СПИИРАН. 2010. Вып. 2 (13). С. 119–133.
25. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи клик-собственников владений // Труды СПИИРАН. 2010. Вып. 4 (15). С. 193–212.
26. *Фильченков А.А.* Алгоритмы выявления ацикличности первичной структуры алгебраической байесовской сети // Научная сессия НИЯУ МИФИ-2012 (30 января–4 февраля 2012 г., Москва). Аннотации докладов. В 3 т. Т.2 Проблемы фундаментальной науки. Стратегические информационные технологии. М.: НИЯУ МИФИ, 2012. С. 276–277.
27. *Фильченков А.А.* Алгоритмы построения третичной структуры алгебраической байесовской сети // Труды СПИИРАН. 2011. Вып. 2(17). С. 197–218.
28. *Фильченков А.А.* Алгоритмы построения элементов третичной полиструктуры алгебраической байесовской сети // Труды СПИИРАН. 2011. Вып. 3(18). С. 237–266.
29. *Фильченков А.А.* Визуальная инструментальная платформа для работы с алгебраическими байесовскими сетями // Сборник докладов XV Международной конференции по мягким вычислениям и измерениям, 25-27 июня 2012, (SCM – 2012). Санкт-Петербург. 2012. С. 195–199.
30. *Фильченков А.А.* Математическое моделирование диагностической модели защищенности информационной системы на основе комбинирования неполной и неточной аналитической информации // VII Санкт-Петербургская межрегиональная конференция «Информационная безопасность регионов России (ИБРР-2011)». (26–28 октября 2011 г., Санкт-Петербург.) Материалы конференции. СПб.: СПОИСУ, 2011. С. 175–176.
31. *Фильченков А.А., Тулупьев А.Л.* Алгоритм выявления ацикличности первичной структуры алгебраической байесовской сети по ее четвертичной структуре // Труды СПИИРАН. 2011. Вып. 4(19). С. 128–145.
32. *Фильченков А.А., Тулупьев А.Л.* Анализ циклов в минимальных графах смежности алгебраических байесовских сетей // Труды СПИИРАН. 2011. Вып. 17. С. 151–173.
33. *Фильченков А.А., Тулупьев А.Л.* Понятие торакса в применении к исследованию графов смежности алгебраических байесовских сетей // Труды СПИИРАН. 2011. Вып. 16. С. 186–205.
34. *Фильченков А.А., Тулупьев А.Л.* Совпадение множеств минимальных и нередуцируемых графов смежности над первичной структурой алгебраической байесовской сети // Вестник Санкт-Петербургского государственного университета. Серия 1. Математика. Механика. Астрономия. 2012. Вып. 2. С. 65–74.
35. *Фильченков А.А., Тулупьев А.Л.* Структурный анализ систем минимальных графов смежности Труды СПИИРАН. 2009. Вып. 11. С. 104–127.

36. *Фильченков А.А., Фроленков К.В., Тулупьев А.Л.* Устранение циклов во вторичной структуре алгебраической байесовской сети на основе анализа ее четвертичной структуры // Труды СПИИРАН. 2012. Вып. 2(21). С. 143–156.
37. *Фильченков А.А., Тулупьев А.Л.* Третичная структура алгебраической байесовской сети // Труды СПИИРАН. 2011. Вып. 18. С. 164–187.
38. *Фильченков А.А., Тулупьев А.Л., Сироткин А.В.* Компаративный анализ клик минимальных графов смежности алгебраических байесовских сетей // Труды СПИИРАН. 2010. Вып. 2 (13). С. 87–105.
39. *Фильченков А.А., Тулупьев А.Л., Сироткин А.В.* Мощность множества минимальных графов смежности // Труды СПИИРАН. 2010. Вып. 4 (15). С. 136–161.
40. *Фильченков А.А., Тулупьев А.Л., Сироткин А.В.* Особенности анализа вторичной структуры алгебраической байесовской сети // Труды СПИИРАН. 2010. Вып. 1 (12). С. 97–118.
41. *Фильченков А.А., Тулупьев А.Л., Сироткин А.В.* Ребра графов смежности в контексте компаративного анализа клик минимальных графов смежности алгебраических байесовских сетей // Труды СПИИРАН. 2010. Вып. 3 (14). С. 132–149.
42. *Фильченков А.А., Тулупьев А.Л., Сироткин А.В.* Структурный анализ клик минимальных графов смежности // Вестн. Тверск. гос. ун-та. Сер.: Прикладная математика. 2011. №20. С. 139–151.

Фильченков Андрей Александрович — аспирант кафедры информатики математико-механического факультета С.-Петербургского государственного университета (СПбГУ), младший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН. Область научных интересов: автоматическое обучение вероятностных графических моделей. Число научных публикаций — 56. aaafil@mail.ru, СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450. Научный руководитель — А.Л. Тулупьев.

Filchenkov Andrey Alexandrovich — PhD student of Computer Science Department, SPbGU, junior researcher, Theoretical and Interdisciplinary Computer Science Laboratory, SPIIRAS Research area: machine learning of probabilistic graphical models. The number of publications — 56. aaafil@mail.ru, SPIIRAS, 14-th line V.O., 39, St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450. Scientific advisor — A.L. Tulupyev.

Тулупьев Александр Львович — д.ф.-м.н., профессор; заведующий лабораторией теоретических и междисциплинарных проблем информатики СПИИРАН, доцент кафедры информатики математико-механического факультета С.-Петербургского государственного университета (СПбГУ). Область научных интересов: представление и обработка данных и знаний с неопределенностью, применение методов математики и информатики в социокультурных исследованиях, применение методов биостатистики и математического моделирования в эпидемиологии, технология разработки программных комплексов с СУБД. Число научных публикаций — 250. ALT@iias.spb.su, www.tulupyev.spb.ru; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-3337, факс +7(812)328-4450.

Tulupyev Alexander Lvovich — PhD in Computer Science, Dr. of Sc. Professor; Head of Theoretical and Interdisciplinary Computer Science Laboratory, SPIIRAS, Associate Professor of Computer Science Department, SPbSU. Research area: uncertain data and knowledge representation and processing, mathematics and computer science applications in socio-cultural studies, biostatistics, simulation, and mathematical modeling applications in epidemiology, data intensive software systems development technology. Number of publications — 250. ALT@iias.spb.su, www.tulupyev.spb.ru; SPIIRAS, 14-th line V.O., 39, St. Petersburg, 199178, Russia; office phone +7(812)328-3337, fax +7(812)328-4450.

Поддержка исследования. Работа выполнена при финансовой поддержке РФФИ, проекты № **09-01-00861**-а «Методология построения интеллектуальных систем поддержки принятия решений на основе баз фрагментов знаний с вероятностной неопределенностью» и № **12-01-00945**-а «Развитие теории алгебраических байесовских сетей и родственных им логико-вероятностных графических моделей систем знаний с неопределенностью».

Рекомендовано ТиМПИ СПИИРАН, зав. лаб. А.Л. Тулупьев, д.ф.-м.н., проф.

Работа поступила в редакцию 01.06.2012.

РЕФЕРАТ

Фильченков А.А., Тулупьев А.Л. **Алгоритм выявления ацикличности первичной структуры алгебраической байесовской сети на основе оценки числа ребер в минимальном графе смежности.**

Условием работы алгоритмов глобального логико-вероятностного вывода в алгебраической байесовской сети (АБС) является отсутствие циклов в ее вторичной структуре. Однако не над любой первичной структурой возможно построение ациклической вторичной структуры. Первичная структура, над которой можно построить ациклическую вторичную, называется ациклической.

Построение над данной первичной структурой даже одной вторичной с последующим поиском циклов в ней требует слишком много времени. Поэтому актуален вопрос о выявлении ацикличности первичной структуры АБС по косвенным признакам. Известно два метода, один из которых основывается на анализе четвертичной структуры АБС, а другой — на оценке числа ребер вторичной структуры. Цель работы — предложить алгоритм выявления ацикличности первичной структуры на основе оценки числа ребер в ее вторичной структуре без непосредственного построения вторичной структуры, а также оценка сложности этого алгоритма.

В работе предложен алгоритм выявления ацикличности первичной структуры на основе оценки числа ребер в минимальном графе смежности полным перебором, который опирается на вторую теорему о циклах в минимальных графах смежности и предполагает построение множества сильных сужений для данной первичной структуры и вычисления числа компонент связности в каждом. Оценена сложность работы алгоритма, которая зависит от ряда численных характеристик глобальной структуры АБС, а именно от числа МФЗ, числа значимых весов, числа ребер каждого значимого веса и числа вершин в каждой значимой клики.

Предложено ускорение алгоритма, которое основывается на том, что ребра следует группировать по их весам уже при переборе всех сепараторов набора вершин. Вычислена сложность работы алгоритма, которая совпадает со сложностью работы исходного. Доказана корректность работы обоих алгоритмов.

Рассмотрена возможность создания алгоритмов выявление ацикличности первичной структуры, использующих алгоритмы синтеза того или иного элемента третичной полиструктуры.

Summary

Filchenkov A.A., Tulupyev A.L. **Algorithm for Detection Algebraic Bayesian Network Primary Structure Acyclicity Based on Number of Minimal Join Graph Edge Estimating.**

The condition for algebraic Bayesian networks (ABN) global logical-probabilistic inference algorithms performance is the absence of cycles in its secondary structure. But not every primary structure allows synthesis of an acyclic secondary structure over it. The primary structure, on which an acyclic secondary can be synthesized is called an acyclic one.

Synthesis of even the only secondary structure over given primary structure with further cycles search is highly time-cost. Therefore, the ABN primary structure acyclicity detection with indirect indications is an actual problem. There are the two known techniques, one of which is based on the ABN quaternary structure analysis and the other on the estimation of the number of edges in the secondary structure.

Algorithm for detection algebraic Bayesian network primary structure acyclicity based on number of minimal join graph edge estimating via brute force is proposed in the paper. It is based on the second theorem for minimal join graph cycles and requires synthesis of strong narrowings set of a given primary structure and number of each strong connected components calculation. The algorithm complexity is estimated. It depends on a number of numerical characteristics of the ABN global structure, namely the number of MKPs, the number of useful weights, the number of edges of each useful weight and the number of vertices in each useful clique.

Acceleration for the algorithm is proposed that is based on the grouping of edges according to their weights simultaneously with exhaustive search of all separators of the vertices set. The algorithm complexity is estimated, the estimation coincides with the complexity of the source algorithm. The correctness of both algorithms is proven.

The possibility of creating algorithms to detect the primary structure acyclicity based on the algorithms of synthesis of a tertiary polystructure particular element is regarded.