УДК 004.8

К.В. Фроленков, А.А. Фильченков, А.Л. Тулупьев

СИБЛИНГОВЫЙ КРИТЕРИЙ ЦИКЛИЧНОСТИ МИНИМАЛЬНЫХ ГРАФОВ СМЕЖНОСТИ

Фроленков К.В., Фильченков А.А., Тулупьев А.Л. **Сиблинговый критерий цикличности минимальных графов смежности.**

Аннотация. В связи с невозможностью применения некоторых алгоритмов глобального логико-вероятностного вывода над цикличной вторичной структурой алгебраической байесовской сети (АБС) и относительно значительной временной сложностью алгоритма построения такой структуры, целесообразно предъявить критерий, который позволит проверять цикличность АБС до процесса построения вторичной структуры. Статья предлагает один из таких критериев, основывающийся на анализе вспомогательной структуры (полусиблингового графа) на предмет наличия циклов особого класса.

Ключевые слова: алгебраические байесовские сети, вероятностные графические модели систем знаний, глобальная структура, ацикличность первичной структуры.

Frolenkov K.V., Filchenkov A.A., Tulupyev A.L. **Sibling criterion for cyclicity of minimal join graphs.**

Abstract. Due to the inability of some algorithms for global logical and probabilistic inference on cyclic secondary structure of algebraic Bayesian network (ABN) and a relatively large time complexity of the algorithm for constructing such a structure it's appropriate to present a criterion which allows checking if the ABN is cycling before building the secondary structure. The paper offers such criterion, based on the analysis of the support structure (semi-sibling graph).

Keywords: algebraic Bayesian networks, probabilistic graphical knowledge models, global structure, primary structure acyclicity.

1. Введение. Объектом исследования данной статьи является алгебраическая байесовская сеть (АБС). АБС принадлежит к классу логико-вероятностных графических моделей, которые позволяют представлять знания в виде оценок вероятностей пропозициональных формул в условиях ограничения памяти [28].

Если описывать данную модель, выделяя уровни представления, вывода и обучения, как предложено в [27], то представлением будет первичная и вторичная структуры АБС: множество фрагментов знаний (ФЗ) и граф смежности (ГС), отражающий связи между ними. При этом для представления неопределенности могут быть использованы как скалярные, так интервальные оценки вероятностей [2, 4, 7–10] (что является преимуществом АБС, так как для родственных моделей, таких как байесовские сети доверия и скрытые марковские модели, такое представление не разработано [25, 26]). Алгоритмы логико-вероятностного вывода также разделяются на локальные (выполняемые для одного ФЗ) и глобальные (для всей сети) [3, 5]. В теории АБС разделяют априорный вывод, состоящий в получении априорной вероятности формулы, а также апостериорный вывод, у которого выделяется две задачи: задача оценки вероятности поступившего свидетельства при условии задаваемого сетью вероятностного распределения (семейства таких распределений) и задача оценки вероятности заданных сетью вероятностных оценок при условии поступившего свидетельства [3, 7]. Задача обучения также состоит из нескольких подзадач: выделяют локальное обучение (построение вероятностных оценок во фрагменте знаний) и глобальное обучение (построение первичной и вторичной структур АБС) [12].

Задачу глобального обучения обычно [12] разделяют на две последовательные задачи: задачу построения первичной структуры АБС по статистическим данным и задачу построения вторичной структуры АБС над заданной первичной. Однако обучение первичной структуры должно также учитывать и особенности вторичной структуры, поскольку, как будет показано в рамках данной статьи, в противном случае сеть, построенная последовательным решением двух указанных задач, рискует оказаться неспособной выполнять определенные виды логико-вероятностного вывода.

Так, не любой граф с вершинами, соответствующими максимальным фрагментам знаний первичной структуры АБС, может быть использован в качестве вторичной структуры. Это связано с особенностями алгоритмов распространения поступивших в сеть свидетельств (алгоритмов глобального логико-вероятностного вывода) [6, 11], согласно которым за пересчетом вероятностей фрагмента знаний следует передача виртуального свидетельства, содержащего информацию обо всех элементах подыдеала конъюнктов — пересечения данного ФЗ и ФЗ, смежного с ним в графе [6].

На сегодняшний день известны лишь алгоритмы распространения свидетельств, опирающиеся на ациклическую вторичную структуру [6, 11]. Это означает, что для работоспособности АБС необходимо выбирать минимальный граф смежности в качестве вторичной структуры, поскольку известно [21], что при фиксированной первичной структуре АБС все графы смежности, не являющие минимальными по числу ребер, цикличны, а минимальные графы смежности, построенные по данному набору фрагментов знаний, цикличны или ацикличны одновременно. Это позволяет говорить о цикличности первичной структуры АБС безотносительно к методу построения графа смежности (цикличными АБС будем называть такие, первичная структура которых не допускает построение ацикличной вторичной структуры). Последнее накладывает дополнительные ограничения на алгоритмы обучения. Одним из таких ограничений, очевидно, является требование, согласно которому первичная структура АБС должна быть ацикличной, поскольку иначе алгоритмы глобального логико-вероятностного вывода окажутся неприменимыми к построенной над ней сети.

Выявление ацикличности первичной структуры АБС непосредственно — путем построения и проверки ацикличности всех или некоторых вторичных структур — чрезвычайно затратно по времени [13–16]. По этой причине существует необходимость в критерии, использование которого для выявления цикличности АБС будет требовать меньше времени (окажется менее трудоемким с вычислительной точки зрения). Применение такого критерия на этапе обучения первичной структуры позволит выбирать из множества подходящих структур только те, которые являются ациклическими.

Идеи, которые могут сформировать основу такого критерия, были представлены в [18],однако их недостаточно для проверки всех возможных первичных структур. Поэтому требуется развитие данных идей, так, чтобы стало возможным построить алгоритмы, обеспечивающие проверку критерия.

Цель работы — на основе положений из [18], сформировать критерий, позволяющий определять ацикличность первичной структуры по косвенным признакам, и доказать его корректность.

2. Описание предметной области. Для формализации задач и изложения результатов потребуется система терминов, введённая в [6, 20, 22, 23].

В теории АБС математическая модель фрагмента знаний определяется как идеал конъюнктов с заданными на нем вероятностными оценками. Однако при рассмотрении графов смежности вероятностные распределения, задаваемые фрагментами знаний, пока не играют никакой роли, поэтому в контексте целей данной работы будут рассматриваться лишь нагрузки фрагментов знаний.

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

Нагрузка $\mathrm{W}(U)$ *фрагмента знаний* U — подалфавит заданного алфавита $\mathrm{W}(U) \subseteq A$.

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*: $\forall i \neq j(W(V_i) \not\subset W(V_j))$ и $W((V_j) \not\subset W(V_i))$.

Сепаратор двух МФЗ — подалфавит, являющийся пересечением нагрузок этих фрагментов знаний. Два МФЗ называются *сочлененными*, если их сепаратор непуст.

Значимой нагрузкой будем называть нагрузку на непустой сепаратор каких-либо двух МФЗ. Множество значимых нагрузок обозначим как W.

Граф максимальных фрагментов знаний — ненаправленный граф, вершины которого соответствуют фрагментам знаний, вошедшим в АБС, ребра в котором возможны только между вершинами, соответствующими сочлененным ФЗ.

Нагрузка $W(\{V_i, V_j\})$ ребра $\{V_i, V_j\} \in E(G)$ графа G определяется как сепаратор его концов: $W(\{V_i, V_j\}) = W(V_i) \cap W(V_j)$. *Нагрузка* $W(H)$ подграфа $H \subseteq G$ — наибольший по включению подалфавит, который входит в нагрузки всех его вершин: $W(H) = \bigcap_{V \in H} W(V)$.

Магистральный путь между двумя сочлененными вершинами — это такой путь между ними, что нагрузка каждой его вершины содержит сепаратор концов этого пути.

Граф *магистрально связен*, если между каждой парой сочлененных вершин, существует магистральный путь.

Граф смежности — магистрально связный граф МФЗ.

Минимальный граф смежности (МГС) — граф смежности, число ребер которого минимально. Известно [1, 21], что минимальные графы смежности, и только они, являются *нередуцируемыми графами смежности*, то есть такими графами смежности, которые перестают быть магистрально связными при удалении любого ребра.

Вторичной структурой АБС может выступать некоторый граф смежности. Так как для корректной работы существующих на данный момент алгоритмов логико-вероятностного вывода требуется именно минимальный граф смежности [11, 17], в дальнейшем будем действовать в рамках предположения, что вторичная структура АБС — это МГС.

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 сочлененные вершины, максимальный граф смежности вовсе необязательно совпадает с полным подграфом. Для заданного множества вершин существует и при этом единственный максимальный граф смежности [21].

Сужение $G \downarrow U$ графа G на нагрузку U — это граф, в который входят только те вершины и ребра исходного графа G, веса которых содержат или равны U. *Значимое сужение* — сужение на нагрузку, являющуюся сепаратором (для какой-то пары МФЗ из заданного набора).

Каждой значимой нагрузке U, сопоставим *клику* $C(U) = G_{\max} \downarrow U$. Пустой нагрузке будет сопоставлена *праклика* $G_{\max}$.

Замкнутым сверху множеством нагрузок будем называть объединение W с одноэлементным множеством, содержащим пустую нагрузку. Заметим, что на таком множестве существует частичный порядок — отношение включения.

Родительским графом будем называть диаграмму Хассе такого порядка.

Высота нагрузки — максимальная длина пути до данной нагрузки от пустой в родительском графе.

Для данного подграфа H *накрывающая клика* — это клика максимальной высоты нагрузки, полностью содержащая H. Нетрудно заметить, что, поскольку праклика содержит любой подграф максимального графа смежности, то для любого подграфа существует хотя бы одна накрывающая клика. При этом не для любого подграфа накрывающая клика единственна. Если вес H — сепаратор, то вес накрывающей клики совпадает с весом H.

Для нагрузки Q и ее потомка D множество сыновей Q, которые являются предками D, будем называть *разделяющими сыновьями* и обозначать как $\mathrm{SepSon}_Q(D)$.

Утверждение 1. Множество потомков $\{D_i\}$ нагузки Q, для которых $\mathrm{SepSon}_Q(D_i) = \emptyset$, совпадает с множеством сыновей Q.

Доказательство. Заметим, что для каждого сына Q по определению не существует другого сына Q, который бы являлся потомком первого. Поэтому для каждого сына Q множество его разделяющих сыновей пусто.

С другой стороны, если для нагрузки D и ее предка Q нельзя указать нагрузку R, которая бы являлась предком D и потомком (в том числе сыном) Q, то нагрузка D по определению является сыном Q.

Рассмотрим произвольный путь $P = (e_1, \dots, e_n)$ в графе смежности G. Каждому ребру e_i поставим в соответствие клику с нагрузкой $W(e_i)$, а P — произвольную накрывающую ее клику Q. Рассмотрим кортеж $T = (S_1, \dots, S_n)$, где либо $S_i \in \mathrm{SepSon}_Q(e_i), \mathrm{SepSon}_Q(e) \neq \emptyset$, либо $S_i = e_i$ (т. е. каждому ребру сопоставлен какой-либо сын Q, в который это ребро входит). Будем при этом говорить, что сын S_i *накрывает* ребро e_i. В этом кортеже для всех i заменим все максимальные подпоследовательности одинаковых соседних элементов $S_i, \dots, S_i$ на элемент S_i (уточним, что при $n > 2$ элементы S_1 и S_n считаются соседними в том и только том случае, если ребра e_1 и e_n смежны, т. е. когда P — цикл). В полученном кортеже $T^* = (S_1^*, \dots, S_k^*)$ никакая пара соседних элементов не совпадает. Такой кортеж назовем *накрывающим кортежем сыновей*.

Минимальный накрывающий кортеж сыновей — такой накрывающий кортеж сыновей, число входящих сыновей в который минимально для данного пути P.

Для данного пути P в и его накрывающего кортежа сыновей T *разноцветной* называется вершина, два инцидентных ребра которой накрываются различными сыновьями.

Полусиблинговый граф для нагрузки — это граф, построенный над множеством её сыновей, ребро между двумя вершинами которого проведено тогда и только тогда, когда клики, соответствующие данным вершинам, пересекаются.

Полусиблинговый цикл — цикл полусиблингового графа. *Братский цикл* — полусиблинговый цикл, пересечение по вершинам клик всех элементов которого непусто. *Небратский цикл* — полусиблинговый цикл, не являющийся братским.

3. Контрпример. Идеи, представленные в [18], о связи цикличности вторичной структуры и наличия небратского полусиблингового цикла хотя и дают основу для критерия, но не образуют завершенной системы для его построения.

Например, для следующих нагрузок ФЗ первичной структуры: $\{ae, abcf, bcdg, bdh, cdi\}$ на рисунке 1 изображен минимальный граф смежности:

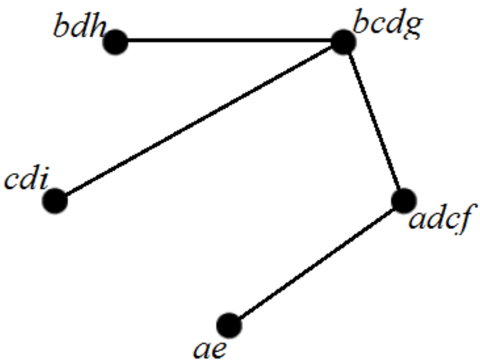

Рис. 1. Минимальный граф смежности.

Для такого набора нагрузок родительский граф будет выглядеть, как показано на рисунке 2:

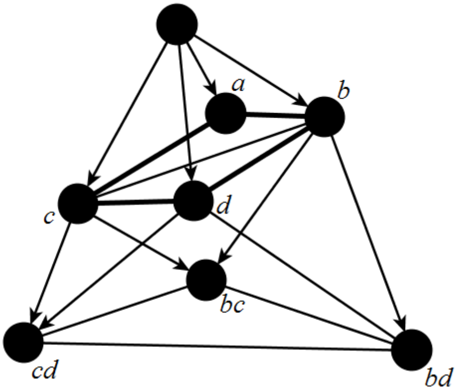

Рис. 2. Родительский граф с небратским полусиблинговым циклом $\boldsymbol{cabd}$.

Можно заметить, что существует полусиблинговый цикл (a, b, d, c, a) из потомков пустой нагрузки. При этом пересечение клик a и d пусто. Это означает, что цикл (a, b, d, c, a) — небратский. Пример показал, что существует ацикличная АБС, полусиблинговый граф, который содержит небратский полусиблинговый цикл. Таким образом, изначальная система идей из [18] недостаточна, чтобы построить искомый критерий.

4. Критерий цикличности АБС.

Теорема (1-я теорема о циклах минимальных графов смежности). Для фиксированной первичной структуры АБС любой минимальный граф смежности является ациклическим тогда и только тогда, когда не существует полусиблинговых небратских простых циклов в полусиблинговых графах, построенных над нагрузками данной первичной структуры.

Для доказательства потребуется лемма из [19].

Лемма (о независимом пути) [19]**.** Для любого несвязного подграфа G любого значимого сужения G^* любого графа смежности существует пара вершин v и u из G, таких, что существует магистральный путь между v и u, не пересекающийся по ребрам с G.

Доказательство теоремы. Для начала докажем, что если существует полусиблинговый небратский простой цикл, то любой МГС цикличен.

Будем рассматривать произвольный граф смежности и полусиблинговый небратский простой цикл $V_1, \ldots, V_n$, построенные над одним набором фрагментов знаний.

Заметим, что $C(V_1) \cap (C(V_2) \cup C(V_n))$ — несвязный подграф $C(V_1)$ (иначе в соответствующем полусиблинговом графе существует ребро между $C(V_2)$ и $C(V_n)$, и цикл не простой, или, в случае $n = 3$, $C(V_1) \cap C(V_2) \cap C(V_3) \neq \emptyset$, и этот цикл братский). Тогда по лемме о независимом пути существует пара вершин $(v, u) \in C(V_1) \cap (C(V_2) \cup C(V_n))$ и магистральный путь между ними, все вершины которого, за исключением v и u, лежат во множестве $C(V_1) \backslash (C(V_2) \cup C(V_n))$. С другой стороны, так как любая клика связна, $\bigcup_{i=2}^{n} C(V_i)$ — связный граф, и существует путь между v и u, принадлежащий $C(V_1) \backslash (C(V_2) \cup C(V_n))$. Объединение таких путей будет циклом.

Осталось доказать, что для цикличного МГС существует полусиблинговый небратский простой цикл.

Рассмотрим произвольный простой цикл $C = (e_1, \ldots, e_n)$ в МГС. По определению $W(C) = \bigcap_{i=1}^{n} W(e_i)$. Заметим, что ни один из $W(e_i)$ не совпадает с $W(C)$. В противном случае это ребро e_i можно было бы удалить без нарушения магистральной связности. Циклу C сопоставим произвольную накрывающую его клику Q_C.

Теперь рассмотрим множество M, составленное из таких клик Q_C для всех простых циклов МГС. В множестве M укажем подмножество M', состоящее из клик наибольшей высоты нагрузки. Из M' выберем такую клику Q_C, для которой минимальный накрывающий кортеж сыновей T содержит наименьшее число сыновей среди всех таких минимальных накрывающих кортежей сыновей. Обозначим количество разноцветных вершин T через p.

Заметим, существует хотя бы две разноцветные вершины, иначе T состоял бы из единственной нагрузки, являющейся сыном Q_C, клика которой содержит простой цикл C. Высота такой нагрузки хотя бы на единицу больше, чем высота нагрузки ее родителя. Отсюда следует, что Q_C не может входить в множество M', что приводит к противоречию.

Разберем 2 случая: $p = 2$ и $p > 2$.

Если $p = 2$, то в T всего две клики: S_1 и S_2, — тогда обе разноцветные вершины принадлежат $S_1 \cap S_2$, их веса содержат $S_1 \cap S_2$. В свою очередь, между этими разноцветными вершинами существует магистральный путь, веса всех ребер которого также содержат $S_1 \cap S_2$. Одно из ребер $e_s \notin S_1 \cap S_2$, в противном случае весь цикл лежит в $S_1 \cap S_2$. Не умаляя общности, предположим, что $e_s \notin S_2$. Тогда в S_1

существуют 2 не включающих друг друга пути между разноцветными вершинами. Следовательно, в S_1 есть цикл, и, следовательно, простой цикл. Получаем противоречие, поскольку высота нагрузки S_1 больше, чем у ее родителя Q_C.

Случай $p > 2$.

Заметим, что нагрузки $T = (S_1, \dots, S_n)$ образуют полусиблинговый цикл, поскольку все соседние в нем клики пересекаются по разноцветным вершинам, и, следовательно, приходятся друг другу братьями.

Пусть этот цикл не простой. $S_i \cap S_j = Y \neq \emptyset$ и $j \neq (i+1) \bmod n$. Для вершины $y \in Y$ существуют магистральные пути к разноцветной вершине x_1 из $S_i \cap S_{(i+1) \bmod n}$ и к разноцветной вершине x_2 из $S_j \cap S_{(j-1) \bmod n}$. Объединение этих путей не может содержать весь цикл, иначе любое ребро рассматриваемого цикла МГС принадлежит $S_i \cup S_j$ и, следовательно, $p = 2$. Видно, что из x_1 в x_2 существуют 2 пути в клике Q_C. Следовательно, в ней есть простой цикл и список клик-сыновей в котором не более $n - 1$ элементов (можно добраться из x_1 в x_2 по ребрам $e_1, \dots, e_n$, проходя не более $n - 2$ разноцветных вершин, включая x_1 и x_2, а путь из x_1 в x_2 через y включает только одну y, при $S_i \neq S_j$, и не включает их в случае $S_i = S_j$), что противоречит минимальности n.

Пусть цикл T — не братский. $\bigcap_{i=1}^{n} S_i = Y \neq \emptyset$. Для вершины $y \in Y$ есть магистральные пути к разноцветным вершинам x_1 и x_2из $S_1 \cap S_2$ и $S_2 \cap S_3$. Первый из них лежит в S_1, второй — в S_3. Заменив в изначальном цикле ребра между x_1 и x_2 на объединение этих путей, получим цикл, в котором y — разноцветная, а x_1 и x_2 — нет. Получили цикл с $n - 1$ разноцветной вершиной.

Следовательно, полусиблинговый небратский простой цикл найден.

5. Заключение. В работе представлен критерий цикличности вторичной структуры АБС, основанный на анализе четвертичной структуры АБС: семейства полусиблинговых графов, носителем которого является приведенная теорема. Построение полусиблингого графа может быть выполнено путем перебора вершин родительского графа, синтез которого является только первым этапом построения вторичной структуры [14, 13]. Таким образом, данный критерий применим для отсеивания цикличных первичных структур из множества предполагаемых первичных структур АБС на этапе обучения.

Также нужно отметить, что данный критерий не является единственным. Помимо тривиального способа построения какого-либо минимального графа смежности и определения его цикличности на данный момент существует алгоритм выявления цикличности путем оценки числа ребер в минимальном графе смежности [1].

Помимо определения цикличности АБС данный критерий может быть применён для преобразования цикличной первичной структуры в ацикличную путем добавления фрагментов знаний [24].

Вопросы о лучшем алгоритме поиска полусиблинговых простых небратских циклов и оценки сложности модифицированного на основе изменений, внесенных данной работой, алгоритма устранения циклов остаются открытыми. Кроме того, требуется сравнение эффективности существующих алгоритмов выявления цикличности АБС и, возможно, разработка новых.

Литература

1. *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.
2. *Тулупьев А.Л.* Алгебраические байесовские сети: логико-вероятностные графические модели баз фрагментов знаний с неопределенностью: Диссертация на соискание ученой степени д-ра физ.-мат. наук. СПб. 2009. 670 с.
3. *Тулупьев А.Л.* Алгебраические байесовские сети: локальный логико-вероятностный вывод: Учеб. пособие. СПб.: СПбГУ; ООО Издательство «Анатолия», 2007. 80 с. (Сер.Элементы мягких вычислений).
4. *Тулупьев А.Л.* Алгебраические байесовские сети: система операций локального логико-вероятностного вывода // Информационно-измерительные и управляющие системы. 2009. №4. С. 41–44.
5. *Тулупьев А.Л.* Ациклические алгебраические байесовские сети: логико-вероятностный вывод // Нечеткие системы и мягкие вычисления: Научный журнал Российской ассоциации нечетких систем и мягких вычислений. 2006. Том 1, № 1. С. 57–93.
6. *Тулупьев А.Л.* Байесовские сети: логико-вероятностный вывод в циклах. СПб.: Изд-во С.-Петербургского ун-та, 2008. 140 с. (Элементы мягких вычислений).
7. *Тулупьев А.Л., Николенко С.И., Сироткин А.В.* Байесовские сети: логико-вероятностный подход. СПб.: Наука, 2006. 607 с.
8. *Тулупьев А.Л.* Непротиворечивость оценок вероятностей в идеалах конъюнктов и дизъюнктов. // Вестник СПбГУ. Сер. 10. 2009. Вып. 2. С. 121–131
9. *Тулупьев А.Л., Сироткин А.В.* Алгебраические байесовские сети: принцип декомпозиции и логико-вероятностный вывод в условиях неопределенности // Информационно-измерительные и управляющие системы. 2008. № 10. т. 6. С. 85–87.
10. *Тулупьев А.Л.* Согласованность данных и оценка вероятности альтернатив в цикле стохастических предпочтений // Известия высших учебных заведений: Приборостроение. 2009. № 7. С. 3–8.
11. *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.

12. *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.
13. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи клик-собственников владений // Труды СПИИРАН. 2010. Вып. 15. С. 193–212.
14. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи самоуправляемых клик// Труды СПИИРАН. 2007. Вып. 4. СПб.: Наука, 2007. С. 87–118.
15. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи клик владений // Труды СПИИРАН. 2010. Вып. 13. С. 67–86.
16. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи самоуправляемых клик-собственников // Труды СПИИРАН. 2010. Вып. 14. С. 150–169.
17. *Фильченков А.А.* Требования к вторичной структуре алгебраической байесовской сети, индуцированные особенностями логико-вероятностного вывода // Юбилейная XIII Санкт-Петербургская международная конференция «Региональная информатика-2012 (РИ-2012)». Материалы конференции. СПОИСУ. СПб., 2012. С. 54.
18. *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.
19. *Фильченков А.А., Тулупьев А.Л.* Понятие торакса в применении к исследованию графов смежности алгебраических байесовских сетей // Труды СПИИРАН. 2011. Вып. 16. С. 87–105.
20. *Фильченков А.А., Тулупьев А.Л. Сироткин А.В.* Компаративный анализ клик минимальных графов смежности алгебраических байесовских сетей // Труды СПИИРАН. 2010. Вып. 13. С. 186–205.
21. *Фильченков А.А., Тулупьев А.Л.* Совпадение множеств минимальных и нередуцируемых графов смежности над первичной структурой алгебраической байесовской сети. Вестник СПбГУ. Серия 1. Математика. Механика. Астрономия. 2012. Вып. 2. С. 69–78.
22. *Фильченков А.А., Тулупьев А.Л.* Структурный анализ систем минимальных графов смежности Труды СПИИРАН. 2009. Вып. 11. С. 104–127.
23. *Фильченков А.А., Тулупьев А.Л.* Третичная структура алгебраической байесовскойсети. // Труды СПИИРАН. 2011. Вып. 18. С. 164–187.
24. *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. 21. С. 143–156.
25. *Darwiche A.* Modeling and Reasoning with Bayesian Networks. Cambridge University Press. 2009.
26. *Fraser A. M.*. Hidden Markov Models and Dynamical Systems. Society for Industrial and Applied Mathematics. 2011.
27. *Koller D., Friedman N.* Probabilistic Graphical Models: Principles and Techniques. Cambridge: The MIT Press. 2009.
28. *Pearl J.* Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Francisco: Morgan-Kaufman. 1988.

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 — 80. 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 — 80. 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.

Фроленков Константин Владиславович — аспирант кафедры информатики математико-механического факультета С.-Петербургского государственного университета (СПбГУ), младший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН. Область научных интересов: машинное обучение, вероятностный вывод. Число научных публикаций — 11. frolenk@mail.ru. Научный руководитель — А.Л. Тулупьев.

Frolenkov Konstantin Vladislavovich — PhD student of Computer Science Department, SPbGU, junior researcher, Theoretical and Interdisciplinary Computer Science Laboratory, SPIIRAS. Research area: machine learning, probabilistic inference. The number of publications — 11. frolenk@mail.ru. 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.

Рекомендовано ТиМПИ СПИИРАН, зав. лаб. А.Л. Тулупьев, д.ф.-м.н., проф.
Работа поступила в редакцию 11.01.2013.

Поддержка исследования. работа выполнена при финансовой поддержке РФФИ, гранты №№ 12-01-00945-а и 12-01-31202-мол_а.

РЕФЕРАТ

Фроленков К.В., Фильченков А.А., Тулупьев А.Л. **Сиблинговый критерий цикличности минимальных графов смежности.**

Алгебраическая байесовская сеть (АБС) принадлежит семейству вероятностных графических моделей.

Наиболее распространенным представлением АБС является представление в виде графа смежности. К графам смежности предъявляется несколько требований, среди которых можно выделить требование ацикличности. Ацикличность графа смежности инвариантна относительно метода построения последнего, поэтому возникает вопрос о выявлении ацикличности первичной структуры.

Очевидным методом проверки предполагаемой первичной структуры АБС на ацикличность и, следовательно, возможность её практического применения является построение вторичной структуры с последующим выявлением циклов в ней. Однако такая стратегия выбора порождает слишком большие временные затраты. В работе предложен критерий, позволяющий проводить такую проверку на основе более дешевой для построения структуры: семейства полусиблиговых графов, то есть графов построенных на множестве значимых нагрузок с ребрами соответствующими непустым пересечениям клик данных нагрузок. Согласно данному критерию цикличность графа смежности равносильна наличию хотя бы одного небратского полусиблингого цикла.

Описанный критерий может быть применен как для целей выбора необходимой первичной структуры АБС до этапа синтеза графа смежности, так и для проверки перед применением алгоритма распространения свидетельства по диаграмме Хассе.

Помимо приведенного критерия существует алгоритм выявления цикличности вторичной структуры через подсчет числа рёбер графа смежности без его построения. Преимуществом использования данного критерия является то, что помимо определения цикличности АБС его можно применять для преобразования цикличной первичной структуры в ацикличную путем добавления фрагментов знаний.

SUMMARY

Frolenkov K.V., Filchenkov A.A., Tulupyev A.L. **Sibling criterion for cyclicity of minimal join graphs.**

Algebraic Bayesian network (ABN) belongs to the family of probabilistic graphical models.

The most common representation of ABN is a representation in the form of the join graph. For join graphs several requirements are presented, among which we highlight the requirement of acyclicity. Acyclicity of a join graph is invariant under a construction method, so there is the question of the identification of acyclicity of a primary structure.

The obvious way to determine if the proposed primary structure of the ABN is acyclic and hence the possibility of its practical application is the construction of the secondary structure with the subsequent identification of cycles in it. However, this strategy of choice is too time-consuming. In this paper we propose a criterion to conduct such a test based on the cheaper to build structures: family of semi-sibling graphs, which is built on a set of graphs of significant workloads with edges corresponding nonempty intersection of clicks of these workloads. According to this criterion, the join graph is cycling if and only if at least one non-brotherly semi-sibling cycle exists.

Described criteria can be used both for the purpose of selecting the required primary structure of the ABN before the stage of a join graph synthesis, and for checking the primary structure before applying the algorithm of propagation an evidence on the Hasse diagram.

In addition to the above criterion there exists an algorithm for identifying cyclic secondary structure by counting the number of edges in the join graph without its construction. The advantage of using this criterion is that in addition to determining cyclicity of ABN it can be used to convert a cyclic primary structure to an acyclic one by adding fragments of knowledge.