УДК 004.8

А.А. Фильченков, А.Л. Тулупьев

ПОНЯТИЕ ТОРАКСА В ПРИМЕНЕНИИ К ИССЛЕДОВАНИЮ ГРАФОВ СМЕЖНОСТИ АЛГЕБРАИЧЕСКИХ БАЙЕСОВСКИХ СЕТЕЙ

Фильченков А.А., Тулупьев А.Л. **Понятие торакса в применении к исследованию графов смежности алгебраических байесовских сетей.**
Аннотация. Предложен новый терминологический поход для формализации работы с графами смежности, основанный на понятии торакса, обозначающего множество ребер. Предложена новая система уточненных понятий теории графов смежности: вес, сужение, жила, магистральная связность, минимальный граф смежности. Уточнены также понятие графа смежности и формулировка теоремы о множестве минимальных графов смежности. Сформулирована и доказана лемма о независимом пути, утверждающая, что из набора непересекающихся множеств ребер найдутся два таких, что магистральный путь между ними не пересекается ни с каким множеством из набора.
Ключевые слова: алгебраические байесовские сети, вторичная структура, машинное обучение, вероятностно-графические модели систем знаний, глобальная структура, теория графов.

Filchenkov A.A., Tulupyev A.L. **Term of Rib Cage in Application to the Algebraic Bayesian Network Join Graphs Analysis**
Abstract. A new terminological approach to join graph processing formalization based on a term of rib cage that is an edge set is suggested. A new system of the join graphs theory accurate terms such as weight, narrowing, sinew, backbone connectivity, minimal join graph is suggested. Also the term of join graph and Minimal Join Graph Set Theorem formulating are defined more preciously. A lemma of independent backbone path stating that there are two such edge sets from given collection of non-overlapping edge sets, that a backbone path between them is non-overlapping with any another edge set is formulated and proven.
Keywords: algebraic Bayesian networks, secondary structure, machine learning, probabilistic graphical knowledge models, global structure, graph theory.

1. Введение. Особое место в искусственном интеллекте занимают алгебраические байесовские сети (АБС), которые относятся к классу вероятностных графических моделей. АБС позволяют работать с неопределенностью, оперируя интервальными оценками вероятностей [1, 2, 15, 29]. Как и родственные им байесовские сети доверия (БСД), АБС представляются в виде графа, однако, в отличие от БСД, в АБС граф (называемый графом смежности) ненаправленный [7, 15, 17]. Вершинами этого графа являются максимальными фрагментами знаний, также называемые первичной структурой, тогда как связи между ними, соответствующие ребрам графа, называются вторичной структурой.

Над одной и той же первичной структурой может быть построено множество вторичных структур. От выбора того или иного графа смежности в качестве вторичной структуры зависит работа алгоритмов сети: на глобальном уровне логико-вероятностные ввод и вывод, а также поддержка непротиворечивости выдвигают особые требования к графу смежности [10–14, 16].

Одним из наиболее интересных вариантов для рассмотрения в качестве вторичной структуры являются минимальные графы смежности[1] (МГС), которые имеют структуру деревьев в том случае, если возможна ациклическая вторичная структура; в обратном случае они также являются наиболее вероятными кандидатами для эффективной работы алгоритмов логико-вероятностного вывода АБС [4–8, 23].

МГС стали объектом изучения в теории АБС сравнительно недавно, тем не менее уже получены существенные результаты. Так, в основополагающей работе [23] создана терминологическая база для исследований и доказаны многие важные утверждения, наиболее существенным из которых является теорема о декомпозиции множества МГС, которая стала теоретической предпосылкой для создания набора алгоритмов построения этого множества [19–22]. В дальнейшем теория была углублена [24, 26–28], за счет чего явно выражена мощность множества МГС [25], сформулирована и доказана теорема о перечислении особых компонент связности максимального графа смежности [28] и исследованы вторичные объекты (такие как феоды, вассалы и др.) [28], занимающие важно место в анализе вторичной структуры.

В некоторых аспектах исследований, связанных с МГС и вторичной структурой АБС вообще, много внимания уделяется ребрам, тогда как вершины играют подчиненную роль, что не очень характерно для теории графов. В некоторых случаях множества ребер являются основным рассматриваемым объектом (например, жилы, которые определяются именно как множества ребер). Это побудило нас выделить множество ребер в отдельное понятие и пытаться построить теорию, результаты которой формулируются в отношении множества ребер, а не графов.

Данный подход изначально имеет две цели:

1) мы хотели бы понятнее изложить материал, связанный с операцией сжатия и возникающими благодаря ей объектами, оказавшийся достаточно трудным для понимания в той форме, в которой он был

[1] Как было доказано [3], минимальность по числу ребер влечет минимальность в смысле нередуцируемости, обратное также верно.

изложен ранее [26–28] (особенно это касается понятий, связанных с операцией сжатия);

2) аппарат торакса (т. е. множества ребер, находящегося в рассмотрении, его строгая формулировка приведена ниже) использован для доказательства леммы, которая востребована для анализа циклов в графах смежности.

Целью данной работы является новая систематизация результатов исследований вторичной структуры АБС через понятие торакса, что само по себе представляет интерес с точки зрения теории графов, потому что предлагается новый взгляд на свойства уже известных объектов, тем самым становятся нагляднее операции и объекты, связанные с ребрами графов. Кроме того, известные результаты дополнены леммой о независимом магистральном пути.

2. Основные понятия и обозначения. Прежде всего определимся с обозначениями:

$\subseteq$ — *(нестрогое) включение*:

$$(A \subseteq B) \Leftrightarrow (\forall a \in A \;\; a \in B);$$

$\subset$ — *строгое включение*:

$$(A \subset B) \Leftrightarrow (\forall a \in A, a \in B \;\; \exists b \in B{:}\, b \notin A).$$

Алфавитом будем называть множество атомарных пропозициональных формул $A = \{x_1, \dots, x_n\}$, над которым заданы максимальные фрагменты знаний (МФЗ).

Слово V — подмножество алфавита: $V \subseteq A$.

Первичная структура или множество главных конъюнктов МФЗ, вошедших в байесовскую сеть, — такое множество слов

$$V^* = \{V_i \subseteq A\}_{i=1}^{i=m},$$

что:

1) оно не содержит несобственное подмножество алфавита:

$$V_i \neq A, V_i \neq \emptyset;$$

2) никакое слово полностью не содержит никакое другое слово:

$$\forall\, i \neq j \; (V_i \not\subset V_j) \text{ и } (V_j \not\subset V_i).$$

Вторичная структура или граф МФЗ (граф МФЗ) — ненаправленный граф, вершины которого соответствуют элементам множества главных конъюнктов МФЗ, вошедших в АБС, а ребра возможны только между теми вершинами, для которых пересечение соответствующих МФЗ непусто.

Вес $W(V_i)$ вершины V_i графа МФЗ — множество атомов алфавита, вошедших в V_i:

$$W(V_i) = \{x_i | x_i \in V_i\}.$$

3. Понятие торакса.

Определение 1. *Торакс*[1] — некоторое множество ребер, где под ребром понимается объект, для которого определено множество его концов, состоящее из двух вершин, причем эти вершины не совпадают, и не существует другого ребра, отличного от данного, множество концов которого совпадает с множеством концов данного ребра.

Определение 2. *Вершины* $V(t)$ торакса t — множество вершин, являющихся концами ребер, входящих в торакс.

За счет введенных определений можно для потребностей нашего исследования переформулировать определение графа следующим образом:

Определение 3. *Граф* $G(t)$ над t — пара $\langle V(t), t\rangle$.

Таким образом, граф является понятием, производным от торакса. Здесь отметим, что граф в данном понимании серьезно отличается от графа в понимании теории графов — в нашем определении в графе невозможно существование вершин, степень которых равна нулю. Для исследования графов смежности такое определение вполне подходит из-за дополнительных предположений, которые делаются относительно фрагментов знаний[2] (ФЗ). Однако в общем оно более ограничено, чем граф в привычном понимании, и мы еще столкнемся с этим ограничением, когда будет говорить про владения. Тем не менее аналогичное предположение можно сделать и во многих других областях применения графов, потому что обычно изолированные вершины могут быть успешно исключены из модели и в последующем рассмотрены отдельно, независимо от вершин из других компонент связности. Можно подытожить, что предложенное определение является более

---

[1] Торакс — транслитерация латинского слова *thorax*, обозначающего грудную клетку. На латыни существует термин *cavea thoracis*, более подходящий для данного понятия, поскольку обозначает группу костей, составляющих грудную клетку. На английский язык это переводится как *rib cage*, что совсем близко по смыслу к понятию множества ребер. Однако, к сожалению, на русский язык оба этих термина переводят как уже упомянутую грудную клетку, что неправильно, так как грудная клетка содержит еще и мышцы. Однако мы не будем углубляться в анатомическую лингвистику и лингвистическую анатомию и, за неимением адекватного русского аналога, воспользуется короткой и удобной транслитерацией латыни.

[2] Так, обособленный ФЗ, ни один атом которого не содержится ни в каких других ФЗ, не несет в себе никакого смысла в контексте рассматриваемой сети, и его нужно включать в сеть.

инструменталистским и прикладным, ориентированным на конкретные задачи, решаемые в теории АБС, чем традиционное определение графа.

Определение 4. *Оболочка* $C(t)$ торакса t — подмножество вершин торакса, являющихся концами ровно одного ребра, входящего в торакс.

Нетрудно заметить, что оболочка торакса представляет собой множество висячих вершин графа над этим тораксом. В частности, справедливы следующие определения.

Определение 5. *Путь* — торакс, мощность оболочки которого равна 2.

Теперь мы можем иначе взглянуть на понятие связности:

Определение 6. Две вершины *связны*, если образуют оболочку какого-либо пути.

Определение 7. *Связный торакс* t — такой торакс, любые две вершины которого связны.

Впрочем, то же самое можно определить, не прибегая к понятию связности, которое напрямую пришло из теории графов.

Определение 8. *Связный торакс* t — такой торакс, что для любого его разбиения на два непересекающихся непустых подторакса их вершины пересекаются:

$$\forall u \subset t: u \neq \emptyset \;\; V(u) \cap V(t \backslash u) \neq \emptyset.$$

Замечание 1. Одноэлементный торакс по определению связен (поскольку множество таких разбиений пусто).

Утверждение 1. Определения 7 и 8 эквивалентны.

Доказательство.

$\Rightarrow$. Рассмотрим торакс произвольного связного графа $G(t)$. Пускай существует такое разбиение на два непересекающихся непустых подторакса t_1 и t_2, что их вершины $V_1 = V(t_1)$ и $V_2 = V(t_2)$ не пересекаются. Так как любое ребро из t лежит либо в t_1, либо в t_2, то любая вершина из $V(t)$ является концом ребер из t_1 или t_2, поэтому лежит либо в V_1, либо в V_2 (так как они не пересекаются). Из этого следует, что граф $G(t)$ разбивается на две компоненты связности: $\langle V(t_1), t_1 \rangle$ и $\langle V(t_2), t_2 \rangle$, что противоречит предложению о его связности.

$\Leftarrow$. Пускай верно, что для любого разбиения торакса t на два непустых подторакса их вершины пересекаются. Предложим при этом, что граф $G(t)$ несвязен, т. е. существуют две такие вершины a и b, что между ними не существует пути.

Сначала выберем торакс, состоящий из ребер, одним из концов которых является a. Назовем его t_a. Очевидно, что b не входит в его вершины (иначе a и b были бы связны). Рассмотрим пересечение $V_i = V(t_a) \cap V(t \backslash t_a)$. Оно непусто. При этом каждая из вершин из V_i соединена с a, поэтому ни одна из этих вершин не должна быть связна с b. Тогда добавим к тораксу t_a торакс t_i, содержащий все ребра, один из концов которых совпадает с вершиной из V_i. Отметим, что $t_i \nsubseteq t_a$, потому что вершины из V_i являются концами каких-то ребер из $t \backslash t_a$, поэтому мощность t_a увеличится.

Продолжим те же действия, рассматривая вершины пересечения тораксов t_a и $t \backslash t_a$ и добавляя к t_a не входящий в него полностью торакс t_i. Поскольку мощность t конечна, а мощность t_a с каждым действием увеличивается, то после очередного шага t_a будет совпадать с t. Поскольку любая вершина из t_a по построению связна с a, то и b будет связна с a.

Мы доказали эквивалентность двух определений.

Утверждение 2. Любой торакс можно единственным способом представить объединением непересекающихся максимальных по включению связных тораксов.

Доказательство. Пускай существуют два способа представления торакса t в виде объединения непересекающихся максимальных по включению связных тораксов: $d^* = \{t_1^d, t_2^d, \dots\}$ и $e^* = \{t_1^e, t_2^e, \dots\}$. Положим $d = d^* \backslash e^*$ и $e = e^* \backslash d^*$. Заметим, что d и e оба не пусты.

Рассмотрим наибольший по мощности подторакс из d и e. Пускай это будет $t^d \in d$. Существуют хотя бы два таких подторакса t_1^e и t_2^e, что $t^d \cap t_1^e \neq \emptyset$ и $t^d \cap t_2^e \neq \emptyset$ (иначе единственный пересекающийся с t^d торакс из e содержал бы все ребра из t^d, а его мощность не превосходила бы t^d, т. е. они бы совпадали, что противоречит построению d и e). Так как t^d связен, то $V(t^d \cap t_1^e) \cap V(t^d \backslash t_1^e) \neq \emptyset$, откуда следует, что подторакс t_1^e пересекается с какими-то тораксами из e, что приводит нас к противоречию.

4. Сужение торакса и граф веса. В АБС первичной структурой называется набор МФЗ, а вторичной — граф над ними. Для работы АБС необходимы обе структуры, но создание АБС начинается с создания первичной структуры. Это несколько ограничивает выразительную силу торакса для определения графов смежности и некоторых других понятий. Именно поэтому мы вынуждены определить торакс МФЗ и вес торакса через МФЗ.

Определение 9. *Торакс МФЗ* над данным набором — торакс, состоящий из всех ребер, возможных в графе МФЗ (т. е. тех ребер, пересечение весов концов которых непусто).

По аналогии с максимальным графом смежности будет обозначать торакс МФЗ как $t_{\max}$.

Определение 10. *Вес торакса* $W(t)$ — пересечение весов всех его вершин:

$$W(t) = \bigcap_{v \in V(t)} W(v).$$

Утверждение 3. Для набора непересекающихся тораксов $\{t_i\}$ верно

$$W\left(\bigcup t_i\right) = \bigcap W(t_i).$$

Доказательство. По определению веса торакса

$$W\left(\bigcup t_i\right) = \bigcap_{v \in V(\cup t_i)} W(v) = \bigcap_{t_i} \bigcap_{v \in V(t_i)} W(v) = \bigcap W(t_i).$$

Под *сужением* $G \downarrow U$ графа смежности G на вес U понимается подграф G, в который входят все ребра и вершины графа G, вес которых содержит U.

Под *значимым сужением* графа G на вес U понимается сужение $G \downarrow U$, в котором вес U является значимым, т. е. существует ребро в графе G, вес которого строго равен U.

Определение 11. *Значимый вес* — такой вес, что существует подторакс $t \in t_{\max}$ с таким весом.

Нетрудно заметить, что это определение совпадает первым определением значимого веса.

Определение 12. *Сужение* $t \downarrow U$ *торакса* t на вес U — максимальный по включению подторакс, вес которого равен U.

Утверждение 4. $G(t \downarrow U) = G(t) \downarrow U$.

Доказательство. $G \downarrow U = G_U$. Докажем, что

$$E(G_U) = t \downarrow U \text{ и } V(G_U) = V(t \downarrow U).$$

1) $E(G_U) = \{e | e \in t, U \subseteq W(e)\}$.

1а) $U \subseteq W\big(E(G_U)\big)$, поэтому $E(G_U) \subseteq t \downarrow U$.

1б) Пускай $\exists e\colon e \in t \downarrow U, U \nsubseteq W(e)$, тогда $U \nsubseteq W\big(V(e)\big)$, тогда поскольку $W(V(t \downarrow U)) \subseteq W(V(e))$, то $U \nsubseteq W(V(t \downarrow U))$, что приводит к противоречию. Отсюда следует $t \downarrow U \subseteq E(G_U)$.

Объединяя результаты подпунктов 1а) и 1б), получим $E(G_U) = t \downarrow U$.

2) $V(G_U) = \{v | v \in V(t), U \subseteq W(v)\}$.

Так как $E(G_U) = t \downarrow U$, то $V\big(E(G_U)\big) = V(t \downarrow U)$. Поскольку G_U — значимое сужение, то оно состоит более чем из одной вершины, и любая вершина является концом хотя бы одного ребра, потому что соединена ребрами со всеми остальными вершинами сужения, поэтому $V\big(E(G_U)\big) = V(G_U)$, откуда следует $V(G_U) = V(t \downarrow U)$.

Утверждение доказано.

Под *кликой* веса U понимается значимое сужение максимального графа смежности на вес U.

Определение 13. *Полный торакс* — торакс t, мощность которого является наибольшей среди всех тораксов, вершины которых совпадают с вершинами t.

Замечание 2. Граф полного торакса полный.

Утверждение 5. Сужение торакса является полным тораксом.

Доказательство. Это следует из того, что вес тораксов с одинаковым множеством вершин совпадает.

Определение 14. *Весовой торакс* — сужение торакса МФЗ на значимый вес.

Замечание 3. Граф весового торакса является кликой.

Определение 15. *Граф веса* — направленный граф, вершинами которого являются весовые тораксы над данным набором ФЗ. Ребро из вершины P в вершину Q проведено, если весовой торакс P содержит весовой торакс Q, и не существует весового торакса R, такого, что весовой торакс P содержит весовой торакс R и весовой торакс R содержит весовой торакс Q.

Определение 16. Будем называть весовые тораксы, в которые входят ребра из весового торакса P в графе веса, *сыновьями* весового торакса P.

Так мы ввели объект, эквивалентный графу клик, который определялся как ориентированный граф над кликами с указанными свойствами.

5. Позвоночник, строгие сужения и жилы.

Определение 17. *Позвонок* V — это подмножество вершин графа (возможно, одноэлементное):

$$\mathrm{V} \subset V(G).$$

Множество всех позвонков будет обозначать как **V**.

Определение 18. *Позвоночник* CV — это подмножество множества позвонков, в котором позвонки попарно не пересекаются:

$$\mathrm{CV} \subset \mathbf{V}.$$

$$\mathrm{CV} = \{V_1, \dots, V_n\}, V_i \in \mathbf{V};\ \forall i, j < n, i \neq j\ V_i \cap V_j = \emptyset.$$

Множество вершин позвоночника будет обозначать как $V(\mathrm{CV})$.

Определение 19. Торакс t *возводи́м над позвоночником* CV, если его вершины содержатся в вершинах этого позвоночника:

$$t \text{ ad CV } \Leftrightarrow V(t) \subseteq V(\mathrm{CV}).$$

Замечание 4. Если торакс возводи́м над позвоночником, то любой его подторакс возводи́м над этим же позвоночником.

Определение 20. *Позвонки* $\mathrm{V}_{\mathrm{CV}}(t)$ торакса t, возводимого над данным позвоночником CV — минимальное по включению подмножество CV, такое, что $V(t) \subseteq V(\mathrm{V}_{\mathrm{CV}}(t))$.

По сути, позвонки торакса содержат только те позвонки позвоночника, которые содержат вершины торакса.

Определение 21. Торакс *строго возводи́м над позвоночником*, если множество позвонков этого торакса совпадает с этим позвоночником:

$$t \ \underline{\mathrm{ad}}\ \mathrm{CV} \ \Leftrightarrow V(t) = V(\mathrm{CV}).$$

Определение 22. *Пространство* $S(\mathrm{CV})$ *позвоночника* CV — множество всех таких тораксов, возводимых над позвоночником CV, что для каждого торакса из $S(\mathrm{CV})$ и любого его разделения на два непустых непересекающихся подторакса, их позвонки пересекаются.

$$S(\mathrm{CV}) = \{t | t \ \underline{\mathrm{ad}}\ \mathrm{CV}; \forall t^* \subset t, t^* \neq \emptyset, \mathrm{V}_{\mathrm{CV}}(t^*) \cap \mathrm{V}_{\mathrm{CV}}(t \backslash t^*) \neq \emptyset\}.$$

Определение 23. *Минимальное подпространство* $M(\mathrm{CV})$ *позвоночника* CV — множество минимальных по мощности тораксов, входящих в пространство CV:

$$M(\mathrm{CV}) = \left\{ t \middle| t \in S(\mathrm{CV}), |t| = \min_{t^* \in S(\mathrm{CV})} |t^*| \right\}.$$

Под *строгим сужением* графа смежности на вес U понимается его сужение на U, из которого удалили все ребра, в точности равного U. Под *владениями* понимаются компоненты связности такого графа.

К сожалению, к анализу владений нельзя применить тораксы, что обусловлено естественной ограниченностью последнего понятия: вершины вторичны по отношению к тораксу, но даже если существует способ перейти от множества вершин к тораксу, то способов выразить единственную вершину через торакс нет[1]. Это является основным препятствием, так как доменные вершины (т. е. владения, состоящие ровно из одной вершины) не могут быть выражены через торакс.

В связи с вышесказанным нам придется для удобства определить строгое сужение только для торакса ФЗ, причем определим его как позвоночник (потому что как торакс мы его определить не можем).

Определение 24. *Строгое сужение* $t_{\max} \not\downarrow U$ на значимый вес U — позвоночник CV, позвонки которого являются либо вершинами наибольших по включению связных подтораксов t_i^c торакса

[1] Задача представить одну единственную вершину через торакс кажется авторам идейно близкой к задаче представить хлопок одной ладони.

$(t_{\max} \downarrow U) \backslash \{e | e \in t_{\max}, W(e) = U\}$, либо одноэлементными множествами вершин из $V(t_{\max} \downarrow U) \backslash V(\bigcup t_i^c)$.

Легко заметить, что первый тип позвонков соответствует компонентам связности, число вершин в которых больше одной, а второй тип — одноэлементным компонентам связности (т. е. доменным вершинам). Конечно, можно представлять доменные вершины таким образом, но это кажется достаточно неудобным.

Под *жилой* веса U мы будем понимать множество ребер МГС над данным набором ФЗ, вес которых равен U. Обычно это рассматривается как свойство, но мы решили не воспроизводить определение жилы через оммажи, с которым можно ознакомиться в [26].

Утверждение 6. Множество жил веса U над данным набором ФЗ совпадает с минимальным подпространством строго сужения на вес U (над тем же набором ФЗ).

Доказательство. Любая жила является прообразом некоторого оммажа (см. [26]), т. е. дерева на владениях. Строгое сужение соответствует набору владений, а пространство над ними — всем связным графам над ними, такими, что каждые два владения (позвонка) соединяет ровно одно ребро. Минимальное же подпространство соответствует минимальном связным графам, т. е. деревьям. По сути, мы просто переформулировали определение жилы.

Определение 25. *Жила веса* U — торакс, принадлежащий минимальному подпространству строгого сужения на вес U.

Теперь мы можем переформулировать основную теорему о множестве МГС [22, 26]:

Теорема (о множестве минимальных графов смежности). Множество МГС совпадает с декартовым произведением минимальных подпространств всех строгих сужений, причем МГС равен графу объединения всех элементов соответствующего кортежа декартового произведения.

6. Торакс смежности. После того, как мы переформулировали большинством понятий графов смежности, вернемся к основе и попытаемся определить торакс смежности.

Под *магистральным путем* от вершины a до вершины b, пересечение весов которых не пусто, понимается такой путь от вершины a до вершины b, что вес любой принадлежащей ему вершины содержит пересечение весов начальной и конечной вершин.

Определение 26. *Магистральный торакс вершин* a и b — множество ребер магистрального пути между a и b.

То же самое можно определить иначе, не прибегая к понятию магистрального пути.

Определение 27. *Магистральный торакс вершин* — это путь (см. определение 5), вес которого содержит вес пересечения вершин оболочки.

Утверждение 7. Определения 26 и 27 эквиваленты.

Доказательство.

$\Rightarrow$. Магистральный путь между a и b является также и путем в понимании определения 5, т. е. число его висячих вершин равно 2. Так как вес каждой вершины этого пути содержит вес $a \cap b$, то пересечение весов всех вершин содержит вес $a \cap b$.

$\Leftarrow$. Так как в оболочку пути входит только две вершины, и пересечение их весов содержится в весе магистрального торакса, то оно содержится и в весе всех вершин этого торакса, а значит, и в весе каждого ребра.

Таким образом, определения эквивалентны.

Под *магистрально связным графом* понимается граф, между каждой парой несовпадающих вершин которого, пересечение весов которых не пусто, существует магистральный путь.

Определение 28. *Магистрально связный торакс* — торакс, для любых двух вершин которого существует магистрально связный подторакс, лежащий в этом тораксе.

Под *графом смежности* понимается магистрально связный граф МФЗ.

Определение 29. *Торакс смежности* — магистрально связный торакс МФЗ.

Однако мы можем определить торакс смежности и не прибегая к понятию магистральной связности.

Определение 30. Граф МФЗ является *графом смежности*, если любое его сужение связно.

Утверждение 8. Определение 30 совпадает с изначальным пониманием графа смежности.

Доказательство.

$\Rightarrow$. Пускай между любыми двумя вершинами, пересечение весов которых не пусто, существует магистральный путь. Если две вершины попадают в сужение, то связывающий их магистральный путь также попадает в это сужение, потому что вес каждого ребра этого пути содержит пересечение весов вершин, которое, в свою очередь, содержит вес сужения. Поэтому в любом сужении любые две вершины будут связны посредством магистрального пути.

$\Leftarrow$. Пускай в графе МФЗ любое сужение связно. Докажем, что между двумя вершинами тогда существует магистральный путь. Рассмотрим две вершины a и b, $W(a) \cap W(b) = U \neq \emptyset$. Тогда путь, связывающий a и b в сужении на вес U и будет магистральным путем от a до b, потому что все его ребра содержат вес U.

Определение 31. *Пространство* $S^t(\mathrm{CV})$ *позвоночника* CV при условии t — множество всех тораксов из $S(\mathrm{CV})$, пересеченных с тораксом t.

Определение 32. *Позвоночник фрагментов знаний* КР — позвоночник, каждый позвонок которого равен одноэлементному множеству, состоящему из ФЗ.

Определение 33. *Торакс смежности* — это торакс пространства $S^{t_{\max}}(\mathrm{КР})$.

Утверждение 9. Определения 33 и 29 эквивалентны.

Доказательство. Достаточно заметить, что граф МФЗ является графом торакса МФЗ, а пространство позвоночника ФЗ — это просто все возможные тораксы, построенные над набором ФЗ и составляющие связный граф. Таким образом, пересечение будет, с одной стороны, магистрально связно, а с другой, — каждый граф смежности можно представить как такое пересечение.

Определение 34. *Граф смежности* — граф торакса смежности.

Под МГС понимается граф, число ребер которого минимально. Это можно переопределить.

Определение 35. *Минимальный торакс смежности* — минимальный по числу ребер торакс смежности.

Приведем эквивалентную формулировку.

Определение 36. *Минимальный торакс смежности* — торакс минимального пространства $M^{t_{\max}}(\mathrm{КР})$.

Определение 37. *Минимальный граф смежности* — граф минимального торакса смежности.

Утверждение 10. Определение 37 эквивалентно изначальному пониманию графа смежности.

Доказательство. Эквивалентность следует из утверждений 6, 8 и 9, а также из того, что в минимальном пространстве по определению мощность тораксов минимальна.

После новых определений теорему о множестве МГС можно переформулировать:

Теорема о множестве МГС**.** Минимальное пространство

$$M^{t_{\max}}(\mathrm{КР})$$

совпадает с декартовым произведением минимальных пространств всех строгих сужений $t_{\max}$, причем каждый элемент $M^{t_{\max}}(\text{КР})$ равен объединению всех элементов соответствующего ему кортежа декартового произведения.

7. Лемма о независимом магистральном пути. Сформулируем и докажем лемму, которая применяет понятие торакса и потребуется для выявления циклов в графах смежности без их построения.

Пускай торакс t^* — некоторое сужение некоторого торакса смежности.

Утверждение 11. Для любых двух непересекающихся подтораксов x и y торакса t существуют такие вершины $p \in V(x)$ и $q \in V(y)$, что либо существует магистральный торакс b вершин p и q, не пересекающийся ни с x, ни с y, либо $p = q$.

Доказательство. Будем считать, что $V(x) \cap V(y) = \emptyset$, иначе просто возьмем произвольную вершину из этого пересечения, которая будет соответствовать данному предположению.

Рассмотрим две произвольные вершины из $V(x)$ и $V(y)$ (назовем их p_1 и q_1 соответственно), и магистральный торакс b_1, их связывающий. Обозначим $u_1 = W(b_1)$. Из сделанного в предыдущем абзаце предположения следует, что $c_1 = b_1 \backslash x \backslash y \neq \emptyset$, иначе бы вершины x и y пересекались.

Рассмотрим вершины из $C(c_1)$. Ту, которая в $V(x)$, назовем p_2, а ту, которая в $V(y)$ — q_2 (если $p_1 = p_2$ и $q_1 = q_2$, то мы нашли искомые вершины, потому что тогда $c_1 = b_1$). Если c_1 — магистральный торакс вершин p_2 и q_2, то мы нашли искомые вершины. Если же c_1 — не магистральный, то найдем магистральный торакс, обозначив его b_2, а вес его, соответственно, u_2. Так как магистральный торакс b_1 содержит c_1, а c_1 — не магистральный торакс, то

$$u_1 = W(b_1) \supseteq W(c_1) \supset W(b_2) = u_2.$$

Продолжим рассматривать пары вершин из $V(x)$ и $V(y)$, выбираемые как оболочки подторакса магистрального торакса предыдущей пары вершин, не лежащего ни в $V(x)$, ни в $V(y)$. Так как вес магистрального торакса каждой такой пары строго содержится в весе магистрального торакса предыдущей пары, а число пар вершин ограничено, то однажды мы найдем искомую пару вершин, магистральный торакс которых не пересекается ни с $V(x)$, ни с $V(y)$.

Определение 38. Магистральный торакс b из утверждения 11 назовем *магистральным тораксом тораксов* t_1 и t_2.

Лемма (о независимом магистральном пути). Для любого несвязного подторакса $t \in t^*$ существует пара вершин $v_1, v_2 \in V(t)$, таких, что существует магистральный торакс b вершин v_1 и v_2:

$$b \cap t = \emptyset.$$

Доказательство. По утверждению 2, любой торакс можно единственным образом представить объединением максимальных по включению непересекающихся связных подтораксов. Так как t — несвязный, то имеется не менее двух таких подтораксов.

Пускай $t = \bigcup t_i$, где t_i — указанные подтораксы. Рассмотрим магистральный торакс подтораксов t_1 и t_2, назовем его b_2, пускай его вес равен u_2, и он соединяет вершины v_1 и v_2 из $V(t_1)$ и $V(t_2)$ соответственно. Если он не пересекается ни с каким другим подтораксом, то искомыми вершинами будут вершины v_1 и v_2. В противоположном случае выберем какой-нибудь подторакс, с которым пересекается b_2, и обозначим его как t_3. Вершину из $V(t_3)$, ближайшую по b_2 к v_1 (в том смысле, что мощность связного подторакса b_2 с оболочкой, состоящей из v_1 и вершиной из $V(t_3)$, минимальна), обозначим как v_3. Если подторакс с оболочкой v_1, v_2 является магистральным тораксом этих вершин, то мы нашли искомые вершины. Если же нет, то это значит, что между ними существует магистральный путь, отличный от этого.

Обозначим первый подторакс, не являющийся магистральным тораксом, как c_2, а второй, являющийся магистральным тораксом, как b_3, а его вес как u_3. В результате получим формулу, похожую на формулу из утверждения 11:

$$u_2 = W(b_2) \supseteq W(c_2) \supset W(b_3) = u_3.$$

Продолжим выбирать подторакс, пересекающий магистральный торакс, а в нем — очередную вершину, и уже для нее будем строить магистральный торакс с вершиной t_1. Вес каждого такого торакса будет содержаться в весе предыдущего, а поскольку число пар всех вершин ограничено, однажды мы выберем такой магистральный торакс, что он не будет пересекаться ни с каким подтораксом.

8. Заключение. В работе к анализу графов смежности был подключен новый аппарат, основанный на понятии торакса — множества ребер. Через понятие торакса переформулировано большинство терминов графов смежности.

Мы решили не касаться владений, потому что достаточно сложно определять доменные вершины, тем не менее необходимые для исследования объекты были выделены при помощи понятия строго сужения торакса.

Переформулировка старых объектов теории в новых терминах сделана в первую очередь ради удобства и емкости терминологической базы этой теории. Так, аппарат, связанный с декомпозицией графов смежности (сжатия, оммажи, курии, жилы, пучки), удалось выразить более лаконично через применение понятий торакса и позвоночника. Кроме того, часть теории, связанная с сужениями, также получила качественный аналог с точки зрения понятийной структуры.

Помимо удобства и емкости введенных понятий переформулирование позволяет по-новому взглянуть на уже привычные объекты. Так, приведено несколько формулировок понятий графа смежности и самой теоремы о множестве МГС. Выражение объекта через новый смысловой конструкт ослабевает или даже разрушает инертность в понимании этого объекта, освобождает его из ограничений, накладываемых лингвистическими формулировками.

Дополнительной целью статьи было доказательство леммы о независимом магистральном пути, которую можно применить в свою очередь для доказательства теоремы о выявлении циклов в МГС. Выявление циклов в графах смежности без их непосредственного построения позволит лучше оценивать «качество» вторичной структуры, выраженное в том числе и в отсутствии циклов в МГС.

Применение аппарата тораксов не ограничивается исследованиями МГС, его можно использовать для решения более общих задач теории графов.

Литература

1. *Городецкий В.И.* Алгебраические байесовские сети — новая парадигма экспертно-вычислительных систем // Юбилейный сборник трудов институтов Отделения информатики, вычислительной техники и автоматизации РАН. Т. 2. М.: РАН, 1993, С. 120–141.
2. *Городецкий В.И., Тулупьев А.Л.* Формирование непротиворечивых баз знаний с неопределенностью // Известия РАН. Сер. Теория и системы управления. 1997. №5. С. 33–42.
3. *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.
4. *Павельчук А.В., Тулупьев А.Л., Тотмянина С.А.* Подход к объектно-ориентированному представлению данных алгебраических байесовских сетей в java-коде и реляционных СУБД // Региональная информатика-2008 (РИ-2008). XI Санкт-Петербургская международная конференция. Санкт-Петербург, 22–24 октября, 2008 г.: Материалы конференции / СПОИСУ. СПб.: 2009. С. 68–76.

5. *Сироткин А.В.* Модели, алгоритмы и вычислительная сложность синтеза согласованных оценок истинности в алгебраических байесовских сетях // Информационно-измерительные и управляющие системы. 2009. №11. С. 32–37.
6. *Тотмянина С.А., Павельчук А.В., Тулупьев А.Л.* Алгебраические байесовские сети: структуры данных в СУБД и Java- коде // Интегрированные модели, мягкие вычисления, вероятностные системы и комплексы программ в искусственном интеллекте. Научно-практическая конференция студентов, аспирантов, молодых ученых и специалистов (Коломна, 26–27 мая 2009 г.). Научные доклады. В 2-х т. Т. 2. М.: Физматлит, 2009, С. 123–131.
7. *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).
8. *Тулупьев А.Л.* Алгебраические байесовские сети: реализация логико-вероятностного вывода в комплексе java-программ // Труды СПИИРАН. СПб.: Наука, 2009. Вып. 8. С. 191–232.
9. *Тулупьев А.Л.* Алгебраические байесовские сети: система операций глобального логико-вероятностного вывода // Информационно-измерительные и управляющие системы. 2010. № 11. С 65–72.
10. *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.
11. *Тулупьев А.Л.* Байесовские сети: логико-вероятностный вывод в циклах. СПб.: Изд-во С.-Петербургского ун-та, 2008. 140 с. (Элементы мягких вычислений.)
12. *Тулупьев А.Л.* Непротиворечивость оценок вероятностей в алгебраических байесовских сетях. Вестник СПбГУ. Сер. 10. 2009. Вып. 3. С. 144–151.
13. *Тулупьев А.Л.* Преобразование ациклических байесовских сетей доверия в алгебраические байесовские сети // Известия высших учебных заведений: Приборостроение. 2009. № 3. С. 21–23.
14. *Тулупьев А.Л.* Согласованность данных и оценка вероятности альтернатив в цикле стохастических предпочтений // Известия высших учебных заведений: Приборостроение. 2009. № 7. С. 3–8.
15. *Тулупьев А.Л., Николенко С.И., Сироткин А.В.* Байесовские сети: логико-вероятностный подход. СПб.: Наука, 2006. 607 с.
16. *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.
17. *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.
18. *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.
19. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи клик владений // Труды СПИИРАН. 2010. Вып. 1 (13). С. 67–86.
20. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи клик-собственников владений // Труды СПИИРАН. 2010. Вып. 4 (15). С. 193–212.

21. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи самоуправляемых клик // Труды СПИИРАН. 2010. Вып. 1 (12). С. 119–133.
22. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи самоуправляемых клик-собственников // Труды СПИИРАН. 2010. Вып. 3 (14) С. 150–169.
23. *Фильченков А.А., Тулупьев А.Л.* Структурный анализ систем минимальных графов смежности Труды СПИИРАН. 2009. Вып. 11. С. 104–127.
24. *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.
25. *Фильченков А.А., Тулупьев А.Л., Сироткин А.В.* Мощность множества минимальных графов смежности // Труды СПИИРАН. 2010. Вып. 4 (15). С. 136–161.
26. *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.
27. *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. 2 (14). С. 132–149.
28. *Фильченков А.А., Тулупьев А.Л., Сироткин А.В.* Структурный анализ клик минимальных графов смежности // Вестник Тверского государственного университета. Сер. Прикладная математика. 2011. [в печати].
29. *Gorodetsky V.I., Drozdgin V.V., Jusupov R.M.* Application of Attributed Grammar and Algorithmic Sensitivity Model for Knowledge Representation and Estimation // Artificial Intelligence and Information, Control Systems of ROBOTSA. Amsterdam: Elsevier Science Publishers B. V., 1984, P. 232–237.

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 — 20. 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 — 20. 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; заведующего 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 — 230. 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 — 230. 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 28.03.2011.

Поддержка исследования. Работа выполнена при финансовой поддержке РФФИ, проект № **09-01-00861**-а «Методология построения интеллектуальных систем поддержки принятия решений на основе баз фрагментов знаний с вероятностной неопределенностью», а также грантом правительства Санкт-Петербурга для победителей конкурса грантов Санкт-Петербурга для студентов, аспирантов, молодых ученых, молодых кандидатов наук 2010 г., диплом ПСП№10697.

РЕФЕРАТ

Фильченков А.А., Тулупьев А.Л. **Понятие торакса в применении к исследованию графов смежности алгебраических байесовских сетей.**

Введено понятие торакса, которым обозначается множество ребер графа. На основе этого понятия введено понятие графа (как пары множества концов ребер торакса и самого торакса). Предложена формулировка связности торакса, не основывающаяся на понятии связности графа.

Введено понятие веса торакса как пересечение весов его вершин. Понятие сужение перенесено на торакс.

Введен термин *позвоночник*, определенный как набор непересекающихся множеств вершин, пространство позвоночника введено для обозначения тораксов, определенных над этим позвоночником, любое разделение которых на два непустых непересекающихся подторакса связно по множествам вершин, входящих в позвоночник. Минимальным подпространством позвоночника названо его подпространство, состоящие из тораксов с наименьшей мощностью.

Показано, что множество жил совпадает с минимальным подпространством особых позвоночников, которые названы строгими сужениями. Благодаря доказанному, теорема о множестве минимальных графов смежности (МГС) переформулирована: множество МГС совпадает с декартовым произведением минимальных подпространств всех строгих сужений, причем МГС равен графу объединения всех элементов соответствующего кортежа декартового произведения.

Введено понятие магистрального торакса, определяемого как торакс, состоящий из ребер магистрального пути, а также как торакс с двумя висячими вершинами, вес которого содержится в пересечении их весов. Граф смежности переопределен как граф, любое сужение которого связно, а также как граф торакса, являющегося элементом пространства позвоночника, состоящего из одноэлементных фрагментов знаний и пересеченного с тораксом фрагментов знаний. Последний определяется как торакс максимальной мощности, вершинами которого являются эти фрагменты знаний и сохраняется свойство смежности.

Сформулирована и доказана лемма о независимом магистральном пути, утверждающая, что для любого несвязного подторакса сужения торакса смежности существует такая пара его вершин, что их магистральный торакс не пересекается с этим подтораксом.

ABSTRACT

Filchenkov A.A., Tulupyev A.L. **Term of Rib Cage in Application to the Algebraic Bayesian Network Join Graphs Analysis.**

Term of rib cage that designate an edge set is introduced. A term of graph is reformulated by means of the term as a pair of set of edges vertices and the rib cage). Rib cage connectivity is defined without using a graph connectivity term.

Rib cage weight is defined as intersection of its vertices weights. The term of narrowing is applied to the rib cage.

Vertebral column is defined as a non-overlapping vertex set collection, vertebral column is introduces to designate rib cages that are defined on this vertebral column and such that any its partition into two not empty non-overlapping rib subcages is connected by the vertex set belong to the vertebral column. A vertebral column minimum subspace is defined as the vertebral column subspace that consists of rib cages with minimal cardinal number.

It is shown that sinews set coincides with minimum subspace of the special vertebral column that is called a strict narrowing. The theorem of minimal join graph set is reformulated: minimal join graph set is coincide with a Cartesian product of minimal subspaces of all the strict narrowings, and a minimal join graph equals to the graph of the intersection of all the elements of the corresponding tuple.

A term of backbone rib cage that is defined as a rib cage that consists of a backbone path edges and also as a rib cage with only two dangling vertecies and its weights is in the intersection of the vertices weights is introduced. A joint graph is redefined as a graph, such that every its narrowing is connected and also as a graph of a rib cage from such a vertebral column space, that is a set of a singleton knowledge pattern sets and that is intersected with the rib cage of knowledge patterns defined as the rib cages with minimal cardinal number and satisfies to the request of jointy.

The lemma of independent backbone path that states that for every unconnected rib subcage of a join rib cage narrowing there are two its vertices such that a backbone ribcage conneting them do not intersect the rib subcage is proven.