УДК 004.8

А.А. Фильченков, В.Ф. Мусина, А.Л. Тулупьев

АЛГОРИТМ РАНДОМИЗИРОВАННОГО СИНТЕЗА МИНИМАЛЬНОГО ГРАФА СМЕЖНОСТИ

Фильченков А.А., Мусина В.Ф., Тулупьев А.Л. **Алгоритм рандомизированного синтеза минимального графа смежности.**

Аннотация. В теории алгебраических байесовских сетей стоит задача построения вторичной структуры сети по известной первичной структуре. Для осуществления логико-вероятностного вывода в качестве вторичной структуры может выступать только минимальный граф смежности. В статье сформирован алгоритм рандомизированного синтеза минимального графа смежности. Доказана теорема о том, что выбор любого возможного для заданной первичной структуры алгебраической байесовской сети минимального графа смежности имеет положительную вероятность.

Ключевые слова: алгебраические байесовские сети, вторичная структура, графы смежности, автоматическое обучение, случайные графы.

Filchenkov A.A., Musina V.F., Tulupyev A.L. **Algorithm for minimal join graph's randomized synthesis.**

Abstract. The problem of algebraic Bayesian network's (ABN) secondary structure construction based on known primary structure was outlined in the ABN's theory. Logic-probabilistic inference can be performed only when secondary structure is a join graph. The algorithm for randomized synthesis of minimal join graph is formulated in the paper. The theorem was proved that the selection of every probable for the ABN's primary structure minimal join graph has positive probability.

Keywords: algebraic Bayesian networks, secondary structure, join graphs, automated learning, random graphs.

1. Введение. Логико-вероятностные графические модели используются для представления знаний с неопределенностью [21, 35]. В качестве меры неопределённости в таких моделях выступает вероятностная мера, определяющая либо вероятности истинности (скалярные оценки вероятности истинности), либо интервальные оценки вероятности истинности, как поступающих свидетельств, так и элементов фрагментов знаний, составляющих логико-вероятностную графическую модель.

Узлами алгебраической байесовской сети (АБС) [5, 6, 7, 10, 11, 13] выступают сложные случайные элементы [27] (и даже семейства таких случайных элементов, единых по структуре, но различающихся по значениям параметров) — математические модели фрагментов знаний (фрагменты знаний) [6, 11, 13]. Набор фрагментов знаний формирует первичную структуру алгебраической байесовской сети. Апостериорный вывод — получение вероятностных оценок истинности пропозиций при условии поступивших свидетельств — один из ключевых процессов, алгоритмизация которого развивается и изучается в теории алгебраических байесовских сетей [4, 11, 12]. В рамках апостериорного вывода поступившее свидетельство и порожденные им свидетельства (виртуальные свидетельства) передаются между фрагментами знаний [5, 8, 9, 13]. Графическая структура модели должна не только отображать, но и обеспечивать возможность передачи свидетельства. Граф, вершинами которого служат фрагменты знаний, а ребра отражают возможность передачи свидетельства между фрагментами знаний, называется вторичной структурой алгебраической байесовской сети. В силу особенностей алгоритмов апостериорного логико-вероятностного вывода [5, 8, 9, 13], вторичной структурой могут выступать только графы с особыми свойствами — графы смежности (в частности: минимальные графы смежности, МГС), причём указанное утверждение ниже будет уточнено дополнительно.

Таким образом, для реализации алгоритмов логико-вероятностного вывода в АБС, необходимо синтезировать ее вторичную структуру (т.е. конкретный минимальный граф смежности, вершинами которого являются фрагменты знаний) по заданной первичной структуре такой сети (т.е. набору фрагментов знаний) [14, 15, 22].

Кроме того, для исследования количественных характеристик результатов работы алгоритмов логико-вероятностного вывода требуется большая выборка графов смежности, на которой можно было бы проводить эксперименты. Несмотря на то, что известны алгоритмы генерации всего множества минимальных графов смежности [16, 17, 18, 19], а также произвольного минимального графа смежности [2, 3, 24], актуальной является также задача построения минимального графа смежности, который выбирается из всего множества минимальных графов смежности согласно распределению вероятности, явно или неявно задаваемому исследователем. К примеру, если в результате работы алгоритма получается граф, который выбран из множества минимальных графов смежности равновероятно, то такой алгоритм может использоваться для реализации генетических алгоритмов, которые могут использоваться, как предполагают авторы, при различных видах автоматического обучения алгебраических байесовских сетей.

Цель статьи — сформировать алгоритм рандомизированного синтеза минимального графа смежности на основании имеющейся первичной структуры алгебраической байесовской сети; при этом должно быть обеспечено условие, что выбор любого возможного для заданной первичной структуры алгебраической байесовской сети минимального графа смежности должен иметь положительную вероятность.

2. Графы и случайные графы. *Графом* называется пара $\langle V, E\rangle$, где V — множество вершин графа, а E — множество ребер. Через V и E принято обозначать функции от графа, возвращающие множество его вершин и множество его ребер соответственно.

Случайные графы ввели в 1959 году Эрдеш и Реньи [30, 31, 32, 33] и одновременно и независимо Гильберт [34]. Ими под случайным графом понимается случайный элемент, заданный на некотором множестве графов с фиксированным числом вершин, алгеброй на множестве графов выступает множество всех подмножеств множества графов с фиксированным числом вершин [1][1].

Случайные графы используются для доказательства детерминистических свойств графов. Так, если показано, что случайный граф обладает некоторым свойством A с ненулевой вероятностью, то должен существовать граф, удовлетворяющий свойству A. Такой подход получил название *вероятностного метода* [29, 36]. Случайные графы используются для моделирования сложных сетей, таких как социальные сети (в т.ч. сети распространения инфекционных заболеваний), телефонные сети, сети интернет, сети цитирований научных публикаций и другие [37]. В сложных и больших сетях практически невозможно описать связи между всеми узлами и, как следствие, невозможно точно представить графическую структуру сети. Поэтому при моделировании графической структуры сложной сети задается вероятностное распределение степени вершины [36]. При моделировании динамики сложной сети предполагается, что число узлов растет во времени согласно некоторым закономерностям [28], поэтому модель случайных графов тесно связана с моделированием случайных процессов.

3. Элементы глобальных структур алгебраической байесовской сети. В данном разделе будет приведен необходимый минимум понятий теории АБС, необходимый для описания алгоритмов синтеза МГС. Изложение будет согласовано с терминологией, введенной в работах [20, 22, 25, 23, 26].

Для краткости изложения зафиксируем первичную структуру АБС, и все вводимые далее объекты будем определять для нее. *Алфа-*

---

[1] Вообще говоря, случайный граф как вид случайного элемента может быть устроен еще более сложно за счет рассмотрения динамики случайных графов, в которых с течением времени увеличивается число вершин. Хотя такая конструкция не используется в настоящей работе, она представляется тесно связанной с другими аспектами автоматического обучения вероятностных графических моделей.

вит — множество атомарных пропозициональных формул: $A = \{x_1, \dots, x_n\}$. В рамках рассмотрения графов смежности будем рассматривать графы, построенные не над фрагментами знаний, а над подалфавитами, над которыми построены идеалы соответствующих фрагментов знаний.

Нагруженным графом называется ненаправленный граф, вершины которого соответствуют элементам первичной структуры АБС, для которого введена функция нагрузки: *нагрузкой* $W(v)$вершины v называется соответствующий ей подалфавит. *Сепаратором* двух вершин называется пересечение нагрузок соответствующих вершин: $\mathrm{Sep}(v, u) = W(v) \cap W(u)$. Множество сепараторов будем обозначать как $\mathrm{Separator}$. Две вершины называются *сочлененными*, если их сепаратор непуст. *Согласованным нагруженным графом* называется нагруженный граф, в котором нагрузка ребра между вершинами равна их сепаратору: $W\big((u, v)\big) = \mathrm{Sep}(u, v)$.

Граф максимальных фрагментов знаний (*граф МФЗ*) — согласованный нагруженный граф, в котором ребра возможны только между сочлененными вершинами (т. е. нагрузка любого ребра непуста). *Магистральный путь* между двумя вершинами u и v в графе МФЗ — такой путь $(u = w_0, w_1, \dots, w_{n-1}, w_n = v)$, что нагрузка любой вершины содержит сепаратор u и v: $\forall w_i\colon \mathrm{Sep}(u, v) \subset W(w_i)$. Согласованный нагруженный граф *магистрально связен*, если между каждой парой сочлененных вершин существует магистральный путь.

Граф смежности — магистрально связный граф МФЗ. В графе смежности возможны ребра только между сочлененными вершинами и между любой парой сочлененных вершин существует магистральный путь. *Минимальный граф смежности* (*МГС*) — граф смежности, число ребер которого минимально. *Максимальный граф смежности* — граф смежности, число ребер которого максимально.

Сильное сужение $\downarrow U$ на сепаратор U, это граф, в который входят все вершины первичной структуры, нагрузка которых содержит U, и ребра проведены между каждой парой вершин, сепаратор которых строго содержит U.Сильное сужение $\downarrow U$ разбивается на компоненты связности, которые называются *владениями*.

Сначала рассмотрим произвольный МГС G и произвольный сепаратор U. Известно, что ребра G, нагрузка которых равна U, задают дерево на владениях сильного сужения $\downarrow U$ [22, 25]. Затем рассмотрим набор ребер, образующих дерево на владениях сильного сужения $\downarrow U$. Граф, множество ребер которого совпадает с рассмотренным, а множество вершин совпадает с множеством концов этих ребер, называется *жилой с нагрузкой* U. Поскольку это определение играет важную роль в задаче синтеза МГС, введем эквивалентное определение: *жила с нагрузкой* U — граф, множество ребер которого представляет набор ребер с нагрузкой U, число которых равно числу компонент сильного сужения $\Downarrow U$, уменьшенного на единицу, а граф, полученный добавлением к $\Downarrow U$ ребер жилы, является связным [26], и каждая вершина жилы имеет ненулевую степень.

Рассмотрим кортеж жил, где для каждого сепаратора встречается ровно одна жила с такой нагрузкой. Объединение такого кортежа называется *пучком*. Известна следующая теорема:

Теорема 1 (о множестве МГС) [22, 25**].** Множество минимальных графов смежности совпадает с множеством пучков.

Теорема о множестве МГС используется для построения как множества МГС, так и произвольного МГС. В первом случае следует строить все возможные жилы над владениями $\Downarrow U$, во втором случае — либо произвольные, либо обладающие конкретными свойствами. Подробнее это будет рассмотрено в следующем разделе.

4. Алгоритм рандомизированного синтеза минимального графа смежности. Алгоритм рандомизированного синтеза минимального графа смежности строит по первичной структуре АБС случайный минимальный граф смежности:

1. Построить множество $\mathrm{Separator} = \{U_1, U_2, \dots, U_n\}$ всех сепараторов первичной структуры АБС.
2. Для каждого $U_a \in \mathrm{Separator}$ построить сильное сужение $\Downarrow U_a$.
3. Разложить сужения $\Downarrow U_a$на компоненты связности (владения) $\{c_1^a, c_2^a, \dots, c_{m_a}^a\}$.
4. Перебрать все владения, на каждом шаге добавляя ребро между какой-либо вершиной v рассматриваемого владения и вершиной v', лежащей в одном из прежде выбранных владений. Набор ребер и их концы образуют жилу $S_{U_a}^{l_a}$.
5. Объединить жилы $S_{U_1}^{l_1}, S_{U_2}^{l_2}, \dots, S_{U_n}^{l_n}$.

Рандомизация синтеза минимального графа смежности заключается в случайном выборе последовательности перебора владений для каждого сужения, а также в случайном выборе пар вершин v и v'.

Граф, полученный на последнем шаге работы алгоритма, является минимальным графом смежности, по теореме о множестве минимальных графов смежности.

Для корректного описания алгоритма необходимо ввести следующие обозначения.

- **Components** для заданного графа возвращает множество подграфов, являющихся его компонентами связности;
- **ChooseRandom** выбирает случайным образом элемент из заданного множества.

RandomizedMJG

```
Require: Workloads
Ensure: ∄w₁, w₂ ∈ Workloads: w₁ ⊆ w₂
Separators ← ∅
for all w₁, w₂ ∈ Workloads, w₁ ≠ w₂ do
  Separators ← Separators ∪ {w₁ ∩ w₂}
end for
Edges ← ∅
forall u ∈ Separators do
  SNWorkloads ← ∅
  for all w ∈ SNWorkloads do
    if u ⊂ w do
      SNWorkloads ← SNWorkloads ∪ {u}
    end if
  end for
  SNEdges ← ∅
  for all v₁, v₂ ∈ SNWorkloads, v₁ ≠ v₂ do
    if w ∈ v₁ ∩ v₂ do
      SNEdges ← SNEdges ∪ {(v₁, v₂)}
    end if
  end for
  Holdings ← COMPONENTS(⟨SNWorkloads, SNEdges⟩)
  H ← CHOOSERANDOM(Holdings)
  Holdings ← Holdings\{H}
  AlreadyConnected ← V(H)
  Edges ← ∅
  while Holdings ≠ ∅ do
    H ← CHOOSERANDOM(Holdings)
    v ← CHOOSERANDOM(AlreadyConnected)
    u ← CHOOSERANDOM(H)
    Edges ← Edges ∪ {v, u}
    NotYetConnected ← NotYetConnected\{H}
    AlreadyConnected ← AlreadyConnected ∪ V(H)
```

```
31:   end while
32: end for
33: return Workloads, Edges
```

Листинг 1. Алгоритм рандомизированного синтеза минимального графа смежности.

На шагах (2–4) строится множество сепараторов.

В цикле (6–23) перебираются все сепараторы, и для каждого строится множество ребер жилы. На шагах (7–18) сильное сужение последовательно строятся множества вершин и ребер такого сужения. На шаге (19) строится множество за счет поиска компонент связности в сильном сужении. На шагах (20–21) выбирается случайное владение, которое удаляется из этого множества, а его вершины на шаге (22) делегируются строящемуся компоненту связности. В цикле (24–31), который выполняется до тех пор, пока множество владений не опустеет, на шаге (25) выбирается случайное владение, в котором на шаге (26) выбирается случайная вершина, которая на шаге (28) соединяется ребром со случайно выбранной на шаге (27) вершиной из уже строящегося компонента связности. На шаге (29) выбранное владение удаляется из множества, а его вершины на шаге (30) добавляются к множеству вершин строящегося компонента связности.

Утверждение. Алгоритм рандомизированного синтеза минимального графа смежности строит минимальный граф смежности.

Доказательство. Докажем, что в ходе работы цикла (24–30) для выбранного циклом (6–31) сепаратора U строится множество ребер жилы. В ходе работы цикла строится множество ребер, число которых равно числу владений, уменьшенному на единицу, и которые при последовательном добавлении к соответствующему сильному сужению сделают граф связным, поскольку каждое ребро уменьшает число компонент связности на один. Таким образом, в ходе работы цикла (21–30) к множеству Edges добавляется множество ребер какой-то жилы с нагрузкой U. По теореме 1, объединение таких жил, выбранных по одной для каждого сепаратора, является минимальным графом смежности.

Проверим, что любой граф из множества минимальных графов смежности может быть получен при помощи алгоритма рандомизированного синтеза минимального графа смежности.

Теорема 2. Каждый элемент множества минимальных графов смежности может быть с ненулевой вероятностью получен в результате работы алгоритма[2].

Доказательство. Пусть G является элементом множества минимальных графов смежности. Каждый минимальный граф смежности представляет объединение жил выбранных по одной для каждого сепаратора. Обозначим случайный граф, получаемый в результате работы всего алгоритма как $\boldsymbol{G}$. Поскольку жилы строятся независимо, достаточно показать, что каждая возможная жила строится с ненулевой вероятностью. Рассмотрим произвольную такую жилу S_{U_a}.

В ходе работы цикла (24–30) последовательно перебираются владения, то есть явно задается перестановка этих владений π_a порядка m_a. Эта перестановка выбирается случайным образом, причем вероятность выбрать каждую такую перестановку ненулевая. Возможность выбора конкретных ребер зависит только от перестановки.

Всего возможно $m_a!$ таких перестановок. Для выбранной жилы можно указать $(m_a - 1)!$ таких перестановок, при выборе которых эта жила может быть построена, т. е. для этих перестановок существует возможность выбрать ребра этой жилы.

Пусть $\boldsymbol{\pi}$ есть случайная величина, которая является генератором случайной перестановки порядка $m_a, \forall a \in \{1..n\}$, $\boldsymbol{S_{U_a}}$ есть случайная величина, генерирующая жилу в алгоритме.

$$P\left(\boldsymbol{G} = \bigcup_{a=1}^{n} S_{U_a}^{l_a}\right) = \sum_{a=1}^{n} P(\boldsymbol{\pi_a} = \pi_a) P\left(\boldsymbol{S_{U_a}} = S_{U_a}^{l_a} \middle| \boldsymbol{\pi_a} = \pi_a\right).$$

Так как существуют перестановки π_a, такие, что $P\left(\boldsymbol{S_{U_a}} = S_{U_a}^{l_a} \middle| \boldsymbol{\pi_a} = \pi_a\right)$ больше нуля, и вероятность любой такой перестановки также больше нуля то каждая конкретная жила $S_{U_a}^{l_a}$ при работе алгоритма генерируется с ненулевой вероятностью.

Значит, искомая вероятность $P\left(\boldsymbol{G} = \bigcup_{a=1}^{n} S_{U_a}^{l_a}\right) > 0$.

Таким образом, полученный в результате работы алгоритма минимальный граф смежности является реализацией случайного минимального графа смежности.

5. Заключение. В работе приведен алгоритм рандомизированного синтеза минимального графа смежности, то есть автоматической генерации с ненулевой вероятностью реализации случайного минимально-

[2]Иначе говоря, реализация любого элемента множества минимальных графов смежности имеет строго положительную вероятность при описанном алгоритме.

го графа смежности, неявно задавая сам случайный граф смежности. Возможность строить минимальные графы смежности случайным образом является важной в двух аспектах: во-первых, создания тестовых выборок для получения числовых характеристик работы алгоритмов глобального логико-вероятностного вывода, а во-вторых, для привлечения генетических алгоритмов для решения задачи автоматического обучения (в том числе, синтеза оптимальной вторичной структуры) алгебраической байесовской сети.

Естественно, что «случайность» — эмпирическое распределение вероятностей, складывающееся в результате вычислительных экспериментов — минимальных графов смежности напрямую зависит от используемых алгоритмов генерации случайных чисел, которые на самом деле являются псевдослучайными.

Отметим, что распределение, которое неявным образом задается приведенным алгоритмом, во-первых, не выписано явно, во-вторых, в общем случае не является равномерным, поэтому актуальным является получение алгоритма, который бы строил минимальные графы смежности с одинаковой вероятностью, а еще лучше — с произвольным наперед заданным распределением вероятностей.

Литература

1. *Райгородский А.М.* Модели случайных графов и их применение // Тр. МФТИ. 2010. Т.2. №4. С. 130–140.
2. *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.
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. *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.
5. *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).
6. *Тулупьев А.Л.* Алгебраические байесовские сети. Логико-вероятностный подход к моделированию баз знаний с неопределенностью. СПб.: СПИИРАН, 2000. 292 с.
7. *Тулупьев А.Л.* Алгебраические байесовские сети: локальный логико-вероятностный вывод: Учеб.пособие. СПб.: СПбГУ; ООО Издательство «Анатолия», 2007. 80 с. (Сер.Элементы мягких вычислений).
8. *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.

9. *Тулупьев А.Л.* Байесовские сети: логико-вероятностный вывод в циклах. СПб.: Изд-во С.-Петербургского ун-та, 2008. 140 с. (Элементы мягких вычислений).
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., Николенко С.И., Сироткин А.В.* Байесовские сети: логико-вероятностный подход. СПб.: Наука, 2006. 607 с.
12. *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.
13. *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.
14. *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.
15. *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.
16. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи самоуправляемых клик // Труды СПИИРАН. 2010. Вып. 1 (12). С. 119–133.
17. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи самоуправляемых клик-собственников // Труды СПИИРАН. 2010. Вып. 3 (14) С. 150–169.
18. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи клик владений // Труды СПИИРАН. 2010. Вып. 2 (13). С. 119–133.
19. *Фильченков А.А.* Алгоритм построения множества минимальных графов смежности при помощи клик-собственников владений // Труды СПИИРАН. 2010. Вып. 4 (15). С. 193–212.
20. *Фильченков А.А.* Иерархия глобальных структур алгебраической байесовской сети как система графов и гиперграфов // Научно-технический вестник информационных технологий, механики и оптики. 2013. Вып. 1(83). С. 75–80.
21. *Фильченков А.А.* Меры истинности и вероятностные графические модели для представления знаний с неопределенностью // Труды СПИИРАН. 2012. Вып. 4(23). С. 254–295.
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 байесовской сети. // Труды СПИИРАН. 2011. Вып. 18. С. 164–187.
24. *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.
25. *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.
26. *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.

27. *Ширяев А.Н.*Вероятность. М.: Наука, 1989. 640 с.
28. *Barabási A.-L., Albert R.* Emergence of Scaling in Random Networks// Science. Vol. 286. №.5439. P. 509-512.
29. *Bollobás B.* Random Graphs. 2nd ed. Cambridge University Press. 2001. 498 p.
30. *Erdős P., Rényi A.* On random graphs. I // Publ. Math.1959. P. 290–297.
31. *Erdős P., Rényi A.* On the evolution of random graphs // Magyar Tud. Akad. Mat. Kutató Int. Közl. 1960. №5. P. 17–61.
32. *Erdős P., Rényi A.* On the evolution of random graphs // Bull. Inst. Internat. Statist. 1961. №38. P. 343–347.
33. *Erdős P., Rényi A.* On the strength of connectedness of a random graph // Acta Math. Acad. Sci. Hungar. 1961. №12. P. 261–267.
34. *Gilbert E.N.* Random Graphs // Annals of Mathematical Statistics.1959. № 30 (4). P. 1141–1144.
35. *Koller D., Friedman N.* Probabilistic Graphical Models: Principles and Techniques . The MIT Press, 2009. 1208 p.
36. *van der Hofstad R.* Random graphs and complex networks. Lecture notes. // Available on http://www.win.tue.nl/rhofstad/NotesRGCN.pdf
37. *Strogatz S.H.* Exploring complex networks // Nature. 2001. №410. P. 268–276. doi:10.1038/35065725

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 обучении вероятностных графических моделей. Число научных публикаций — 90. 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 — 90. 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.

Мусина Валерия Фуатовна — младший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН, студент магистратуры экономического факультета СПбГУ. Область научных интересов: случайные процессы, вероятностное и статистическое моделирование, биостатистика, вероятностные графические модели. Число научных публикаций — 13. valery.musina@gmail.ru, 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.

MusinaValeriyaFuatovna — junior research fellow Theoretical and Interdisciplinary Computer Science Laboratory, SPIIRAS, graduate student of Faculty of Economics at Saint Petersburg State University. Research area: stochastic processes, probabilistic and statistic modelling, biostatistics, probabilistic graphical models. Number of publications — 13. valery.musina@gmail.ru, 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.

Тулупьев Александр Львович — д.ф.-м.н., доцент; заведующий лабораторией теоретических и междисциплинарных проблем информатики СПИИРАН, профессор 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 социо-культурных исследованиях, применение методов биостатистики и математического моделирования в эпидемиологии, технология разработки программных комплексов с СУБД. Число научных публикаций — 250. ALT@iias.spb.su; 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., Associate Professor; Head of Theoretical and Interdisciplinary Computer Science Laboratory, SPIIRAS, 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 исследований. Работа выполнена при финансовой поддержке РФФИ, гранты № 12-01-00945-а, 12-01-31202-мол_а.

Рекомендовано ТиМПИ СПИИРАН, зав. лаб. А.Л. Тулупьев, д.ф.-м.н., доцент.
Статья поступила в редакцию 15.02.2013.

РЕФЕРАТ

Фильченков А.А., Мусина В.Ф., Тулупьев А.Л. **Алгоритм рандомизированного синтеза минимального графа смежности.**

Одной из основных задач логико-вероятностного вывода в теории алгебраических байесовских сетей является апостериорный вывод — получение вероятностных оценок истинности пропозиций при условии поступивших свидетельств. В силу особенностей алгоритмов апостериорного логико-вероятностного вывода, вторичной структурой АБС могут выступать только графы с особыми свойствами — графы смежности (в частности: минимальные графы смежности, МГС). Для реализации этих алгоритмов, необходимо синтезировать ее вторичную структуру (т.е. конкретный МГС, вершинами которого являются фрагменты знаний) по заданной первичной структуре такой сети (т.е. набору фрагментов знаний).

Цель статьи — сформировать алгоритм рандомизированного синтеза минимального графа смежности на основании имеющейся первичной структуры алгебраической байесовской сети; при этом должно быть обеспечено условие, что выбор любого возможного для заданной первичной структуры алгебраической байесовской сети минимального графа смежности должен иметь положительную вероятность.

В статье сформирован алгоритм рандомизированного синтеза минимального графа смежности по первичной структуре АБС. Показано, что сформированный алгоритм строит минимальный граф смежности. Кроме того доказано, что каждый элемент множества минимальных графов смежности может быть с ненулевой вероятностью получен в результате работы алгоритма.

Возможность строить минимальные графы смежности случайным образом является важной в двух аспектах: во-первых, создания тестовых выборок для получения числовых характеристик работы алгоритмов глобального логико-вероятностного вывода, а во-вторых, для привлечения генетических алгоритмов для решения задачи автоматического обучения (в том числе, синтеза оптимальной вторичной структуры) алгебраической байесовской сети.

SUMMARY

Filchenkov A.A., Musina V.F, Tulupyev A.L. **Algorithm for minimal join graph's randomized synthesis.**

A posterior inference is one of the main problems of logic-probabilistic inference in the theory of algebraic Bayesian networks (ABN). A posterior inference is a process of recalculation of truth probability estimates of propositions given the probability of incoming evidence. Special features of the logic-probabilistic inference algorithms require graphs with special characteristics — join graphs (in particular minimal join graphs, MJG). One needs to synthesize secondary structure (i.e. specific MJG with knowledge patterns as vertices) based on known primary structure of the network (i.e. set of knowledge patterns) to apply this algorithms.

Purpose of the paper is formulation of the minimal join graph randomized synthesis algorithm on the base of known primary structure of ABN. The condition should be ensured that every MJG probable for the primary structure of ABN have positive probability of being selected.

In the paper the algorithm for minimal join graph randomized synthesis based on known primary structure was formulated. It was showed that this algorithm constructs minimal join graph. Besides this it was proved that every element of the MJGs' set can be constructed by the algorithm with non-zero probability.

The possibility of randomized construction of minimal join graphs is important in two angles: firstly in randomized test sampling for investigation of numerical calculating characteristics of global logic-probabilistic inference algorithms; secondly in using genetic algorithms in automated ABN's learning problem solving (including problem of optimal secondary structure synthesis).