

УДК 004.8

СУБОПТИМАЛЬНАЯ ЗВЕЗДЧАТАЯ СТРУКТУРА АЛГЕБРАИЧЕСКОЙ БАЙЕСОВСКОЙ СЕТИ

А. А. Фильченков¹,

аспирант

Санкт-Петербургский государственный университет

Выделен подкласс минимальных графов смежности — звездчатые графы смежности. Доказано, что множество таких графов содержит оптимальные вторичные структуры, т. е. графы смежности с минимальным диаметром. Представлен алгоритм синтеза такого множества. На основе этого предложен способ ускорения синтеза оптимальной вторичной структуры алгебраической байесовской сети.

Ключевые слова — алгебраическая байесовская сеть, звездчатый граф, обучение глобальной структуры, граф смежности, машинное обучение, субоптимальная вторичная структура.

Введение

Алгебраическая байесовская сеть (АБС) — активно развивающаяся теоретико-компьютерная модель, относящаяся к классу логико-вероятностных графических моделей, которая позволяет использовать как скалярные, так и интервальные оценки вероятности для представления неопределенности [1–6]. АБС может быть применима как одна из математических моделей анализа информационной безопасности [7], в частности, для оценки защищенности системы от социо-инженерных атак [8]. Кроме того, АБС могут применяться в задаче распознавания изображений [9] — традиционной области, в которой используются вероятностные графические модели [10]. Обработка интервальных оценок также оказывается востребованной в анализе гранулярных данных о рекордных интервалах между последними эпизодами [11]. Известные алгоритмы глобального логико-вероятностного вывода для АБС опираются на ее вторичную структуру, которая представляет собой граф, построенный над ее первичной структурой [1, 2, 6, 12]. Первичная структура АБС в свою очередь представляет собой набор математических моделей фрагментов знаний (далее — фрагментов знаний) — сложным образом устроенном представлении случайных элементов.

¹ Научный руководитель — доктор физико-математических наук, доцент, заведующий лабораторией теоретических и междисциплинарных проблем информатики СПИИРАН А. Л. Тулупьев.

Вторичной структурой могут выступать только особые графы — минимальные (по числу ребер) графы смежности [6, 12–14], формальное определение которых будет дано в следующем разделе. В силу специфики алгоритмов логико-вероятностного вывода, предложенных в работах [1, 2, 6, 12], одной из важных характеристик минимальных графов смежности является их диаметр. Глобальное обучение АБС [15] — это один из видов машинного (автоматического) обучения, развитие методов, моделей и алгоритмов которого — одна из самых актуальных задач в искусственном интеллекте [6, 16, 17]. В рамках задачи обучения глобальной структуры было бы желательно строить графы смежности с минимальным диаметром [18]. Следует отметить, что кроме минимальности диаметра существуют и другие требования [18], поэтому минимальные графы смежности с минимальным диаметром являются оптимальной вторичной структурой.

Построение минимального графа смежности, одновременно минимального по диаметру, возможно за счет построения всего множества минимальных графов смежности и выбора из него графа с минимальным диаметром. В рамках данной работы мы рассмотрим звездчатые графы — особый класс минимальных графов смежности, который, как будет показано, характеризуется тем, что множество звездчатых графов смежности содержит в себе некоторые минимальные графы смежности с минимальным диаметром. Мощность этого множества меньше, чем мощность множества всех минимальных графов смежности, поэто-

му поиск минимального графа смежности с минимальным диаметром в таком множестве будет быстрее. Будет также предложен алгоритм синтеза указанного множества, который благодаря свойствам звездчатых графов отличается от алгоритма синтеза множества всех минимальных графов смежности лишь небольшими изменениями.

Графы смежности

В качестве вторичной структуры АБС выступает минимальный (по числу ребер) граф смежности. Дадим необходимые понятия, следуя введенной в работах [13, 14, 19] терминологии.

Согласованный нагруженный граф G — граф $G = (V, E)$ без петель, на вершинах и ребрах которого задана функция нагрузки W , множество значений которой состоит из подмножеств некоторого алфавита A :

$$W : V \cup E \rightarrow 2^A,$$

причем

$$\forall e = (u, v) \in E \quad W(e) = W(u) \cap W(v).$$

Сепаратором двух вершин называется пересечение их нагрузок. В согласованном нагруженном графе, как видно из определения, вес ребра равен сепаратору его концов. Множество сепараторов для данной первичной структуры обозначим как *Sepарator*. Две вершины называются *сочлененными*, если их сепаратор непуст.

Магистральный путь между двумя вершинами согласованного нагруженного графа — это такой ненаправленный путь между ними, на котором нагрузка любой его вершины содержит сепаратор исходных вершин. Согласованный нагруженный граф магистрально связан, если между каждой парой сочлененных вершин существует магистральный путь, но не обязательно ребро.

Граф смежности — это магистрально связанный согласованный нагруженный граф, ребра в котором возможны только между сочлененными вершинами (т. е. нагрузка ребер непуста). *Минимальный граф смежности* — минимальный по числу ребер граф смежности.

В рамках исследования графов смежности вместо математических моделей фрагментов знаний (достаточно сложных конструкций, описание которых можно найти в работах [5, 6]) обычно рассматривается база таких моделей — подалфавиты некоторого алфавита A . Набор таких подалфавитов PS образует разбиение A , причем подалфавиты не поглощают друг друга:

$$PS = \{A_i\}_{i=1, \dots, n} : \bigcup_{i=1..n} A_i = A \text{ и} \\ \forall A_i, A_j \in PS, i \neq j \Rightarrow A_i \not\subseteq A_j.$$

Введенный таким образом PS задает значения функций нагрузки (либо просто нагрузку) для вершин графа. Задачей глобального обучения вторичной структуры АБС является построение минимального графа смежности для заданного множества вершин — именно на них осуществляются алгоритмы логико-вероятностного вывода.

Синтез множества минимальных графов смежности

Далее необходимо ввести систему терминов, связанную с синтезом множества минимальных графов смежности, поскольку при помощи нее будут определены звездчатые графы смежности. Будем следовать работам [13, 19, 20].

Для каждого сепаратора определим множество его *сыновей*: U' является сыном сепаратора U тогда и только тогда, когда $U' \subset U$ и $\nexists U'' : U \subset U'' \subset U'$. Другими словами, если на множестве сепараторов задан частичный порядок, индуцированный отношением включения, то сыновья для сепаратора будут множеством его последователей (в терминах теории частично-упорядоченных множеств). Множество сыновей сепаратора U будем обозначать как $Sons(U)$.

Сужение $G \downarrow A'$ согласованного нагруженного графа G на подалфавит A' — это ненаправленный граф, в который входят только те вершины и ребра исходного графа G , нагрузки которых содержат или равны A' :

$$G \downarrow A' = (\{V_i \mid V_i \in V(G), A' \subseteq W(V_i)\}; \\ \{E_i \mid E_i \in E(G), A' \subseteq W(E_i)\}).$$

Сильное сужение $G \downarrow U$ — сужение $G \downarrow U$ на сепаратор U , из которого удалили все ребра с нагрузкой U :

$$G \downarrow U = (\{V_i \mid V_i \in V(G), U \subseteq W(V_i)\}; \\ \{E_i \mid E_i \in E(G), U \subset W(E_i)\}).$$

Рассмотрим *максимальный по числу ребер* граф смежности G_{\max} . Сильное сужение $G_{\max} \downarrow U$ разбивается на компоненты связности, которые называются *владениями* для нагрузки U , их будем обозначать как $Holdings(U)$. Вершина, которая образует одноэлементное владение для нагрузки U , называется *доменной вершиной* для этой нагрузки. Известно, что множество вершин во владении для нагрузки U является либо множеством вершин, входящим в сужение G_{\max} на какого-либо ее сына, либо объединением таких множеств, либо доменной вершиной.

Жила S_U с нагрузкой U — набор ребер с нагрузкой U , число которых равно числу владений для нагрузки U , уменьшенному на единицу,

а граф, полученный добавлением к $G_{\max} \downarrow U$ ребер из S_U , является связным: $S_U \subseteq E(G_{\max})$:

- 1) $\forall e \in S_U \ W(e) = U$;
- 2) $|S_U| = |\text{Holdings}(U)| - 1$;
- 3) граф $((V(G_{\max} \downarrow U), E(G_{\max} \downarrow U) \cup S_U))$ связан.

В определенном смысле жила является деревом на владениях (строгая формализация этого рассуждения, не требующаяся в данной работе, приведена в работах [13, 14]).

Теорема 1 (о минимальных графах смежности) [13, 14]. Граф смежности является минимальным тогда и только тогда, когда множество ребер каждого значимого веса является в нем жилой.

Теорема о минимальных графах смежности создает основу для конструктивного представления множества минимальных графов смежности как декартова произведения множеств жил, построенных для каждого сепаратора. Последнее лежит в основе алгоритмов синтеза такого множества, схема которого будет приведена ниже.

Звездчатый граф смежности

Введем теперь определение звездчатого графа смежности. *Звездой* называется дерево, одна из вершин которого является концом каждого ребра. Такая вершина называется *центром звезды*, а диаметр такого графа равен двум. Верно также и то, что любое дерево с диаметром, равным двум, является звездой.

Звездчатым графом смежности будем называть граф, в котором каждая жила является звездой. Поскольку звезда является деревом, то, согласно теореме о минимальных графах смежности, звездчатый граф смежности является минимальным. *Звездчатым графом смежности* будем называть такой минимальный граф смежности, что:

- 1) любая его жила является звездой;
- 2) концы любого ребра в любой жиле с нагрузкой U являются либо доменными вершинами для этой нагрузки, либо центрами жилы с нагрузкой U' , $U' \in \text{Sons}(U)$.

Теорема 2 (о структуре жил графа смежности). В жиле графа смежности не существует трех вершин, попарно соединенных ребрами одного веса.

Доказательство: Пускай существуют три такие вершины, попарно соединенные ребрами одной жилы. Они входят не более чем в три разных владения. По определению, в жиле число ребер на единицу меньше, чем число владений. Владения, в которые входят эти вершины, соединенные тремя ребрами, представляют компонент связности. Рассмотрим другое владение. Укажем ребро, которое соединяет его с указанным компонентом связности. Добавим это владение и это ребро к уже рассмотренным компонентам связности. Будем повторять эту операцию со всеми вла-

дениями, добавляя каждое из них вместе с ребром к компоненту связности. Когда мы станем рассматривать последний компонент связности, то окажется, что мы рассмотрели уже достаточное для жилы число ребер, ни одно из которых не соединяет это владение с оставшимися. Следовательно, предположение неверно.

Теорема 3 (о звездчатом графе смежности с минимальным диаметром). Существует минимальный граф смежности с минимальным диаметром, который является звездчатым.

Доказательство: Рассмотрим произвольный минимальный граф смежности G с минимальным диаметром и будем последовательно преобразовывать его к звездчатому графу. Для этого рассмотрим произвольную жилу U графа G . В ней рассмотрим все кратчайшие пути между двумя вершинами в G , длина которых максимальна среди таких путей (и по определению равна диаметру), которые содержат ребра веса U . Множество таких путей обозначим как Shortest_U .

Выберем произвольный путь P из Shortest_U , возьмем какой-либо из двух его концов (назовем его s) и будем последовательно рассматривать вершины P , начиная с выбранного конца. Первую вершину, из которой выходит ребро нагрузки U , обозначим g . Последнюю вершину, в которую входит ребро нагрузки U , обозначим g' . Другой конец пути обозначим s' . Вершины g и g' будем называть *воротами*. Если бы мы обходили путь в обратном направлении, то ворота остались бы прежними, поменялся бы лишь порядок их прохода. По построению все ребра нагрузки U , принадлежащие пути P , принадлежат и пути между g и g' . Ненаправленные пути между s и g и между g' и s' будем называть *тупиками*. Тупики не содержат ребер с нагрузкой U . Тупику мы можем однозначно сопоставить ворота, являющиеся его концом.

Теперь будем для каждого сепаратора менять жилу соответствующей нагрузки на жилу, являющуюся звездой (на самом деле, семейство таких жил), таким образом, чтобы не увеличить диаметр.

Пускай мы на данном шаге выбрали сепаратор U . Рассмотрим множество тупиков для данного сепаратора, среди них выберем один наибольшей длины. Выбранный тупик обозначим b_1 , а его длину — n_1 . Ворота тупика b_1 сделаем центром звезды, соединяя ее ребрами нагрузки U с произвольными вершинами из других владений. Рассмотрим, как изменилась длина кратчайших путей, содержащих ребра веса U . Каждый такой путь состоял из двух тупиков (например, b_2 длины n_2 и b_3 длины n_3), и длина такого пути была не менее $n_2 + n_3 + 1$, поскольку этот путь должен содержать не менее одного ребра веса U , которые не входят в эти тупики. Соответственно, любой путь, содержащий тупик b_1 , по построению со-

держит ровно одно ребро веса U , поэтому его длина не увеличилась.

Предположим, что увеличилась длина кратчайшего пути, не содержащего b_1 (т. е. ни один из тупиков b_2 и b_3 не совпадает с b_1). Обозначим исходный путь $p_{2,3}$, а получившийся — $p'_{2,3}$. Ворота, соответствующие b_2 и b_3 , обозначим g_2 и g_3 . По построению длина $p'_{2,3}$ равна $n_2 + n_3 + 2$, следовательно, длина $p_{2,3}$ была равна $n_2 + n_3 + 1$, следовательно, в G между g_2 и g_3 было ребро. Согласно лемме, в G не могло быть ребер одновременно между g_1 и g_2 , между g_1 и g_3 . Пускай ребра не было между g_1 и g_2 , следовательно, путь между ними состоял не менее чем из двух ребер. Следовательно, путь, включающий тупики b_1 и b_2 , в G имел длину не менее $n_1 + n_2 + 2$, что в силу максимальной n_1 не меньше, чем $n_2 + n_3 + 2$, откуда следует, что $p_{2,3}$ не был максимальным по длине. Следовательно, замена жилы на звездчатую не изменит длины максимальных путей.

Проведя замену жил для каждого сепаратора, мы приходим к звездчатому графу смежности, диаметр которого не больше, чем диаметр рассмотренного минимального графа смежности с минимальным диаметром, а, следовательно, минимален.

Схемы синтеза множеств минимальных графов смежности и звездчатых графов смежности

Звездчатые графы смежности представляют интерес, в первую очередь, потому, что синтез множества таких графов похож на синтез множества минимальных графов смежности, поскольку основан на конструктивном представлении таких графов.

Сначала приведем схему алгоритма синтеза множества минимальных графов смежности. Следует указать, что подобных алгоритмов существует значительное число, но приведем лишь наиболее общий, а затем покажем, как за счет внесения небольших изменений можно привести его к алгоритму синтеза множества звездчатых графов смежности. Подобные изменения будут применимы и для любого другого алгоритма синтеза множества минимальных графов смежности.

На вход такого алгоритма подается множество взаимно непоглощающих подалфавитов некоторого алфавита A , соответствующих первичной структуре алгебраической байесовской сети.

1. Построить множество Separator, содержащее все сепараторы.

2. Для каждого сепаратора U построить соответствующее ему сильное сужение, найти компоненты связности такого сильного сужения H_U .

3. Для каждого набора таких владений H_U перебрать все возможные деревья $T_{U,i}$, построенные над владениями H_U как над вершинами.

4. Для каждого такого дерева $T_{U,i}$ построить все возможные жилы $S_{U,I,j}$, ребра в которых соединяют вершины из двух владений в том и только том случае, если подобные владения соединены ребром в $T_{U,i}$.

5. Объединить все жилы, построенные для одного сепаратора, в множество S_U .

6. Перебрать все возможные кортежи жил, выбранных по одной для каждого сепаратора.

Согласно теореме о минимальных графах смежности, объединение жил в каждом кортеже является множеством ребер какого-то минимального графа смежности, причем каждому минимальному графу смежности соответствует один из кортежей подобного вида. Следовательно, будет построено множество минимальных графов смежности.

Схема алгоритма синтеза множества звездчатых графов совпадает с приведенной выше схемой во всех шагах, кроме четвертого и пятого, поэтому приведем только их.

4. Для каждого набора построенных на шаге 3 владений H_U перебрать все возможные звезды $Z_{U,i}$, построенные над владениями H_U как над вершинами.

5. Для каждой такой звезды $Z_{U,i}$ построить все возможные жилы $Z_{U,I,j}$, ребра в которых соединяют вершины из двух владений в том и только том случае, если подобные владения соединены ребром в $Z_{U,i}$, причем один из концов каждого ребра является одной и той же вершиной (лежащей во владении, соответствующем центру $Z_{U,i}$).

Результатом работы алгоритма станет множество звездчатых графов.

Поскольку все известные алгоритмы синтеза множества минимальных графов смежности различаются в первых трех шагах, то каждый из них можно использовать в качестве основы для синтеза множества звездчатых графов смежности.

Заключение

В работе рассмотрен особый класс минимальных графов смежности — звездчатых графов смежности, содержащий минимальные по диаметру графы смежности. При отсутствии точных алгоритмов построения минимального по диаметру графа смежности поиск такого графа эффективнее осуществлять не среди вообще всех минимальных графов смежности, а среди именно звездчатых графов смежности, которые выступают субоптимальными структурами, т. е. структурами, множество которых содержит оптимальные структуры. Таким образом, представленный в работе результат развивает теорию алгебраических байесовских сетей в срезе более общей проблемы синтеза интеллектуальных систем, основанных на вероятностных графических моде-

лях, — автоматического обучения глобальной структуры таковых.

Для решения задачи синтеза графа смежности с минимальным диаметром необходимо предложить конструктивное описание таких графов. Исследования в этом направлении показывают, что, вероятно, минимальность диаметра не является локальным свойством, т. е. графы с минимальным диаметром не могут быть описаны через особого

вида жилы минимальных графов смежности. Поэтому в рамках решения той же задачи актуален вопрос поиска подкласса минимальных графов смежности, который меньше, чем звездчатые графы смежности, но обладает описанным выше свойством локальности и содержит хотя бы один граф смежности с минимальным диаметром.

Работа выполнена при поддержке РФФИ, гранты № 12-01-00945-а и 12-01-31202-мол_а.

Литература

1. Тулупьев А. Л. Алгебраические байесовские сети: глобальный логико-вероятностный вывод в деревьях смежности: учеб. пособие. — СПб.: СПбГУ; Анатолия, 2007. — 40 с. (Сер. Элементы мягких вычислений.)
2. Тулупьев А. Л. Алгебраические байесовские сети: система операций глобального логико-вероятностного вывода // Информационно-измерительные и управляющие системы. 2010. № 11. С. 65–72.
3. Тулупьев А. Л. Апостериорные оценки вероятностей в алгебраических байесовских сетях // Вестн. С.-Петербург. ун-та. Сер. 10. 2012. Вып. 2. С. 51–59.
4. Тулупьев А. Л. Байесовские сети: логико-вероятностный вывод в циклах. — СПб.: Изд-во СПбГУ, 2008. — 140 с. (Сер. Элементы мягких вычислений.)
5. Тулупьев А. Л., Николенко С. И., Сироткин А. В. Байесовские сети: логико-вероятностный подход. — СПб.: Наука, 2006. — 607 с.
6. Тулупьев А. Л., Сироткин А. В., Николенко С. И. Байесовские сети доверия: логико-вероятностный вывод в ациклических направленных графах. — СПб.: Изд-во СПбГУ, 2009. — 400 с.
7. Котенко И. В., Степашкин М. В., Юсупов Р. М. Математические модели, методы и архитектуры для защиты компьютерных сетей: аналитический обзор перспективных направлений исследований по результатам международного семинара MMM-ACNS-2005 // Тр. СПИИРАН. 2006. Т. 2. Вып. 3. С. 11–29.
8. Азаров А. А., Тулупьев А. Л., Тулупьева Т. В. SQL-представление реляционно-вероятностных моделей социо-инженерных атак в задачах расчета агрегированных оценок защищенности персонала информационной системы // Тр. СПИИРАН. 2012. Вып. 3(22). С. 31–44.
9. Сергеев М. Б., Соловьев Н. В., Стадник А. И. Методы повышения контрастности растровых изображений для систем цифровой обработки видеoinформации // Информационно-управляющие системы. 2007. № 1. С. 2–7.
10. Bishop C. Pattern Recognition and Machine Learning. — Springer, 2006. — 740 p.
11. Суворова А. В. и др. Вероятностные графические модели социально-значимого поведения индивида, учитывающие неполноту информации // Тр. СПИИРАН. 2012. Вып. 3(22). С. 101–112.
12. Тулупьев А. Л. Дерево смежности с идеалами конъюнктов как ациклическая алгебраическая байесовская сеть // Тр. СПИИРАН. 2006. Т. 1. Вып. 3. С. 198–227.
13. Фильченков А. А., Тулупьев А. Л. Структурный анализ систем минимальных графов смежности // Тр. СПИИРАН. Вып. 11. С. 104–127.
14. Фильченков А. А., Тулупьев А. Л., Сироткин А. В. Структурный анализ клик максимальных графов смежности алгебраических байесовских сетей // Вестн. Тверск. гос. ун-та. Сер. Прикладная математика. 2011. № 20. С. 139–151.
15. Тулупьев А. Л., Фильченков А. А., Вальтман Н. А. Алгебраические байесовские сети: задачи автоматического обучения // Информационно-измерительные и управляющие системы. 2011. Т. 9. № 11. С. 57–61.
16. Alpaydin E. Introduction to Machine Learning. 2nd ed. — Cambridge: Mass. MIT Press, 2010. — 581 p.
17. Тихомиров А. В., Шалыто А. А. Применение генетического подхода для генерации клеточных автоматов // Науч.-техн. вестн. С.-Петербург. гос. ун-та информационных технологий, механики и оптики. 2011. Вып. 2(72). С. 62–66.
18. Фильченков А. А., Тулупьев А. Л., Сироткин А. В. Минимальные графы смежности алгебраической байесовской сети: формализация основ синтеза и автоматического обучения // Нечеткие системы и мягкие вычисления: Научный журнал Российской ассоциации нечетких систем и мягких вычислений. 2011. Т. 6. № 2. С. 145–163.
19. Фильченков А. А., Тулупьев А. Л. Совпадение множеств минимальных и нередуцируемых графов смежности над первичной структурой алгебраической байесовской сети // Вестн. С.-Петербург. ун-та. Сер. 1. Математика. Механика. Астрономия. 2012. Вып. 2. С. 65–74.
20. Фильченков А. А. Алгоритмы построения элементов третичной полиструктуры алгебраической байесовской сети // Тр. СПИИРАН. 2011. Вып. 3(18). С. 237–266.