Алгебраические байесовские сети: сложность проверки непротиворечивости

А. Г. Максимов
СПИИРАН, СПбГУ
Санкт-Петербург, Россия
agm@dscs.pro

А. Л. Тулупьев
СПбГУ, СПИИРАН
Санкт-Петербург, Россия
alt@dscs.pro

А. Д. Завалишин
СПИИРАН, СПбГУ
Санкт-Петербург, Россия
adz@dscs.pro

***Аннотация.* Одной из первоочередных задач, возникающих в теории алгебраических байесовских сетей, является задача проверки и поддержания непротиворечивости фрагмента знаний. На данный момент существуют алгоритмы для решения этой задачи, однако в информатике и программировании крайне важно понимать вычислительную трудоемкость используемых методов. Данная работа посвящена оценке числа шагов этих алгоритмов.**

Ключевые слова: алгебраические байесовские сети; фрагмент знаний; локальная непротиворечивость; сложность; линейное программирование

I. Введение

Алгебраические байесовские сети (АБС) являются одним из классов вероятностных графических моделей, которые в свою очередь являются представлением баз фрагментов знаний с неопределенностью [1]. АБС представляет собой неориентированный граф, в вершинах которого находятся фрагменты знаний [2]. Каждый фрагмент знаний представляется идеалом конъюнктов с оценками вероятности истинности каждого элемента [2].

Одной из важнейших задач логико-вероятностного вывода в АБС является задача проверки локальной непротиворечивости фрагмента знаний [4]. На данный момент существуют алгоритмы для решения этой задачи, однако в информатике и программировании крайне важно понимать их вычислительную трудоемкость, поскольку экспоненциальные и близкие к ним алгоритмы неэффективны на практике. Поэтому в случаях, когда вычислительная сложность алгоритмов оказывается слишком большой, необходимо искать альтернативные пути решения. Данная работа посвящена оценке сложности этих алгоритмов. Для непротиворечивости фрагмента знаний со скалярными оценками мы оценим число необходимых арифметических операций. Для непротиворечивости в случае интервальных оценок мы оценим число вспомогательных задач и их сложность.

Работа выполнена в рамках проекта по государственному заданию СПИИРАН № 0073-2019-0003 при финансовой поддержке РФФИ, проект №18-01-00626

II. Релевантные работы

Как было сказано выше, алгебраические байесовские сети относятся к классу вероятностных графических моделей, к которым также относятся байесовские сети доверия [12] и марковские сети [13]. Одним из преимуществ АБС перед родственными байесовскими сетями доверия является возможность работать с неточными, неполными и нечисловыми оценками, формально выражая их в виде интервальных оценок вероятности истинности элементов сети [1]. Эта модель глубоко и разносторонне разобрана в монографии [1] и учебнике [2].

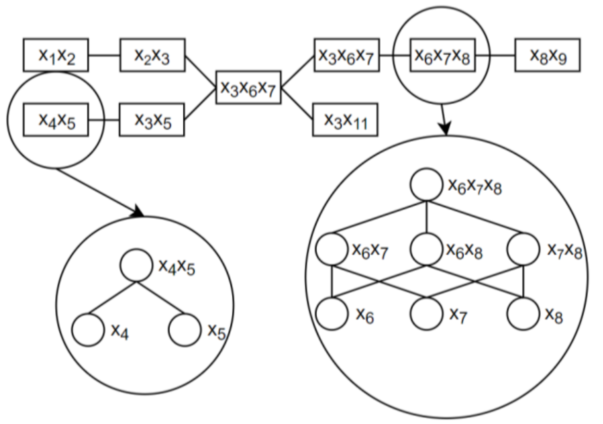

Рис. 1. Алгебраическая байесовская сеть

III. Локальная непротиворечивость: скалярные оценки

Условимся, что здесь и далее n обозначает число переменных.

Под сложением понимается применение групповой операции $\mathbb{R}$, как группы по сложению, то есть и вычитание (прибавление обратного) тоже. «Естественное» сложение будем называть сложением в слабом смысле.

В первую очередь рассмотрим задачу проверки локальной непротиворечивости фрагмента знаний со скалярными оценками, то есть фрагмента знаний, оценки вероятности истинности конъюнктов которого являются числами.

В работе [3] было доказано, что для проверки локальной непротиворечивости в случае ФЗ со скалярными оценками достаточно проверить следующее неравенство:

$$I_n P \geq 0,$$

где P – вектор оценок вероятностей истинности конъюнктов, а I_n – рекурсивно определяемая матрица: $I_1 = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, I_n = I_{n-1} \otimes I_1$, то есть n-ая тензорная степень матрицы I_1. Оценим число операций, необходимых для такого сравнения.

Заметим для начала, что матрица I_n имеет размеры 2^n на 2^n. Вектор оценок вероятностей истинности конъюнктов в свою очередь тоже имеет размер 2^n. Тогда произведение $I_n P$ может быть вычислено за 2^n операций умножения и $2^n - 1$ операций сложения на один элемент результирующего вектора. Всего таких элементов 2^n. Таким образом, произведение может быть вычислено за $2^n(2^n + 2^n - 1) = 2^{2n+1} - 2^n$. Кроме того, потребуется еще 2^n сравнений с нулем. Суммарная оценка 2^{2n+1}, то есть $2 \cdot 4^n$.

Однако эта оценка может быть улучшена, если мы воспользуемся пониманием вида матрицы I_n.

Утверждение: матрица I_n содержит ровно 3^n ненулевых элементов.

Утверждение: любая строка матрицы I_n содержит ненулевой элемент.

Утверждение: любой ненулевой элемент матрицы I_n равен единице по модулю.

Каждое из этих утверждений напрямую следуют из определения матрицы I_n. В самом деле, матрица I_n выглядит следующим образом: $\begin{bmatrix} I_{n-1} & -I_{n-1} \\ 0 & I_{n-1} \end{bmatrix}$. Тогда если в матрице I_{n-1} 3^{n-1} ненулевых элемента, то в ненулевые элементы матрицы I_n каждый из них внесет вклад трижды. Если в каждой строке I_{n-1} есть ненулевой элемент, то "правый столбец" I_n обеспечит сохранение этого свойства. Наконец, если все элементы I_{n-1} равны ± 1 и 0, то при переходе к I_n это свойство не исчезнет, потому что элементы I_{n-1} умножаются только на ± 1 и 0.

Отметим также, что умножение на ± 1, по большому счету не является умножением, а в умножении на 0 нет необходимости.

Теперь мы готовы дать следующую оценку: $3^n - 2^n$ операций сложения и 2^n сравнений с нулем. Оценка числа операций сложения вытекает из первых двух утверждений – так как для вычисления суммы k элементов необходимо всего $k - 1$ операций сложения, мы «экономим» одну операцию за каждую строку с ненулевым элементом, которых, согласно второму утверждению, в точности 2^n. Итоговая оценка – 3^n. Нетрудно заметить, что $3^n < 2 \cdot 4^n$.

Отметим, что приведенная выше оценка на самом деле не является экспоненциальной. Размер входных данных в задаче проверки непротиворечивости в ФЗ со скалярными оценками составляет 2^n. Обозначив его за новую переменную x увидим, что необходимо $x^{\log_2 3}$ операций – даже меньше x^2.

IV. Локальная непротиворечивость: интервальные оценки

Второй рассматриваемой нами задачей станет задача проверки непротиворечивости в ФЗ с интервальными оценками, то есть в ФЗ, для каждого конъюнкта которого известны нижняя и верхняя оценки вероятности истинности. В случае интервальных оценок встает также вопрос о *поддержании* непротиворечивости – уточнении оценок элементов [4].

Пример: если в ФЗ над двумя переменными вероятности истинности переменных x и y равны нулю, а вероятность истинности xy лежит на отрезке [0,1], то эту оценку можно уточнить – конъюнкция xy тоже никогда не бывает истинной.

Для проверки и поддержания локальной непротиворечивости в ФЗ с интервальными оценками необходимо составить систему ограничений R [1]. Она будет складываться из двух подсистем E и D. D содержит ограничения предметной области – вероятность истинности формулы f не превосходит верхней оценки и не опускается ниже нижней. E содержит ограничения аксиоматики теории вероятностей, например, $P(x) \geq P(xy)$. Чтобы получить ограничения E, необходимо умножить уже знакомую нам матрицу I_n на вектор переменных P [3]. Имея в распоряжении систему ограничений R, для вычисления верхней ($p^+(f)$) и нижней ($p^+(f)$) оценок вероятностей формулы f можно решить серию задач линейного программирования [5] следующего вида:

$$p^+(f) = \max_R \{p(f)\}$$

$$p^-(f) = \min_R \{p(f)\}$$

Для решения задач линейного программирования можно воспользоваться, например, симплекс-методом [6]. Именно такой подход традиционно используется в теории алгебраических байесовских сетей [3].

Поскольку методы решения задач линейного программирования могут отличаться, оценим не число операций, а количество необходимых задач, а также их сложность – число ограничений и переменных в них.

Поскольку ФЗ содержит 2^n конъюнкций, одна из которых пустая, нам придется решить $2(2^n - 1)$ задач линейного программирования – по одной на каждую конъюнкцию, кроме пустой, так как ее вероятность всегда равна единице. В каждой из них будет в точности $2^n - 1$ основная переменная — по одной на каждую конъюнкцию, кроме пустой. Наконец, число ограничений складывается из двух факторов. Ограничений предметной области будет $2(2^n - 1)$ – два ограничения (верхняя и нижняя оценка) на каждую переменную. Размер системы ограничений аксиоматики теории вероятностей совпадает с числом строк матрицы I_n. Как говорилось выше, в матрице I_n 2^n строк. Таким образом, необходимо решить $2(2^n - 1)$ задач линейного программирования с $2^n - 1$ переменными и $3 \cdot 2^n - 2$ ограничениями.

Покажем теперь, что приведенная выше оценка вновь не является экспоненциальной. Размер входных данных x в этом случае будет $2(2^n - 1)$. В терминах икса нам потребуется решить x задач линейного программирования с $x/2$ переменными и $\frac{x}{2} \cdot 3 + 1$ ограничениями – для этого нам вполне хватит полинома не слишком большой степени [7, 8].

V. Заключение

В работе рассмотрены основные методы проверки и поддержания локальной непротиворечивости во фрагментах знаний со скалярными и интервальными оценками вероятностей. Даны оценки их сложности. По результатам работы можно сделать вывод о полиномиальной разрешимости задач проверки непротиворечивости фрагмента знаний, что указывает на возможности эффективного применения алгебраических байесовских сетей в практических задачах.

В будущем авторы надеются оценить скорость сходимости симплекс-метода в задачах поддержания непротиворечивости, а также других задачах, возникающих в теории АБС. Кроме того, авторы надеются, что благодаря общему виду систем ограничений, возникающих в этих задачах, удастся найти аналитический вид их решений. В дальнейшем АБС предполагается применять, например, в исследованиях социоинженерных атак [9, 10, 11].

Список литературы

[1] Тулупьев А.Л., Николенко С.И., Сироткин А.В. Байесовские сети: логико-вероятностный подход. СПб.: Наука, 2006. 607 с.

[2] Тулупьев А.Л., Николенко С.И., Сироткин А.В Основы теории байесовских сетей. Санкт-Петербургский государственный университет. 2019. 400 с.

[3] Тулупьев А.Л. Алгебраические байесовские сети: локальный логико-вероятностный вывод: Учеб. пособие. СПб.: СПбГУ; Издательство «Анатолия», 2007. 80 с. (Сер. Элементы мягких вычислений).

[4] Тулупьев А.Л. Алгебраические байесовские сети: система операций глобального логико-вероятностного вывода // Информационно-измерительные и управляющие системы. 2010. №11. С. 65–72.

[5] Канторович Л.В. Об одном эффективном методе решения некоторых классов экстремальных проблем // Доклады Академии Наук СССР. №28. 1940. С. 211–214.

[6] Dantzig G.B. Maximization of a linear function of variables subject to linear inequalities // Activity analysis of production and allocation. 1951. Vol. 13. Pp. 339–347.

[7] Karmarkar N. A new polynomial-time algorithm for linear programming // Proceedings of the sixteenth annual ACM symposium on Theory of computing. 1984. Pp. 302–311.

[8] Khachiyan L.G. A polynomial algorithm in linear programming // Doklady Akademii Nauk. – Russian Academy of Sciences. 1979. Vol. 244. №. 5. Pp. 1093–1096.

[9] Корепанова А.А., Олисеенко В.Д., Абрамов М.В., Тулупьев А.Л. Применение методов машинного обучения в задаче идентификации аккаунтов пользователя в двух социальных сетях // Компьютерные инструменты в образовании. 2019. №3. С. 29–43. doi:10.32603/2071-2340-2019-3-29-43

[10] Suleimanov A., Abramov M., Tulupyev A. Modelling of the social engineering attacks based on social graph of employees communications analysis // 2018 IEEE Industrial Cyber-Physical Systems (ICPS). IEEE, 2018. P. 801-805.

[11] Khlobystova A., Abramov M., Tulupyev A. An approach to estimating of criticality of social engineering attacks traces // Studies in Systems, Decision and Control. 2019. vol. 199. P. 446–456.

[12] Judea Pearl. Probabilistic Reasoning in Intelligent Systems. CA: Morgan Kaufmann. 1988.

[13] Kindermann R. Markov random fields and their applications //American mathematical society. 1980.