**Алгоритмы жадного поиска эквивалентности.**

В основе многих исследований лежит двухфазный алгоритм жадного поиска эквивалентности, основанный на гипотезе, предложенной Миком [5], и изложенный Чикерингом в [2]. Основная идея GES заключается в итеративном поиске наилучшего DAG в пространстве классов эквивалентности. Алгоритм начинает работу с несвязанного графа и итеративно добавляет ребра, увеличивающие значение score-функции до тех пор, пока добавление ребер в прямой фазе GES не перестанет приводить к увеличению оценки BIC. Как только алгоритм заходит в локальный максимум, наступает обратная фаза, в которой последовательно удаляются ребра, так же по пути наибольшего роста score-функции. Алгоритм завершает работу, как только достигается максимум функции оценки, и полученный класс эквивалентности возвращается в качестве решения.

Однако алгоритм GES не эффективен для сетей с большим количеством переменных, поскольку его выполнение требует большого количества времени, особенно при многократном запуске. [8] предложили быстрый алгоритм жадного поиска эквивалентности fGES (Fast Greedy Equivalence Search), позволяющий ускорить процесс обучения больших байесовских сетей, которые могут охватывать до миллиона узлов и ребер. Авторы исследования предлагают усовершенствовать концепцию GES путём распределения вычислений между несколькими процессорами, так как каждый этап поиска может быть реализован параллельно для проверок независимостей между парами переменных. Это обусловлено тем, что ребра между переменными в представленном подходе оцениваются при предположении, что граф пуст, такие оценки независимы, поэтому могут выполняться параллельно.

В течение первой фазы алгоритма, подразумевающей прямой поиск, рёбра добавляются последовательно по пути наибольшего увеличения значения score-функции, при этом на каждом последующем шаге можно распределить оценку возможных ребер. Во второй фазе выполнения алгоритма необходимо оценить меньшее количество оценок, поскольку проверяются только те ребра, которые уже были добавлены в граф, однако потребность в распараллеливании может возникнуть в случае, если граф получился плотным.

**Алгоритмы, основанные на декомпозиции графа независимостей.**

Авторы [4] предлагают решать проблему обучения больших байесовских сетей путем разделения графа на ряд локальных структур и обучения сетей меньшей размерности с последующим их объединением. В работе представлен гибридный алгоритм структурного обучения SAR (Separation And Reunion).

На начальном этапе строится ненаправленный граф, описывающий отношения независимости между переменными. Ребро между узлами добавляется, если значение взаимной информации (MI) или условной взаимной информации (CMI) между ними достаточно велико. Чтобы ошибочно не удалить ребро, являющееся истинным, авторы вводят пороговые значения, позволяющие определять наиболее значимые связи между вершинами.

Далее полученный граф последовательно декомпозируется, для этого необходимо найти минимальный набор узлов, который делит сеть на два подграфа. Эта задача решается путем минимизации функции:

,

где 𝔼(X, Y) – набор ребер, соединяющих две части графа – X и Y.

Далее на полученных локальных структурах обучаются байесовские сети небольшой размерности и объединяются в единую сеть.

Так как глобальный поиск разделителей является экспоненциально сложной задачей при большом количестве переменных в байесовской сети, авторы статьи [2] предлагают искать разделители не глобально во всем графе, а локально в подграфах, что позволяет сузить пространство поиска и тем самым ускорить процесс обучения.

В работе представлен рекурсивный метод локального поиска разделителей, основная идея которого заключается в последовательном разбиении морального графа на множество простых подграфов. Авторы предлагают искать разделители в подграфах, при этом в качестве целевой переменной рекомендуется выбирать узел с наименьшей степенью, чтобы разбить граф на ряд наиболее простых структур.

Изображение выглядит как диаграмма, линия

Автоматически созданное описание

[2]

На рис.1 представлена байесовская сеть (a), её моральный граф (b), а также два способа поиска разделителей в нём: рекурсивный поиск через случайный целевой узел (c) и через узел с минимальной степенью (d). Предполагается, что случайным образом в качестве целевой переменной мольного графа Gm был выбран узел B, разделителем является вершина D, которая делит граф на два подграфа: {A, B, C, D} и {D, E}. Так как метод локального поиска подразумевает поиск разделителей только в подграфе, не содержащем целевую переменную, подграф {A, B, C, D} исключается из рассмотрения, при этом граф {D, E}, является простым и дальнейшая его декомпозиция невозможна.

Рассмотрим второй вариант декомпозиции: согласно представленной концепции, в качестве целевой переменной выступает вершина с наименьшей степенью, в данном случае – E, узел D разбивает граф на {A, B, C, D} и {D, E}. Поиск продолжается в подграфе {A, B, C, D}, наименьшей степенью обладают узлы A и C, разделителем в данном случае выступает ребро между вершинами B и D, а подграф декомпозируется на {A, B, D} и {B, C, D}.

Приведенный пример иллюстрирует наибольшую эффективность поиска разделителей при выборе узла с минимальной степенью в качестве целевой переменной. Подграфы получаются наиболее простыми, а также схожими по размеру.

Далее в работе излагается алгоритм ERDA, реализуемый в два этапа.

1. Построение и декомпозиция неориентированного графа. Авторы используют в работе алгоритм HITON-MB [1] для построения морального графа, однако допускают использование иных алгоритмов обнаружения марковского одеяла. Затем осуществляется декомпозиция в соответствии с методологией, представленной выше.
2. Обучение байесовских сетей меньшего размера в подграфах и объединение в единую сеть.

Другим направлением исследований в области структурного обучения больших байесовских сетей является использование k-деревьев. В основе данной концепции лежит идея о том, что k-дерево является максимальным графом с шириной дерева k, поэтому каждый граф (в том числе моральный граф байесовской сети) c шириной, не превосходящей k, является подграфом k-дерева. В работах [6] [7] авторы описывают алгоритмы S2 и S2+, в основе которых лежит описанная концепция.

В исследовании авторы используют информационную оценку IS, чтобы определить, насколько хорошо k-дерево описывает распределение данных.

,

где *Smi(Tk)* отражает объём потери информации от представления данных в виде k-дерева, а *Sl(Tk)* представляет собой оценку наилучшего подграфа k-дерева.

где – взаимная информация узлов i и j, *m(G)* – моральный граф, а – функция локальной оценки для родительского набора . [6]

Значение IS используется в алгоритме S2+ для отбора k+1 переменных с наибольшими значениями этой оценки. Далее осуществляется поиск k-дерева, максимизирующего IS для данной клики.

Далее для поиска направленного ациклического графа, моральный граф которого является подграфом полученного k-дерева, авторы предлагают приближенный алгоритм, основанный на локальном топологическом упорядочении. Для этого k-дерево представляется в виде дерева, содержащего все его максимальные клики, c корнем в R. Сначала устанавливается порядок для R, затем рассматривается клика C (родительской клике которой уже присвоен порядок) и устанавливается порядок для узлов в С относительно порядка её родительской клики и порядок задаётся порядок для самой клики C. Алгоритм продолжает работу, пока все максимальные клики k-дерева не будут упорядочены.

Далее планируется рассмотреть алгоритмы k-A и k-G, также использующие k-деревья. В работе представлено сравнение результатов работы представленных алгоритмов с рассмотренными ранее S2 и S2+.

(Scanagatta M. et al. Learning Bounded Treewidth Bayesian Networks with Thousands of Variables //arXiv preprint arXiv:1605.03392. – 2016.)

Также, текст будет дополнен сравнением представленных алгоритмов, основанных на декомпозиции (ERDA и SAR).

Список источников

[1] Aliferis C. F. et al. Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: algorithms and empirical evaluation //Journal of Machine Learning Research. – 2010. – Т. 11. – №. 1.

[2] Chickering D. M. Optimal structure identification with greedy search //Journal of machine learning research. – 2002. – Т. 3. – №. Nov. – С. 507-554.

[3] Jia X., Li H., Guo H. A recursive local search method of separators for Bayesian network decomposition structure learning algorithm //Soft Computing. – 2023. – Т. 27. – №. 7. – С. 3673-3687.

[4] Liu H. et al. A new hybrid method for learning bayesian networks: Separation and reunion //Knowledge-Based Systems. – 2017. – Т. 121. – С. 185-197.

[5] Meek C. Graphical Models: Selecting causal and statistical models : дис. – Carnegie Mellon University, 1997.

[6] Nie S., De Campos C. P., Ji Q. Learning bounded tree-width Bayesian networks via sampling //Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 13th European Conference, ECSQARU 2015, Compiègne, France, July 15-17, 2015. Proceedings 13. – Springer International Publishing, 2015. – С. 387-396.

[7] Nie S. et al. Advances in learning Bayesian networks of bounded treewidth //Advances in neural information processing systems. – 2014. – Т. 27.

[8] Ramsey J. et al. A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images //International journal of data science and analytics. – 2017. – Т. 3. – С. 121-129.