**Министерство науки и высшего образования Российской Федерации**

**федеральное государственное автономное образовательное учреждение высшего образования**

**«Национальный исследовательский университет ИТМО»**

**(Университет ИТМО)**

**Факультет цифровых трансформаций**

**Образовательная программа Искусственный интеллект в промышленности**

**Направление подготовки (специальность) 09.04.02 Информационные системы и технологии**

О Т Ч Е Т

о практике (Научно-исследовательская работа)

Тема задания: Исследование и разработка алгоритмов мета-обучения графовых структур для ускорения обучения больших байесовских сетей

Обучающийся Селиванова Виктория Валерьевна, J4151

Согласовано:

Руководитель практики от университета:Деева Ирина Юрьевна, доцент, факультет цифровых трансформаций

Практика пройдена с оценкой **\_\_\_\_**

Дата **\_\_\_\_**

Санкт-Петербург

20**23**

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ. 3](#_Toc156213252)

[1. БАЙЕСОВСКИЕ СЕТИ. 5](#_Toc156213253)

[2. СТРУКТУРНОЕ ОБУЧЕНИЕ БОЛЬШИХ БАЙЕСОВСКИХ СЕТЕЙ. 6](#_Toc156213254)

[2.1. Алгоритм распараллеливания вычислений. 6](#_Toc156213255)

[2.2. Алгоритмы, основанные на декомпозиции графа независимостей. 7](#_Toc156213256)

[2.3. Алгоритмы, основанные на сокращении пространства поиска. 11](#_Toc156213257)

[2.МЕТА-ОБУЧЕНИЕ. 15](#_Toc156213258)

[2.1 Примеры применения мета-обучения для решения задач машинного обучения. 15](#_Toc156213259)

[2.2 Мета-обучение графов. 16](#_Toc156213260)

[3. ХАРАКТЕРИСТИКИ ГРАФОВ И НАБОРОВ ДАННЫХ. 18](#_Toc156213261)

[ЗАКЛЮЧЕНИЕ 21](#_Toc156213262)

[Список источников 22](#_Toc156213263)

# **ВВЕДЕНИЕ.**

Байесовская сеть – это вероятностная графическая модель, позволяющая представлять многомерные данные в виде факторизованного распределения. Байесовские сети используются в различных областях науки, включая медицину, экономику, инженерию, биоинформатику и т.д. Широкое применение они получили и в машинном обучении, поскольку их способность моделировать вероятностные зависимости и учитывать неопределенность делает их полезными инструментами для анализа данных.

Однако нахождение структуры байесовской сети является NP-трудной задачей, поскольку при увеличении количества переменных, пространство поиска возможных структур байесовской сети растёт суперэкспоненциально. Это накладывает ограничения на процесс поиска оптимальной структуры для больших байесовских сетей, которые масштабируются до сотен, тысяч или миллионов узлов. Применение стандартных алгоритмов структурного обучения в таких случаях, во-первых, не всегда возможно, во-вторых, очень затратно по времени и вычислительным ресурсам. В связи с этим актуальной становится задача исследования алгоритмов поиска оптимальной структуры таких графических моделей.

Целью исследование является повышение эффективности обучения больших байесовских сетей с точки зрения времени при сопоставимом качестве обучения.

Для достижения поставленной цели были сформулированы следующие задачи:

* Формулировка проблемы обучения больших байесовских сетей;
* Изложение теоретических основ структурного обучения байесовских сетей;
* Обзор предлагаемых методов обучения больших байесовских сетей;
* Сравнительный анализ предлагаемых методов;
* Обзор литературы, посвященной концепции мета-обучения;
* Обзор существующих алгоритмов мета-обучения байесовских сетей и задач, которые они решают.
* Изложение идеи применения мета-обучения для ускорения обучения больших байесовских сетей.

# **1. БАЙЕСОВСКИЕ СЕТИ.**

Байесовская сеть B представляет собой кортеж, состоящий из направленного ациклического графа G (далее DAG, Directed Acyclic Graph) и набора параметров θ:

При этом DAG G состоит из набора узлов X, каждый из которых соответствует одной из рассматриваемых n переменных X = {X1, X2, …, Xn}, и соединяющих их направленных рёбер E.

Направленное ребро из вершины Xi в вершину Xj свидетельствует о зависимости между данными переменными, при этом Xi идентифицируется как родительский элемент Xj.

Важной особенностью байесовских сетей является возможность факторизации многомерного распределения. Это свойство позволяет представлять многомерное распределение P(X1, X2, …, Xn) с учётом отношений условной зависимости между переменными. Пусть Pa(Xi) представляет набор родителей элемента Xi. Тогда отношения зависимостей между переменными можно выразить с помощью условного распределения вероятностей:

Для построение байесовской сети необходимо определить её структуру, это можно сделать как с помощью экспертных знаний об изучаемой области, так и на основе набора данных. Алгоритмы изучения структуры БС на основе данных можно разделить на три основные группы. Первую группу представляют алгоритмы, основанные на ограничениях (constraint based), основной идеей которых является использование статистических тестов для определения отношений условной независимости между парами переменных. Другим классом алгоритмов структурного обучения являются алгоритмы, основанные на оценке (score based), целью которых является поиск оптимальной структуры графа, максимизирующей значение некоторой score-функции. Третьей категорией являются гибридные алгоритмы, сочетающие как достоинства, так и недостатки constraint based и score based алгоритмов. [6]

Однако в контексте обучения больших байесовских сетей наиболее распространённые алгоритмы структурного обучения оказываются неприменимы по причине сверхэкспоненциально возрастающего пространства поиска возможных структур DAG, которые необходимо оценить. В связи с этим возникает необходимость в обзоре актуальных методов обучения больших байесовских сетей.

# **2. СТРУКТУРНОЕ ОБУЧЕНИЕ БОЛЬШИХ БАЙЕСОВСКИХ СЕТЕЙ.**

## **2.1. Алгоритм распараллеливания вычислений.**

В основе многих исследований лежит двухфазный алгоритм жадного поиска эквивалентности, основанный на гипотезе, предложенной Миком [13], и изложенный Чикерингом в [3]. 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. [16] 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.

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

Авторы [11] 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 [8] 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

Рисунок 1 Иллюстрация локального метода поиска разделителей в графе. [8]

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 [14] [15] 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 .

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.

## **2.3. Алгоритмы, основанные на сокращении пространства поиска.**

Ещё одним направлением развития методов структурного обучения байесовских сетей является разработка алгоритмов, позволяющих сузить пространство поиска возможных DAG. Так, например, в работе [20] представлен пакет BiDAG на языке R, реализующий методы Монте-Карло цепи Маркова (MCMC). Цель представленного подхода состоит в построении такой цепи Маркова, стационарное распределение которой равно апостериорному распределению P(G|D). На первом этапе работы посредством реализации алгоритма PC формируется сокращенное пространство поиска, представляющее собой неориентированный скелет графа. Далее осуществляется оптимизация структуры с помощью алгоритма OBS (ordering based search), основанного на упорядочении. Согласно данному алгоритму n узлов i1, i2, …, in образуют некоторый линейный порядок i1 i2 … in и алгоритм OBS ищет оптимальную структуру не во всех пространстве возможных DAG, а в пространстве порядков марковской цепи гораздо меньшей размерности.

В работе [18] представлен метод обучения больших байесовских сетей, основанный на исследовании пространства возможных родителей и структурной оптимизации на основе упорядочения. Для определения множества возможных родителей узла авторы предлагают использовать аппроксимированную оценку BIC:

где – непересекающиеся множества возможных родителей.

Оценка BIC\* позволяет быстро оценивать большие наборы узлов-кандидатов и оценивать, можно ли исключить часть вершин из дальнейшего рассмотрения, тем самым сократив пространство поиска. После выбора наиболее возможных родителей осуществляется оптимизация структуры посредством метода, основанного на упорядочении, что так же позволяет искать оптимальную структуру в пространстве меньшей размерности.

В таблице 1 представлены результаты обзора предлагаемых методов структурного обучения больших байесовских сетей.

Таблица 1. Предлагаемые методы структурного обучения больших байесовских сетей.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Название алгоритма структурного обучения больших байесовских сетей | Принцип бучения байесовской сети | Масштабируемость | Типы данных | Наличие открытой реализации |
| GES | Итеративный поиск наилуч­шего DAG в пространстве классов эквивалентности | Среднее количество узлов. | Дискретные, непрерывные | PyPI ges |
| fGES | Распределение вычислений между процессорами. | До миллиона вершин и рёбер. | Дискретные, непрерывные | Py-causal |
| SAR | Разделение графа на подгра­фы, обучение подграфов меньшей размерности с по­следующим объединением. | Более 100 узлов. | Дискретные, непрерывные | - |
| ERDA | Разбиение графа на подгра­фы на основе локального по­иска разделителей. | Более 100 узлов. | Дискретные, непрерывные | - |
| S2, S2+ | Использование k-деревьев. | Более 100 узлов. | Дискретные, непрерывные | - |
| BiDAG | Использования алгоритма PC для формирования сокра­щен­ного пространства поис­ка и осуществление струк­турной оптимизации в про­странстве порядков (мень­шей размерно­сти чем про­странство возмож­ных DAG). | Более 100 узлов. | Дискретные, непрерывные | Пакет BiDAG на языке R. |
| ASOBS | Исследование пространства возможных родителей для отсечения части простран­ства поиска и оптимизация в пространстве порядков. | От 100 до 1 500 узлов. | Дискретные, непрерывные | - |

Рассмотренные алгоритмы решают проблему обучения больших байесовских сетей по-разному: путём распределения вычислений оценок условной независимости, разделением графа на ряд локальных структур для обучения подграфов меньшей размерности или путём сокращения пространства поиска возможных структур DAG. Декомпозиционные алгоритмы сталкиваются с такими проблемами как экспоненциальная сложность поиска разделителей при большом количестве узлов в графе или чрезмерно разреженная структура графа при локальном поиске разделителей. Жадный поиск по классам эквивалентности является NP-трудной задачей. Последняя группа методов представляет особый интерес для дальнейшего изучения, поскольку они направлены на решение проблемы возрастания пространства поиска. К тому же, стоит отметить, что из рассмотренных алгоритмов лишь GES [3], fGES [12] и BiDAG [15] имеют открытую реализацию. Для дальнейших исследований методов, позволяющих сократить пространство поиска возможных DAG может быть целесообразным рассмотрение возможности применения мета-обучения.

# **2.МЕТА-ОБУЧЕНИЕ.**

Мета-обучение – это наука о методах, использующих метазнания для эффективного обучения моделей путём адаптации процессов машинного обучения. [2] Цель мета-обучения – накопить опыт во время решения тех или иных задач машинного обучения, а затем применить его для гораздо более быстрого построения новых моделей.

Мета-обучение позволяет решать широкий спектр задач, например, таких как выбор алгоритма на основе характеристик производительности, оптимизация гиперпараметров, обучение модели в условиях дефицита данных, обобщение знаний, полученных из других задач, регуляризация и предотвращение переобучения и т.д.

В контексте обучения больших байесовских сетей мета-обучение может позволить на основе каких-либо характеристик набора данных делать выводы о мета-характеристиках графа, позволяющих сократить пространство поиска и тем самым ускорить процесс построения байесовской сети. Но для начала рассмотрим, как алгоритмы мета-обучения применяются для решения задач машинного обучения.

## **2.1 Примеры применения мета-обучения для решения задач машинного обучения.**

В работе [4] представлен широко применяемый алгоритм независимого от модели мета-обучения MAML (Model Agnostic Meta-Learning), основная идея которого заключается в обучении параметров модели таким образом, чтобы она могла быстро адаптироваться к новой задаче. Цель предложенного подхода состоит в том, чтобы изучить функцию, применимую к ряду задач, путём минимизации общих потерь и найти такой набор параметров инициализации θ, от которого можно быстро оптимизировать новую модель за несколько шагов градиентного спуска.

В статье [10] мета-обучение применяется для выбора лучшего алгоритма классификации на основе характеристик набора данных. В ходе исследования из 34 наборов данных авторы отобрали 32 показателя, описывающих различные характеристики датасетов, после чего применили на каждом из наборов 13 алгоритмов классификации, для сопоставления точности алгоритмов на тех или иных входных данных. В заключительном модуле алгоритма осуществляется рекомендация оптимального алгоритма для нового набора данных на основе сходств его характеристик с характеристиками эталонного набора.

## **2.2 Мета-обучение графов.**

Авторы статьи [7] предлагают вместо того, чтобы обучать DAG на полном графе, представлять узлы локальными подграфами, содержащими вершины, которые их окружают и использовать их для обучения GNN. В работе теоретически обосновано отсутствие потери важной информации при локальном применении GNN. Авторы представляют алгоритм G-Meta, который на начальном этапе создаёт подграфы, после чего применяет нейронное кодирование полученных подграфов и затем использует подход мета-обучения MAML (Model Agnostic Meta-Learning), основанный на оптимизации, для передачи метазнаний о графах. Преимуществами G-Meta является масштабируемость на графы с более чем 1 000 узлов за счет небольших размеров подграфов и быстрая скорость обучения.

В работе [17] представлен алгоритм обучения графов с мета-подкреплением MCD (Meta-Causal Discovery), который обучается на наборе DAG с уже известной структурой. Авторы предлагают проводить вмешательства, такие как добавление, удаление или обращение ребра, а затем проводить оценку соответствия полученного графа истинной структуре. Для этого используется метрика, основанная на структурном расстоянии Хэмминга (dSHD, d-directed), которая подсчитывает, сколько рёбер необходимо удалить или наоборот добавить, чтобы получить истинную структуру DAG.

В статье [12] сформулирована проблема плохой масштабируемости методов глобальных подходов к структурному обучению байесовских сетей и неточности локальных методов, обоснованной возможностью попадания в локальный оптимум. В связи с этим авторы предлагают 4 эвристических фактора для расширения пространства поиска, а затем представляют мета-эвристический подход к структурному обучению DAG. Рёбра в данном подходе рассматриваются как массивы из from и to вершин (пример представлен на рисунке 2).

Изображение выглядит как текст, диаграмма, снимок экрана, линия

Автоматически созданное описание

Рисунок 2. Пример DAG и представление его рёбер в виде массивов. [12]

В работе представлены такие факторы расширения пространства как преобразование узла не являющегося терминальным в терминальный (не имеющий потомков) (S1), преобразование вершины, не являющейся корневым узлом в корневой узел (S2), обмен местами двух вершин в массиве to (S3) и обмен местами всех дуг для двух вершин (S4). Используя эти факторы, авторы предлагают мета-эвристический алгоритм поиска оптимальной структуры DAG, который позволяет выйти из ловушки локального оптимума и при этом не теряет возможности масштабирования.

Авторы [21] представляют алгоритм мета-обучения Meta-GNN, решающий проблему классификации узлов в графовых нейронных сетях на небольшом наборе обучающих примеров путём обучения множеству аналогичных задач с использованием MAML. Основная идея заключается в том, чтобы из обучающего набора данных создать множество схожих задач и обучить на них алгоритм, чтобы в дальнейшем применять его к аналогичным наборам данных. При изучении каждой задачи вычисляется потеря перекрёстной энтропии, затем выполняется обновление параметров модели посредством градиентного спуска:

где – обновлённые параметры модели, – потеря перекрёстной энтропии, – скорость обучения задачи. Цель оптимизации заключается в минимизации потерь, при этом оптимизация осуществляется по параметрам модели с использованием обновлённых параметров, для того чтобы получить хорошие параметры для всех аналогичных задач, а не конкретно для рассматриваемой в текущий момент.

Рассмотренные методы иллюстрируют возможность мета-обучения решать различные задачи как в целом для машинного обучения, так и для графов в частности, что мотивирует к дальнейшему исследованию возможности применения данной концепции и для решения иных задач. Хотелось бы определить, позволит ли мета-обучение решить проблему масштабируемости алгоритмов структурного обучения, путём предсказания каких-либо мате-характеристик графов, которые могли бы позволить сократить пространство поиска DAG. В связи с этим целесообразно рассмотреть, какими характеристиками обладают графы и наборы данных.

# **3. ХАРАКТЕРИСТИКИ ГРАФОВ И НАБОРОВ ДАННЫХ.**

К базовым числовым показателям, характеризующим структуру графа, в первую очередь можно отнести количество вершин и рёбер, число соседей у узлов в графе, а также расстояние между вершинами. Также, широкое применение в анализе структуры графов получили и характеристики, являющиеся производными от перечисленных, к ним относят показатели степеней вершин, диаметр графа, плотность и коэффициент кластеризации.

Степень вершины отражает количество рёбер, которые ей инцидентны, иначе говоря, эта характеристика описывает количество соседей, с которыми соединен данный узел. На основе данного критерия может быть целесообразным определение максимальной и минимальной степени *вершины*, а также, в комбинации с количеством узлов – *среднего количества соседей у вершин* в графе. Перечисленные показатели важны тем, что они позволяют делать выводы о структуре графа с точки зрения связности вершин, а также оценивать, например, максимальное количество родителей у вершин.

Немаловажной метрикой является расстояние между узлами, которое представляет собой длину кратчайшего пути от одного узла до другого. Данная метрика лежит в основе такой характеристики графа как диаметр, который отражает максимальное расстояние между двумя вершинами и позволяет делать предположения о наличии подграфов или изолированных узлов.

Важной характеристикой, также, является плотность графа, определяемая как отношение количества рёбер в рассматриваемом графе к максимально возможному числу рёбер в полном графе с таким же количеством узлов. Плотность, также, является характеристикой связности и позволяет делать выводы о том, является ли граф разреженным (при значении показателя близкому к 0) или более плотным (при значении близкому к 1). Также, выделяют такой показатель как коэффициент кластеризации, позволяющий делать выводы о склонности вершин образовывать кластеры с соседними узлами. Он определяется как отношение количества ребер, соединяющих рассматриваемую вершину с соседними узлами, к максимально возможному количеству ребер между соседними вершинами.

Немаловажной задачей является и выявление характеристик наборов данных, которые могут позволить строить гипотезы о возможной структуре графа. Так, в книге [2] авторы представили различные мета-признаки в зависимости от задачи (классификация, регрессия, анализ временных рядов, кластеризация), а также изложили принципы генерации новых признаков. Критерии разделены на простые, статистические и теоретико-информационные. К простым характеристикам авторы отнесли количество записей, количество атрибутов, долю дискретных и непрерывных факторов, количество или долю выбросов по каждому признаку, для задачи регрессии был выделен такой показатель, как плотность данных, определяемый как отношение количества записей к количеству признаков. Среди статистических характеристик были выделены корреляция между парой признаков и ковариация, описывающие степень взаимосвязи, коэффициент эксцесса, асимметрию, среднее значение, а к теоретико-информационным критериям были отнесены энтропия и взаимная информация.

# **ЗАКЛЮЧЕНИЕ**

В ходе аналитического обзора были рассмотрены различные подходы к структурному обучению байесовских сетей большой размерности, Одним из направлений исследований является разработка алгоритмов поиска оптимальной структуры путём сокращения пространства поиска возможных DAG. Также, в работе изложена концепция мета-обучения, представлены задачи, которые она позволяет решить, а также представлены существующие на текущий момент алгоритмы мета-обучения графов. Для дальнейшего исследования предлагается разработка алгоритма по применению концепции мета-обучения графов для определения характеристик байесовских сетей, которые позволят сократить пространство поиска оптимальной структуры графа и тем самым ускорить процесс обучения.

# **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] Brazdil P. et al. Metalearning: applications to automated machine learning and data mining. – Springer Nature, 2022. – С. 346.

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

[4] Finn C., Levine S. Meta-learning and universality: Deep representations and gradient descent can approximate any learning algorithm //arXiv preprint arXiv:1710.11622. – 2017.

[5] He C., Di R., Tan X. Bayesian Network Structure Learning Using Improved A\* with Constraints from Potential Optimal Parent Sets //Mathematics. – 2023. – Т. 11. – №. 15. – С. 3344.

[6] He C., Liu W., Ren J. Bayesian Network Structure Learning: A Review //2022 6th Asian Conference on Artificial Intelligence Technology (ACAIT). – IEEE, 2022. – С. 1-7.

[7] Huang K., Zitnik M. Graph meta learning via local subgraphs //Advances in neural information processing systems. – 2020. – Т. 33. – С. 5862-5874.

[8] 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.

[9] Koller D., Friedman N. Probabilistic graphical models: principles and techniques. – MIT press, 2009.

[10] Li L. et al. Meta-learning based industrial intelligence of feature nearest algorithm selection framework for classification problems //Journal of Manufacturing Systems. – 2022. – Т. 62. – С. 767-776.

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

[12] Liu X. et al. A metaheuristic causal discovery method in directed acyclic graphs space //Knowledge-Based Systems. – 2023. – Т. 276. – С. 110749.

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

[14] 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.

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

[16] 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

[17] Sauter A. W. M., Acar E., François-Lavet V. A meta-reinforcement learning algorithm for causal discovery //Conference on Causal Learning and Reasoning. – PMLR, 2023. – С. 602-619.

[18] Scanagatta M. et al. Learning Bayesian networks with thousands of variables //Advances in neural information processing systems. – 2015. – Т. 28.

[19] Scanagatta M., Salmerón A., Stella F. A survey on Bayesian network structure learning from data //Progress in Artificial Intelligence. – 2019. – Т. 8. – С. 425-439.

[20] Suter P. et al. Bayesian structure learning and sampling of Bayesian networks with the R package BiDAG //arXiv preprint arXiv:2105.00488. – 2021.

[21] Zhou F. et al. Meta-gnn: On few-shot node classification in graph meta-learning //Proceedings of the 28th ACM International Conference on Information and Knowledge Management. – 2019. – С. 2357-2360.