Байесовские сетевые модели для неполных и динамических данных

Марко Скутари1

1 | ВВЕДЕНИЕ

Байесовские сети (BNS) представляют собой вероятностную графическую модель, которая идеально подходит для изучения сложных явлений посредством взаимодействия их компонентов. Представляя последние в виде узлов графика и соединяя их дугами, которые кодируют то, как эти компоненты взаимодействуют друг с другом, они обеспечивают высокоуровневое качественное представление об исследуемом явлении, построенное на прочных теоретических основах.

BNS были первоначально разработаны в 1980-х годах в контексте экспертных систем, а основополагающая работа была обобщена в Pearl (1988) и Neapolitan (1989). Более поздние монографии Кастильо и др. (1997), Неаполитана (2003), Коуэлла и др. (2007), Корба и Николсона (2010) широко освещают свойства BNs и их использование; Коллер и Фридман (2009) являются наиболее полным справочником на сегодняшний день. В этом контексте “экспертные системы” были задуманы как комбинация формального представления знаний о конкретной предметной области, собранных экспертами в предметной области, и “механизма вывода”, который мог бы использовать это представление для ответа на произвольные запросы. BNS могут представлять сложные явления модульным способом благодаря кодируемым ими предположениям об условной независимости, что позволяет им выходить за рамки тривиальных задач и использоваться для разработки эффективных алгоритмов вывода.

Как отмечается в работе Cowell et al. (2007), к концу 1990-х годов ограничения построения экспертных систем становились все более очевидными и подталкивали исследования к изучению графической структуры и параметров BNS на основе данных. Для многих явлений, которые были и с тех пор представляют интерес для научного сообщества, экспертные системы слишком сложны, чтобы разработчик моделирования мог получить от экспертов требуемые знания в предметной области, даже предполагая, что экспертные знания доступны в первую очередь (или с достаточным консенсусом, которому можно доверять). Однако в то же время доступность все более крупных наборов данных сделала более привлекательным их использование в качестве альтернативного источника информации для построения BNS.

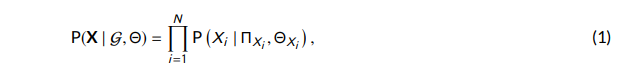
Учитывая ограниченную вычислительную мощность, доступную в то время, этот переход от экспертного выявления к обучению на основе данных стал возможным благодаря принятию ряда упрощающих допущений относительно характера данных (Heckerman et al., 1995). С тех пор в подавляющем большинстве литературы, в частности, были сделаны два предположения. Во-первых, что данные являются полными; то есть, что все переменные полностью соблюдаются для всех выборок. И, во-вторых, эти выборки независимы друг от друга; что, в частности, исключает динамические данные, такие как временные ряды, в которых выборки связаны друг с другом.

Однако проблемы, стоящие на переднем крае исследований и практических приложений, только усложняются, что делает эти предположения слишком сильными во многих ситуациях. В этой статье мы рассмотрим, как BNS можно использовать для моделирования динамических данных и данных с неполными наблюдениями, приведя несколько примеров, в которых они успешно использовались для разработки передовых приложений. Сравнительно простой случай полных статических данных, безусловно, привлек наибольшее внимание в литературе по историческим причинам и из-за его вероятностной и вычислительной простоты; но BN ни в коем случае не ограничиваются моделированием только таких данных.

Содержание этого обзора организовано следующим образом. В разделе 2 мы знакомим с BNS в общих чертах, включая лежащие в их основе допущения о распределении и доступные подходы к обучению. В разделе 3 мы обсудим BNS, ориентированные на динамические (то есть временные) данные, начиная с их определения и основных свойств (раздел 3.1), а затем продемонстрируем, что это общий подход, включающий несколько других моделей, представленных в литературе (раздел 3.2). В разделе 4 мы обсуждаем неполные данные и то, как их можно правильно использовать для изучения параметров (раздел 4.1) и структуры. Наконец, в разделе 5 мы приводим обзор доступных программных реализаций (раздел 5.1) и реальных приложений (раздел 5.2) BNs, которые мы обсуждали.

2 | СПРАВОЧНАЯ ИНФОРМАЦИЯ И ОБОЗНАЧЕНИЯ

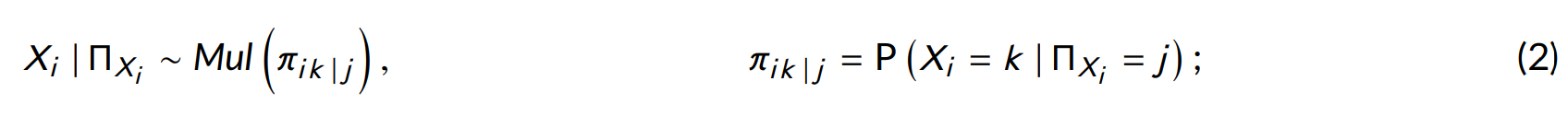
BNS - это класс графических моделей, в которых узлы ориентированного ациклического графа (DAG) G представляют собой набор X ={X1, . . . , XN } случайных величин, описывающих некоторые представляющие интерес величины. Дуги, соединяющие эти узлы, выражают отношения прямой зависимости, при этом графическое разделение в G подразумевает условную независимость по вероятности. В результате G индуцирует факторизацию



в котором совместное распределение вероятностей X (с параметрами Θ) разлагается на одно локальное распределение для каждого Xi (с параметрами ), зависящее от его родителей ΠX . Предполагая, что G равно sparse1, BNs обеспечивает компактное представление как низко-, так и высокомерных распределений вероятностей.

1 Не существует общепринятого порога количества дуг для того, чтобы DAG назывался “разреженным”; обычно принимается значение O (cN ) дуг с c от 1 до 5.

BN также очень гибки с точки зрения предположений о распределении; но хотя в принципе мы могли бы выбрать любое распределение вероятностей для X, литература в основном сосредоточена на трех случаях по аналитическим и вычислительным причинам. Дискретные BNS (Хекерман и др., 1995) предполагают, что и X, и Xi являются полиномиальными случайными величинами. Локальные распределения принимают вид

их параметры представляют собой условные вероятности Xi с учетом каждой конфигурации значений его родительских элементов, обычно представленные в виде таблицы условных вероятностей для каждого Xi . Гауссовская BNS (GBNs; Гейгер и Хекерман, 1994) моделирует X с многомерной нормальной случайной величиной и предполагает, что Xi являются одномерными нормалями, связанными линейными зависимостями. Параметры локальных распределений могут быть эквивалентно записаны (Weatherburn, 1961) как частичные корреляции между Xi и каждым родительским Xj с учетом других родительских параметров; или как коэффициенты βXi модели линейной регрессии

где 

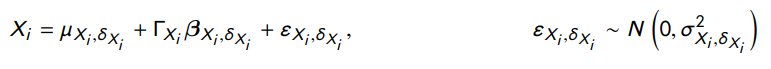
Наконец, условные линейные гауссовы BNS (CLGBNs; Лауритцен и Вермут, 1989) объединяют дискретные и непрерывные случайные величины в смешанной модели:

• дискретным Xi разрешается иметь только дискретных родителей (обозначаемых ∆Xi ), и предполагается, что они следуют полиномиальному распределению, как в (2);

• непрерывным Xi разрешается иметь как дискретных, так и непрерывных родителей (обозначаемых формула), а их локальные распределения являются

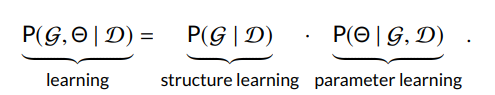


который может быть записан как смесь линейных регрессий

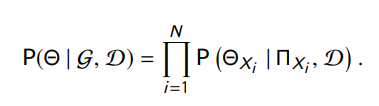


против непрерывных родительских элементов с одним компонентом для каждой конфигурации δXi дискретных родительских элементов . Если Xi не имеет дискретных родителей, смесь возвращается к одиночной линейной регрессии, подобной приведенной в (3).

Задача изучения BN из набора данных D, содержащего n наблюдений, выполняется в два этапа:



Изучение структуры заключается в нахождении базы данных G, которая кодирует структуру зависимости данных, таким образом максимизируя P (G | D) или какую-либо альтернативную меру соответствия; изучение параметров заключается в оценке параметров Θ с учетом G, полученного в результате изучения структуры. Оба этапа могут интегрировать данные с экспертными знаниями за счет использования подходящих предварительных распределений для G и Θ (см., например, Castelo and Siebes, 2000; Mukherjee and Speed, 2008; Druzdzel and van der Gaag, 1995). Если мы предположим, что параметры в разных локальных распределениях независимы и что данные не содержат пропущенных значений (Heckerman et al., 1995), мы можем выполнять изучение параметров независимо для каждого Xi, потому что следующее (1)



Кроме того, предполагая, что G является разреженным, каждое локальное распределение Xi | ΠXi будет включать только несколько переменных и, следовательно, будет иметь маломерное пространство параметров, что делает изучение параметров эффективным с вычислительной точки зрения.

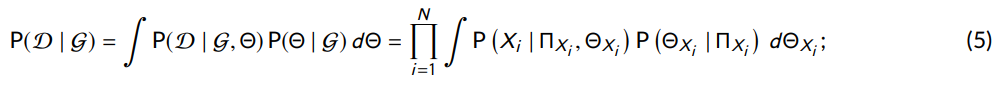
С другой стороны, хорошо известно, что изучение структуры является как NP-сложным (Чикеринг и Хекерман, 1994), так и NP-полным (Чикеринг, 1996), даже при нереально благоприятных условиях, таких как доступность оракула независимости и вывода (Chickering et al., 2004)2.

2 Интересно, что некоторые улучшения в изучении структуры BN не являются NP-сложными; см., Например, Dojer (2006) об изучении структуры причинно-следственных сетей.

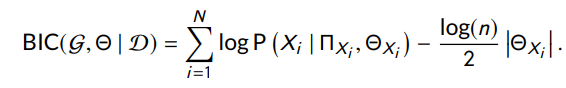
Это несмотря на то, что если мы возьмем



следуя (1), мы можем разложить предельную вероятность P (D | G) на одну составляющую для каждого локального распределения



и это несмотря на то, что каждый компонент может быть записан в замкнутом виде для дискретных BNS (Хекерман и др., 1995), GBNs (Гейгер и Хекерман, 1994) и CLGBNs (Беттчер, 2001). То же самое верно, если мы заменим P (D | G) частотными оценками соответствия, такими как BIC (Schwarz, 1978), который обычно используется при изучении структуры из-за его простого выражения:



По сравнению с предельными вероятностями, BIC также имеет то преимущество, что он не зависит ни от какого гиперпараметра, при этом сходясь к log P (D | G) как n → ∞. Эти функции оценки обладают двумя важными свойствами:

• они допускают локальные вычисления, поскольку, следуя (1), они разлагаются на один компонент для каждого локального распределения;

• они принимают одинаковое значение для всех DAG, кодирующих одинаковое распределение вероятностей (эквивалентность баллов), которые затем могут быть сгруппированы в классы эквивалентности (Чикеринг и Хекерман, 1997).3

3 Все базы данных в одном и том же классе эквивалентности имеют одинаковый базовый неориентированный граф и v-структуры (шаблоны дуг типа Xi → Xj ← Xk , без дуг между Xi и Xk ).

Изучение структуры с помощью максимизации баллов обычно основано на алгоритмах эвристической оптимизации общего назначения, адаптированных для использования преимуществ этих двух свойств для увеличения скорости изучения структуры (Scutari et al., 2019). Наиболее распространенными являются стратегии жадного поиска, такие как восхождение на холм и запретный поиск (Russell and Norvig, 2009), которые используют локальные ходы, предназначенные для воздействия только на одно или два локальных распределения на каждой итерации; другие варианты, рассмотренные в литературе, включают генетические алгоритмы (Larrañaga et al., 1996) и оптимизацию колонии муравьев (de Campos et al., 2002). Непосредственное изучение классов эквивалентности (в отличие от DAG) может быть выполнено аналогично алгоритму жадного поиска эквивалентности (GES; Chickering, 2002). Точная максимизация P (D | G) и BIC также стала возможной в последние годы благодаря все более эффективному сокращению пространства DAG и жестким ограничениям баллов (Cussens, 2012; Suzuki, 2017; Scanagatta et al., 2015).

Другим вариантом изучения структуры является использование тестов условной независимости для изучения ограничений условной независимости из D и, таким образом, для определения того, какие дуги следует включить в G. Полученные алгоритмы называются алгоритмами, основанными на ограничениях, в отличие от алгоритмов, основанных на оценке, которые мы представили в предыдущем параграфе; обзор и сравнение этих двух подходов см. В Scutari et al. (2018). Чикеринг и др. (2004) доказали, что алгоритмы, основанные на ограничениях, также являются NP-жесткими для неограниченных групп доступности данных; и фактически они эквивалентны алгоритмам, основанным на оценке, при фиксированном топологическом порядке узлов в G, когда ограничения независимости оцениваются с помощью статистических тестов, связанных с кросс-энтропией (Cowell, 2001).

Наконец, как только и G, и Θ были изучены, мы можем отвечать на запросы об интересующих нас величинах, используя полученный BN в качестве нашей модели мира. Распространенными типами являются запросы с условной вероятностью, в которых мы вычисляем апостериорную вероятность одних переменных при наличии свидетельств о других; и запросы с наиболее вероятным объяснением, в которых мы определяем конфигурацию значений некоторых переменных, которая имеет наибольшую апостериорную вероятность при данных значениях некоторых других переменных. Последнее особенно подходит для реализации как прогнозирования, так и вменения недостающих данных. Эти запросы могут быть автоматизированы для любого заданного BN с использованием точных или приближенных алгоритмов вывода, которые работают непосредственно с BN без необходимости каких-либо ручных вычислений; обзор таких алгоритмов см. В работе Скутари и Дениса (2014). Учитывая их способность кратко, но осмысленно представлять мир, автоматически отвечать на произвольные запросы и комбинировать данные и экспертные знания в процессе обучения, BNS может эффективно моделировать самые разнообразные явления. Однако их применимость не всегда очевидна для специалистов-практиков в других областях из-за сильного внимания литературы к простому сценарию, в котором данные являются статическими (в отличие от динамических, то есть с временным измерением) и полными (то есть полностью наблюдаемыми). Динамические данные занимают центральное место в ряде передовых приложений и исследований в таких разных областях, как генетика и робототехника; а неполные данные являются обыденным фактом практически при любом анализе данных в реальном мире. BNs может строго обрабатывать и то, и другое, как мы увидим ниже.

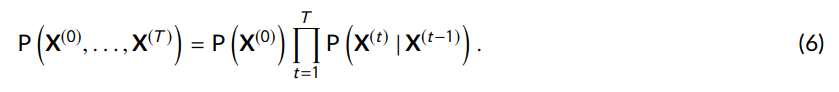
3 | ДИНАМИЧЕСКИЕ БАЙЕСОВСКИЕ СЕТИ

Динамическая BNS (DBNs4) объединяет классическую (статическую) BNs и марковские процессы для моделирования динамических данных, в которых каждый отдельный объект многократно измеряется с течением времени, например, продольных или панельных данных. Доступное введение приведено, например, в работе Murphy (2002). Они находят широкое применение в инженерии (Павлович и др., 1999; Фриго и др., 2008), медицине (ван дер Хейден и др., 2014), генетике и системной биологии (Перрен и др., 2003). Термин “динамический” в данном контексте подразумевает, что мы моделируем динамическую систему, не обязательно, что сеть меняется с течением времени.

4 Сбивчиво, дискретные BN иногда называют DBN, чтобы соответствовать гауссовским BN, которые называются GBN.

3.1 | Определения и свойства

Для простоты давайте сначала предположим, что мы работаем в дискретном времени: наша система состоит из одного набора X (t ) случайных величин для каждого из t = 1, . . . , T моментов времени. Мы можем смоделировать его как DBN с марковским процессом вида



где P(X(0)) задает начальное состояние процесса , а P(X(t) | X(t -1)) определяет переход между временами t - 1 и t.

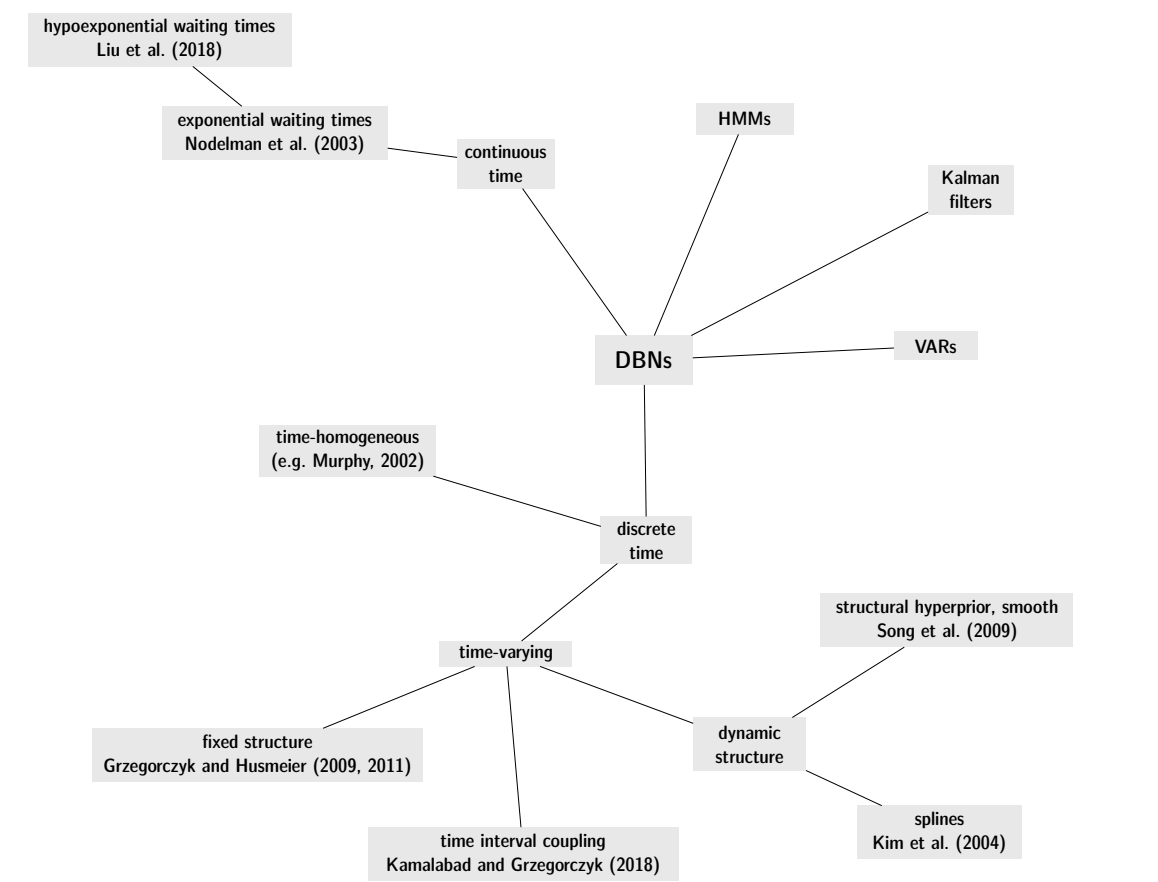
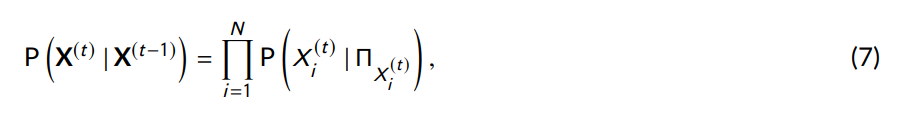


Рисунок 1 – Диаграмма демонстрирующая отношение между статьями и способами указанными в пункте 3.

Мы можем смоделировать этот переход с помощью 2-кратного BN (2TBN), определенного над (X(t -1), X(t)), в котором мы, естественно, предполагаем, что любая дуга между узлом в t - 1 и узлом в t обязательно должна быть направлена в сторону последнего, следуя стрелке времени. При моделировании X(t) узлы в X(t -1) появляются только при кондиционировании; мы считаем, что они по существу фиксированы и не имеют свободных параметров, поэтому оставляем их в качестве корневых узлов. В конце концов, X(t -1) будет случайным в P (X(t -1) | X(t -2)), и было бы несовместимо с (6) рассматривать X(t -1) как случайную величину дважды! Тогда, следуя (1), мы можем записать



и мы обычно предполагаем, что параметры, связанные с локальными распределениями, не меняются со временем, чтобы сделать процесс однородным по времени.

Такой выбор обусловлен простотой вычислений и статистики: в построении DBN нет внутренних ограничений, которые мешали бы им моделировать зависимости, уходящие еще дальше в прошлое, или, если уж на то пошло, тенденции или сезонность. (Смотрите Рисунок 1 для получения обзора моделей, которые мы подробно обсудим ниже.) Гжегорчик и Хусмейер (2009, 2011), например, сконструировали нестационарные неоднородные базы данных для моделирования непрерывных данных с использованием точек изменения для захвата различных режимов в разные промежутки времени; они уравновесили эту повышенную шестнадцатеричность, потребовав, чтобы G было одинаковым для всех режимов, позволяя изменяться только параметрам локальных распределений. Позже это ограничение было смягчено в Камалабаде и Гжегорчике (2018), разрешив изменять как G, так и параметры, но связав их между смежными временными интервалами. Гиперпараметрам, управляющим связью, было разрешено варьироваться между сегментами, следуя обратному Гамма-гиперприорированию. Робинсон и Хартеминк (2010) также определили нестационарную DBN с точками изменения и связанными наборами изменений arc, с усеченными геометрическими параметрами размера наборов изменений, количества и интервала между точками изменения; и Song et al. (2009) построили DBN, которые изменяются плавно (не кусочно) с течением времени как по структуре, так и по параметрам. Ким и др. (2004) представили еще более гибкую модель, которая использовала сплайновую регрессию с B-сплайнами для идентификации ПXi(t).Дополнение временных (панельных) данных нестационарными (поперечными) данными для изучения DBN также было исследовано Ляхдесмяки и Шмулевичем (2008) с использованием оценки, аппроксимирующей результирующую трудноразрешимую вероятность.

Моделирование DBN как процессов с дискретным временем также является выбором, мотивированным математической простотой. Нодельман и др. (2003) первоначально предложили класс DBN непрерывного времени (CTBN) с независимым экспоненциальным временем ожидания и дискретными узлами; они однозначно идентифицируются, поскольку все дуги не мгновенны, а также имеют предельную вероятность замкнутой формы. Совсем недавно эта работа была расширена в Liu et al. (2018) путем замены экспоненциального времени ожидания гипоэкспоненциальностями, чтобы лучше отражать поведение данных в нескольких областях. DBN с дискретным временем, безусловно, проще, чем любая из этих CTBN, но эта математическая простота влечет за собой важные практические последствия. Во-первых, чтобы работать в дискретном времени, мы должны выбрать единый временной шаг (промежуток времени между t − 1 и t ) для всей DBN; но во многих явлениях реального мира разные переменные могут иметь очень разную временную детализацию, и эти временные детализации также могут меняться в ходе сбора данных, что делает какой-либо единый выбор временного шага нецелесообразным. Во-вторых, выбор временного шага может затушевывать динамику явления. Смысл использования дискретного времени заключается в том, что мы агрегируем все изменения состояния в DBN на протяжении всего каждого временного шага. С одной стороны, если переменные развиваются медленнее, чем шаг по времени, мы вынуждены моделировать DBN как марковский процесс более высокого порядка5, что приводит к гораздо более сложной модели. С другой стороны, если переменные развиваются более быстрыми темпами, чем временной шаг, такое усреднение эффективно скроет изменения состояния и их взаимодействие в единой сводной статистике; следовательно, DBN обеспечит очень слабую аппроксимацию лежащего в основе явления. Более того, если мы не сможем правильно идентифицировать , это может привести к рекурсивной привязке к родительским элементам результирующего в DBNs , которые намного сложнее, чем реальное лежащее в основе явление, и которые трудно изучить на основе ограниченных данных.

Частичное решение последней проблемы состоит в том, чтобы позволить родителям находиться либо в одно и то же время, либо в предыдущий момент времени, моделируя DBN как марковский процесс первого порядка. Когда наблюдения представляют собой средние или агрегированные измерения за период времени (скажем, t-й момент времени соответствует объему данных за t-ю неделю), имеет смысл допускать мгновенные зависимости между переменными в один и тот же момент времени, поскольку мгновенность зависимости - это всего лишь фикция, вытекающая из определения нашей модели. С другой стороны, когда наблюдения фактически соответствуют мгновенным измерениям (например, от синхронизированных датчиков), в модели обычно допускаются только не мгновенные зависимости на том основании, что обусловливание должно предшествовать по времени обусловленному событию  С точки зрения причинности мы можем аналогичным образом утверждать, что каждое из может вызывать только в том случае, если оно предшествует по времени; если это Π (t) происходит в тот же момент времени, что и X (t ), тогда то, что мы моделируем, является совпадением, а не причинностью. Это основная идея причинности Грейнджера (Granger, 1968), в которой говорится, что можно сказать, что один временной ряд (такой как ) оказывает причинное влияние на второй временной ряд (такой как ) тогда и только тогда, когда включение прошлых знаний о первом повышает точность прогнозирования для последнего. Следовательно, разрешение мгновенных зависимостей заметно усложняет причинно-следственные рассуждения в DBN. То же самое верно и для изучения структуры DBN в первую очередь: изучение общего DBN является NP-сложным (Чикеринг и Хекерман, 1994; Чикеринг и др., 2004), в то время как изучение DBN, содержащего только не мгновенные зависимости, - нет (Dojer, 2006).

5A Случайный процесс является марковским процессом порядка L, если X(t ) зависит только от X(t -1), . . . X(t −L) и не зависит от X(t −L−1), . . . , X(0). Чем выше порядок, тем дальше в прошлое могут уходить зависимости.

6 Эту проблему часто называют запутанностью.

Интуитивно, пространство число возможных DBN намного меньше, если мы не допускаем мгновенных зависимостей, потому что существует меньше дуг-кандидатов, которые мы можем включить, и потому что их направления фиксированы, чтобы следовать стреле времени и причинности Грейнджера 7.

7 Интересно, что при небольшом количестве доступных временных точек вывод, основанный на DBNS, является более точным, чем методы, непосредственно основанные на причинно-следственной связи Грейнджера; но обратное верно для более длинных временных рядов (Цзоу и Фэн, 2009).

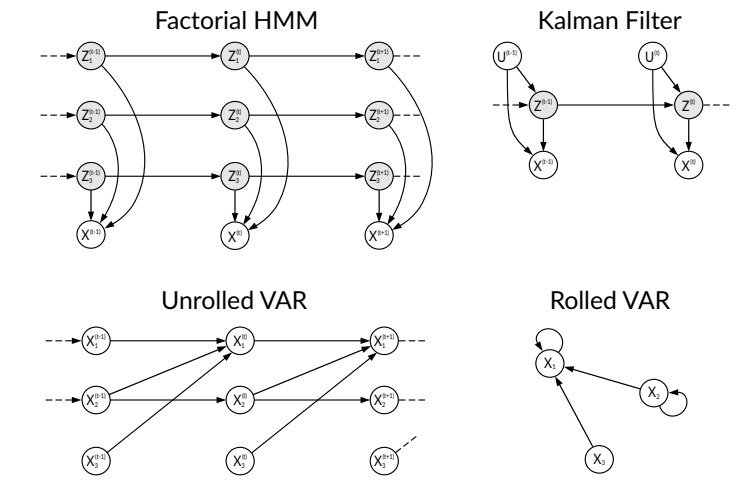


Рисунок 2- Динамические модели, представленные в виде BNS: факториал HMM от Гахрамани и Джордана (1996, вверху слева),

простая модель фильтра Калмана от Мерфи (2002, вверху справа) и модель VAR порядка 1 как в развернутом (внизу слева), так и

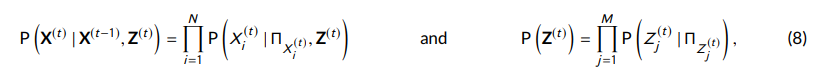
в свернутом виде (внизу справа). Заштрихованные узлы соответствуют скрытым переменным.

3.2 | Модели, которые могут быть выражены в виде DBN

DBN обладают большой выразительной способностью, и они включают в себя и обобщают множество классических моделей, которые по отдельности изучались в литературе. Здесь мы обсудим три такие модели: скрытые марковские модели, векторные авторегрессионные модели и фильтры Калмана. Скрытые марковские модели и фильтры Калмана можно рассматривать как частные экземпляры моделей пространства состояний, и их представления DBN проясняют взаимосвязь между этими двумя классами моделей. Представление этих моделей в виде DBN имеет преимущества, помимо того, что делает их сравнение более удобным; это позволяет использовать вероятностный механизм, который мы рассмотрели в разделах 2 и 3.1, когда он более удобен или эффективен с вычислительной точки зрения, чем альтернативы (как в случае вывода в скрытых марковских моделях; Murphy, 2002). Некоторые примеры показаны на рисунке 2 и будут проиллюстрированы ниже.

Скрытые марковские модели (HMMs; Цуккини и Макдональд, 2009) являются одним из наиболее распространенных подходов к моделированию явлений со скрытым состоянием, то есть, при котором поведение наблюдаемых переменных X зависит от поведения одной или нескольких дискретных скрытых переменных Z, а также от других переменных в X. Этот сценарий обычно возникает, когда технические, экономические или этические соображения делают невозможным полное наблюдение за состоянием, лежащим в основе интересующего явления. Известными примерами являются вменение (Marchini and Howie, 2010) и фазирование (Delaneau et al., 2012) в исследованиях общегеномных ассоциаций из-за ограничений технологии зондирования и маркировки ДНК; отслеживание животных в экологии (Patterson et al., 2008), когда мы можем наблюдать за их перемещениями с помощью радиомаяков, но не за их поведением; и подтверждение массовых миграций на протяжении истории путем объединения археологических артефактов и образцов древней ДНК (Шиффельс и Дурбин, 2014). До появления глубоких нейронных сетей HMM также были предпочтительной моделью для распознавания речи (Гейлс и Янг, 2008) и рукописного ввода (Плетц и Финк, 2009).

В терминах DBN типичная модель HMM с M скрытыми переменными может быть записана как

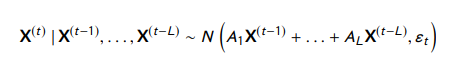


с тем ограничением, что родителями могут быть только другие скрытые переменные. Предполагается, что скрытые переменные являются дискретными; и в подавляющем большинстве литературы наблюдаемые переменные также считаются дискретными. В зависимости от выбора , мы можем получить различные варианты HMM , такие как иерархические HMM (Fine et al., 1998), в которых каждый определяется как HMM сам по себе для создания многоуровневой стохастической модели; и факториальные HMM (Ghahramani and Jordan, 1996), в которых X(t) определяется конфигурацией набора взаимно независимых (показано на рисунке 2, верхняя левая панель).

Векторные авторегрессионные модели (VARs; Box et al., 2016) представляют собой простое многомерное расширение одномерных авторегрессионных временных рядов для непрерывных переменных. Таким образом, их основные области применения - прогнозирование в финансах (Bańbura et al., 2010) и совсем недавно - в анализе данных ФМРТ (Gates et al., 2009). Они определяются как



для некоторого фиксированного марковского порядка L. Мы можем переписать (9) как



а затем ограничить родительские значения каждого X (t ) теми, для которых соответствующие коэффициенты регрессии в A1, . . . , AL отличаются от нуля, используя взаимно однозначное соответствие между коэффициентами регрессии и частичными корреляциями (Weatherburn, 1961). Формально, , что позволяет записать (9) в форме, аналогичной формула (6), и получить гауссовский DBN. Та же конструкция используется Сонгом и др. (2009) для их неоднородных моделей DBN, которые параметризуются по мере обработки VAR и оцениваются с использованием регрессий со штрафом L1. В частном случае, когда L = 1, переменные могут быть графически представлены двумя эквивалентными способами, показанными на нижних панелях рисунка 2: “развернутая” DBN, в которой каждый узел соответствует одному ; и более компактная “свернутая” DBN, в которой каждый узел соответствует переменной Xi, дуга от Xi до Xj подразумевает . Дуга от Xi к самой себе подразумевает



Фильтры Калмана (KFS; Hamilton, 1994) сочетают в себе черты как HMM, так и VARs, как подробно обсуждалось в Roweis и Ghahramani (1999) и Ghahramani (2001): как и VARs, они являются линейными гауссовыми DBN; но у них также есть скрытые переменные, такие как HMM. Они широко используются для фильтрации (то есть шумоподавления) и прогнозирования в системах позиционирования GPS (Work et al., 2008); атмосферного моделирования и прогнозирования погоды (Cassola and Burlando, 2012); и сейсмологии (Sakaki et al., 2010).

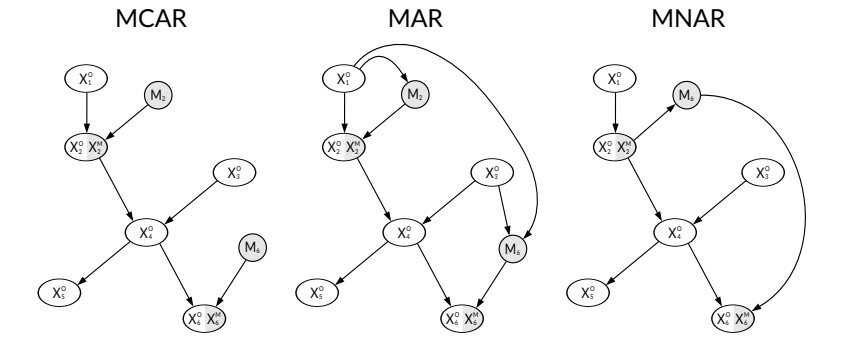


Рисунок 3 - BN - изображения моделей отсутствия CAR (слева), MAR (в центре) и MNAR (справа) из Рубин (1976) и Литтл и Рубин (1987). Заштрихованные узлы и часть узлов соответствуют переменным, которые не наблюдаются в данных.

В своей простейшей форме (см. Рис. 2, верхняя правая панель) KFS включают слой одного или нескольких скрытых переменных, моделирующие ненаблюдаемую часть явления,

учет одной или нескольких наблюдаемых переменных



с добавлением независимого гауссова шума в обоих слоях. Оба уровня часто включают дополнительные (непрерывные) объясняющие переменные U, а также могут быть дополнены (дискретными) переключающими переменными для обеспечения различных режимов, как в Grzegorczyk и Husmeier (2009, 2011). Если мы исключим последнее, предполагается, что система совместно является гауссовой: это позволяет оформить KFs как DBN так же, как мы сделали для VARs.

4 | БАЙЕСОВСКИЕ СЕТИ НА ОСНОВЕ НЕПОЛНЫХ ДАННЫХ

Подавляющее большинство литературы по изучению BNS основано на предположении, что D является полным, то есть набором данных, в котором каждая переменная имеет наблюдаемое значение для каждой выборки. Однако в реальных приложениях нам часто приходится иметь дело с неполными данными; некоторые выборки будут полностью соблюдены, в то время как другие будут содержать недостающие значения для некоторых переменных. Хотя возникает соблазн просто вменить недостающие значения в качестве шага предварительной обработки, давно известно, что даже довольно сложные методы, такие как вменение методом горячей деки, проблематичны в многомерных условиях (Kalton and Kasprzyk, 1986). Простое удаление неполных выборок также может повлиять на процесс обучения, в зависимости от того, насколько сильно отсутствуют недостающие данные.

Рубин (1976) и Литтл и Рубин (1987) формализовали три возможных паттерна (или механизма) отсутствия (показаны на рисунке 3):

• Полное случайное отсутствие (MCAR): когда полные выборки неотличимы от неполных. Другими словами, вероятность того, что значение будет отсутствовать, не зависит как от наблюдаемых, так и от пропущенных значений. Например, на левой панели рисунка 3 и X2, и X6 видны лишь частично; следовательно, формула. Закономерности отсутствия данных формула контролируются M2 (для X2) и M6 (для X6) и являются полностью случайными; скажем, M2 и M6 являются двоичными переменными, которые кодируют вероятность того, что два прибора (независимо) выйдут из строя и, следовательно, не смогут измерить X2 и X6 для некоторых людей.

• Случайное отсутствие (MAR): случаи с неполными данными отличаются от случаев с полными данными, но характер отсутствия можно предсказать по другим наблюдаемым переменным. Другими словами, вероятность того, что значение будет отсутствовать, является функцией наблюдаемых значений. Примером может служить центральная панель на рисунке 3: по сравнению с левой панелью, теперь мы знаем, что два прибора, скорее всего, выйдут из строя, когда (скажем) наблюдаются высокие значения X1 (и X3, в случае X6).

• Отсутствие не случайно (MNAR): характер отсутствия не является случайным или его нельзя предсказать по другим наблюдаемым переменным; вероятность того, что запись будет отсутствовать, зависит как от наблюдаемых, так и от пропущенных значений. Распространенными примерами являются переменные, которые систематически отсутствуют или для которых характер отсутствия зависит от самих отсутствующих значений. На правой панели рисунка 3 скажите, что X2 подвергается цензуре (то есть оно никогда не наблюдается, если его значение выше фиксированного порога); но когда X2 отсутствует, вероятность того, что отсутствует X6 (кодируется M6), также чрезвычайно высока. Поскольку мы никогда не наблюдаем X2, когда отсутствует X6, мы не можем правильно смоделировать эту взаимосвязь; насколько нам известно, и X2, и X6 могут отсутствовать из-за какого-то общего внешнего фактора, поскольку в данных они, по-видимому, отсутствуют вместе.

MCAR и MAR являются игнорируемыми моделями отсутствия; вероятность того, что какое-то значение отсутствует, может зависеть от наблюдаемых значений, но не от отсутствующих значений, и, таким образом, может быть правильно смоделирована. Если мы обозначим с помощью DO и DM наблюдаемую и ненаблюдаемую части D и сгруппируем все бинарные индикаторы отсутствия Mi в M с параметрами Ξ, то мы можем записать

если отсутствуют данные MAR, то M зависит только от DO,



и если отсутствуют данные MCAR, то M не зависит ни от DO, ни от DM,

.

В обоих случаях можно смоделировать M на основе имеющихся данных. Однако это не относится к MNAR, поскольку M зависит от ненаблюдаемого DM.

Моделирование неполных данных является аналитически трудноразрешимым и непомерно трудоемким в вычислительном отношении по сравнению со сценарием с полными данными: точный анализ требует вычисления совместного апостериорного распределения Θ с учетом всех возможных завершений D. Однако выполнение D имеет вычислительную сложность, которая экспоненциально возрастает с увеличением количества пропущенных записей, поскольку оно включает в себя нахождение набора пропущенных завершений данных с наибольшей совместной вероятностью, учитывая наблюдаемые данные. Рассмотрение только наиболее вероятного завершения также может вызвать чрезмерную уверенность в результатах анализа, поскольку завершенные наблюдения будут иметь меньшую вариабельность в зависимости от конструкции; но полное усреднение байесовского вывода по всем возможным Θ и завершенному D настолько сложны с вычислительной точки зрения, что неосуществимы даже в простых настройках.

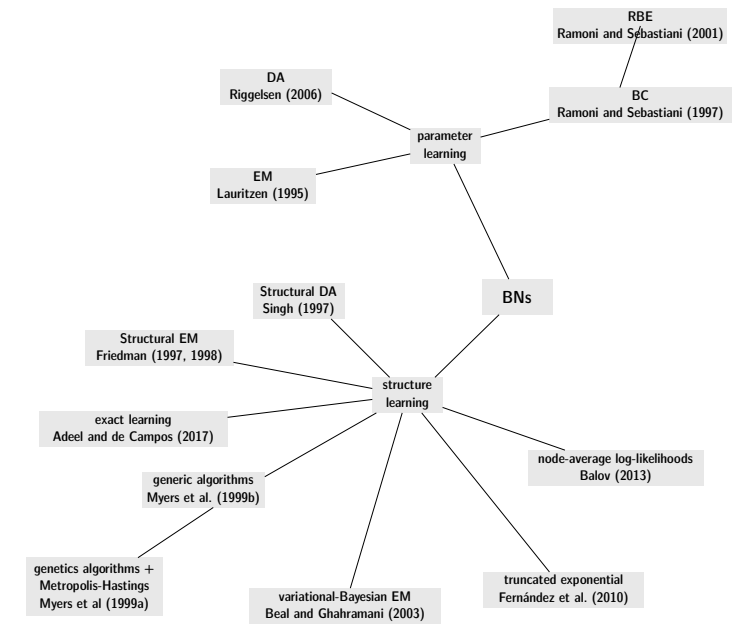


Рисунок 4 – Диаграмма, иллюстрирующая взаимосвязь между документами и подходами, рассмотренными в разделе 4.

Кроме того, эта совместная максимизация нарушает предположение о независимости параметров и делает невозможным определение разложимых оценок для изучения структуры, не прибегая к некоторому приближению.

4.1 | Изучение параметров

Было разработано множество подходов для изучения параметров на основе неполных данных с заданной фиксированной известной структурой8, начиная от итерационных методов, таких как увеличение данных (DA; Таннер и Вонг, 1987) и алгоритм максимизации математического ожидания (EM; Лауритцен, 1995), до методов, основанных на вероятностных интервалах, таких как привязка и Коллапс (BC; Рамони и Себастьяни, 1997) и неопределенных вероятностях (de Campos et al., 1994). (Смотрите Рисунок 4, чтобы получить обзор тех, которые мы подробно обсудим ниже.) Эти методы обычно предполагают, что отсутствующие данные являются MCAR или MAR для работы с P(Θ, DM | G, DO ), и их точность резко снижается, если это не так (Spiegelhalter and Cowell, 1992); однако было обнаружено, что BC надежен и для данных MNAR. Онисько и др. (2002) также отметили, что простые подходы к вменению могут хорошо зарекомендовать себя при изучении параметров BN при фиксированной, разреженной сетевой структуре, что еще больше расширяет доступные варианты.

8 Хотя почти вся литература, посвященная BNs, предполагает наличие дискретных данных.

В контексте изучения параметров алгоритм EM сохраняет свою классическую структуру:

• этап математического ожидания (E) состоит в вычислении ожидаемых значений достаточной статистики (такой как количество nij k в дискретных BN, частичные корреляции в GBN) с использованием точного вывода в соответствии с описанными выше линиями для использования как неполных, так и полных выборок;

• шаг максимизации (M) берет достаточную статистику из шага E и оценивает параметры BN либо с использованием максимального правдоподобия, либо байесовских апостериорных оценок.

Оценки параметров затем используются на следующем электронном шаге для обновления ожидаемых значений достаточной статистики; повторные итерации этих двух шагов в конечном итоге вернут максимальную вероятность или максимальные апостериорные оценки для параметров. Используя обозначение (10), шаг E эквивалентен вычислению E(DM , DO | G, Θ), а шаг M - максимизации P(Θ | G, DO , DM ).

DA очень похож, но вместо итеративного сведения к одному набору оценок параметров он использует выборку Гиббса (Geman and Geman, 1984) для генерации значений из апостериорных распределений как DM, так и Θ. Эти два шага заключаются в следующем:

• на этапе условного вычисления (I) данные дополняются значениями, полученными на основе прогнозных распределений пропущенных значений;

• на этапе оценки параметра (P) значение параметра извлекается из апостериорного распределения Θ в зависимости от завершенных данных с I-го этапа.

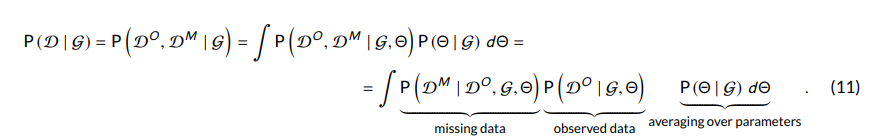
Более формально мы определяем расширенный вектор параметров { DM , Θ}, содержащий как отсутствующие значения, так и параметры BN. Учитывая начальный набор значений, он обновляет каждый элемент { DM , Θ} путем выборки нового значения для каждого пропущенного значения из P(DM | DM , Θ), and by sampling a new value for each parameter from P(Θi | Θ−i , DM ), each формула, по очереди. После начальной фазы выгорания этот процесс сойдет к своему стационарному распределению и вернет реализации параметров из апостериорного распределения Θ в зависимости от наблюдаемых данных. Аналогичный подход к выборке Гиббса был предложен совсем недавно Риггельсеном (2006): он выбирает из более простого, приблизительного прогнозируемого распределения и использует веса для реализации эффективной схемы выборки по важности. Кроме того, веса позволяют использовать выборки, сгенерированные на этапе выгорания, а также выборки из стационарного распределения выборок Гиббса, поскольку они взвешивают выборки в соответствии с их предполагаемой точностью прогнозирования.

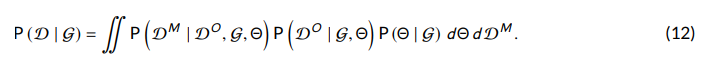
Наконец, BC и его преемник, надежный байесовский оценщик (RBE; Рамони и Себастьяни, 2001), используют дискретную природу категориальных переменных для получения грубых интервальных оценок условных вероятностей, полученных из неполных переменных, которые затем сводятся к точечным оценкам либо с помощью выпуклой комбинации границ интервалов, экспертных знаний, либо и того, и другого. Первый (связанный) шаг использует тот факт, что каждое πi k | j может быть ограничено ниже, предполагая, что ни одно из пропущенных значений для Xi не дополняется их k-м значением при соблюдении j-й родительской конфигурации; и что πi k | j может быть ограничено выше, предполагая, что все пропущенные значения дополняются их k-м значением. Достоинством этого подхода является отсутствие каких-либо предположений относительно распределения отсутствующих данных. Кроме того, ширина каждого интервала обеспечивает четкое представление о надежности оценок, которые могут быть приняты во внимание при выводах и прогнозировании. Второй (свернутый) шаг предполагает, что отсутствующие данные являются MAR или MCAR, чтобы иметь возможность вычислить ожидаемые завершения для неполных выборок, которые затем используются для вычисления среднего значения и дисперсии πi k | j . Интересно, что интервалы от шага привязки могут быть использованы для увеличения выборки как EM, так и Гиббса и получения более точных выводов, но точность прогнозирования RBE оказалась выше, чем у Ramoni и Sebastiani (2001). Аналогичное исследование в контексте классификаторов BN можно найти в работе Peña et al. (2000). Концептуально аналогичный подход был также предложен Ляо и Джи (2009), которые сначала использовали качественные экспертные знания о значениях параметров для их привязки, а затем оценили их значения с помощью выпуклой оптимизации, встроенной в алгоритм EM.

Обратите внимание, что мы можем работать со скрытыми переменными, используя аналогичные подходы, пока значение G фиксировано, и нам просто нужно изучить Θ. Недавний пример приведен в работе Ямадзаки и Мотомуры (Yamazaki and Motomura, 2019), которые показывают, что можно изучить область скрытой дискретной переменной, а также связанные с ней параметры, если за ней наблюдали родители и дочерние элементы. Несколько других примеров обсуждаются в контексте DBNs в работе Murphy (2002).

4.2| Структурированное Обучение

Изучение структуры BN на основе неполных данных во многих отношениях является расширением методов, описанных в разделе 4.1. (См. Рисунок 4 для обзора подходов, которые мы подробно обсудим ниже, а также подходов к изучению параметров.) В своем общем виде это неосуществимо с вычислительной точки зрения, потому что нам нужно выполнить совместную оптимизацию по отсутствующим значениям и параметрам для оценки каждой сети-кандидата. Начиная с (4), мы можем сделать это очевидным, переписав P(D | G) как функцию DO , DM :

Из этого выражения мы можем видеть, что для максимизации P (D | G) мы должны совместно максимизировать вероятность наблюдаемых данных DO и вероятность недостающих данных DM с учетом наблюдаемых данных для каждого кандидата G и усреднения по всем возможным Θ. Это дает нам максимальное апостериорное решение для структурирования обучения; полный байесовский подход также потребовал бы усреднения по всем возможным конфигурациям отсутствующих данных, что привело бы к



По сравнению с (11), (12) содержит одно дополнительное измерение для каждого пропущенного значения (в дополнение к одному измерению для каждого параметра в Θ) и, следовательно, является слишком многомерным для вычисления в практических приложениях. Дополнительная проблема заключается в том, что, хотя P(DO | G, Θ) разлагается, как в (5), P(DM | DO, G, Θ) в общем случае этого не делает.

Чтобы обойти эти вычислительные проблемы, в литературе были рассмотрены два возможных подхода: итеративное заполнение и уточнение данных, использование стандартных алгоритмов и оценок для полных данных; или использование функций оценки, которые приближаются к BIC и P (D | G), но которые поддаются разложению и могут быть эффективно вычислены даже на неполных данных.

Структурный алгоритм EM (SEM9; Friedman, 1997) является наиболее известной реализацией первого подхода; он имеет важные приложения в филогенетике (Friedman et al., 2002), истории болезни (van der Heijden et al., 2014) и анализе клинических испытаний (Liew et al., 2019). SEM делает изучение структуры выполнимым с вычислительной точки зрения за счет поиска наилучшей структуры внутри EM, вместо встраивания EM в алгоритм изучения структуры. Он состоит из двух этапов, подобных классическому ЭМ:

• на этапе E мы дополняем данные, вычисляя ожидаемую достаточную статистику с использованием текущей структуры сети;

• на M-шаге мы находим структуру, которая максимизирует ожидаемую функцию оценки для завершенных данных.

9 Эта аббревиатура является еще одним источником путаницы, поскольку SEM также может обозначать “модели структурных уравнений”, которые тесно связаны с BNs. Обсуждение их сходств и различий см. в работе Гупты и Ким (2008).

Поскольку при подсчете очков на M-шаге используются завершенные данные, изучение структуры может быть эффективно реализовано с использованием стандартных алгоритмов. Первоначальное предложение Фридмана (1997) использовало BIC и жадный поиск; позже Фридман (1998) расширил SEM до полностью байесовского подхода, основанного на апостериорных оценках, и доказал сходимость полученного алгоритма.

Фактически, на M-шаге может использоваться любая комбинация алгоритма изучения структуры и оценки; совсем недавно Scanagatta et al. (2018) предложили обучающие BNS со структурой ограниченной ширины дерева, используя свой алгоритм k-MAX и вариант BIC. (Это имеет двойное преимущество, заключающееся в ускорении M-шага и получении BNS, для которых заполнение данных на E-шаге происходит относительно быстро.) Сингх (1997) предложил аналогичный подход, основанный на DA, генерирующий наборы завершенных наборов данных и усредняющий полученные изученные сети на каждой итерации. Майерс и др. (1999b) также решили итеративно изучать как сетевые структуры, так и недостающие данные одновременно, но сделали это с использованием эволюционных алгоритмов и кодирования обоих в виде “генов”. Следовательно, особи в эволюционирующей популяции включают в себя как завершенный набор данных, так и связанный с ним BN. Это было объединено с Metropolis-Hastings для ускорения обучения, как описано в Myers et al. (1999a).

Совсем недавно Адил и де Кампос (2017) предложили точный алгоритм обучения, который явно моделирует шаблоны пропусков с помощью вспомогательных переменных, которые включены в BN как отдельные узлы, а не просто как вычислительные устройства. Кроме того, они показали, что его вычислительная сложность такая же, как и у других алгоритмов точного обучения для полных данных; и они адаптировали предложенный алгоритм к (более быстрой) эвристике, которая затем оказалась непротиворечивой.

Вторая группа подходов включает вариационно-байесовский EM от Била и Гахрамани (2003), который максимизирует вариационное приближение P (D | G), что, в свою очередь, является нижней границей истинного предельного правдоподобия. Балов (2013) доказал, что изучение структуры с помощью BIC не согласуется даже при MCAR, и предложил заменить его логарифмическими вероятностями со средним значением по узлу, вычисленными на основе локально полных наблюдений. Альтернатива состоит в использовании приближений, основанных на смесях усеченных экспонент, как было продемонстрировано в Fernández et al. (2010): они объединили EM и DA, чтобы соответствовать регрессионным BNS со структурами, которые напоминают наивные байесовские классификаторы и дополненные деревом наивные байесовские классификаторы, и аппроксимируют объясняющие переменные. Доступное введение в эту область исследований представлено в работах Чикеринга и Хекермана (1997) и Хекермана (1997), в которых математически подробно описывается взаимосвязь между P (D | G), ее приближением Лапласа и BIC.

5 | ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ И ПРИКЛАДНЫЕ ПРОГРАММЫ

В заключение этого обзора мы рассмотрим практические аспекты использования моделей BN, рассмотренных выше, для исследования реальных проблем: доступности программных реализаций и примечательных примеров приложений в различных областях.

5.1 | Доступное программное обеспечение

Среди рассмотренных нами статей DBN только Гжегорчик и Хусмейер (2009, 2011) и (Ляхдесмяки и Шмулевич, 2008) предоставляют реализации методов, которые они предлагают, в виде сценариев Matlab, доступных по запросу авторов. Кроме того, Liu et al. (2018) предоставляют ссылки на пользовательское программное обеспечение, которое они используют в статье. Хьюджин (Andersen et al., 1989) реализует однородные DBN с дискретным временем, как определено в (6), как и GeNIe (BayesFusion, 2019).

Программные реализации для обработки неполных данных более доступны, хотя обычно их нет в статьях, предлагающих методы. И Hugin , и GeNIe реализуют EM для изучения параметров, но не Structural EM для изучения параметров. Пакеты R bnlearn (Scutari, 2010) и bnstruct (Franzin et al., 2017) реализуют структурный алгоритм EM; другая реализация в Matlab доступна как часть дополнительного материала (Rancoita et al., 2014). Адил и де Кампос (2017) позже использовали его для реализации своего точного алгоритма обучения вместе с Gobnilp (Бартлетт и Куссенс, 2013); и (Риггельсен, 2006) использует пользовательскую реализацию на C ++ своего подхода к выборке Гиббса. Подходы, предложенные Баловым (2013), реализованы в пакете catnet R (Балов и Зальцман, 2019), а подходы, предложенные Фернандесом и др. (2010), включены в Elvira (Консорциум Elvira, 2002).

5.2| Известные приложения

Динамические и неполные данные распространены во многих прикладных областях, в которых используются BNS; необходимость работы с такими данными является основной мотивацией для разработки подходов, которые мы рассмотрели в разделах 3 и 4. Поэтому многие из них нашли применение в передовых областях, особенно в клинических исследованиях и науках о жизни.

Это особенно верно для DBN: поскольку они включают в себя ряд статистических моделей, которые сами по себе очень популярны (см. Раздел 3.2), было бы справедливо сказать, что они используются в большинстве отраслей науки под разными обличьями. Однако их основными областями применения в литературе остаются генетика и системная биология, и в частности моделирование клеточных сигнальных путей в форме сетей по мере их эволюции с течением времени. Это включает в себя одновременное исследование состояния нескольких белков или других омических соединений во времени и изучение их взаимодействий и того, как они эволюционируют в ответ на внешние раздражители. Эксперименты In vitro, в которых клеточные линии исследуются в контролируемой среде, были смоделированы как однородные DBN (см., например, Hill et al., 2012; Perrin et al., 2003). С другой стороны, более сложные эксперименты in vivo часто включают молекулярные системы, которые не находятся в стационарном режиме и, следовательно, требуют неоднородных DBN. Некоторые примеры, основанные на моделях, представленных в Grzegorczyk and Husmeier (2009, 2011), Kamalabad and Grzegorczyk (2018) и других работах тех же авторов, изучают эмбриональные стволовые клетки (Mathew et al., 2015); сепсис как причину острого повреждения легких (Emr et al., 2014); связь тупой травмы и травматического повреждения спинного мозга с воспалением (Zaaqoq et al., 2014) и гипотензией (Almahmoud et al., 2015). Другие данные omics, включая транскриптомику (López-Kleine et al., 2013) и данные микробиома (McGeachie et al., 2016; Lugo-Martinez et al., 2019), также были изучены с использованием DBNs.

DBN также нашли применение в экологии, в частности, при изучении динамики видов. Подобно человеческому организму, экосистема действует как сложная система, в которой виды взаимодействуют друг с другом (например, посредством хищничества или конкуренции) и с окружающей средой, в которой они живут. Результирующая динамика была смоделирована в работе Aderhold et al. (2012) с использованием неоднородного DBN от Grzegorczyk и Husmeier (2011). DBN, который также корректирует пространственную корреляцию за счет использования скрытых переменных, был разработан Трифоновой и др. (2015) для изучения динамики видов рыб в 7 географически и временно изменяющихся районах Северного моря. Кроме того, экосистемы изучались с использованием DBN также с точки зрения наук об окружающей среде, учитывая, например, влияние изменения климата на подземные воды (Молина и др., 2013) и способы наилучшего управления водохранилищами в условиях нечастых дождей (Роперо и др., 2017).

Наконец, третий класс применений DBNs - моделирование механических систем в инженерных исследованиях, исследованиях надежности и контроля качества. Например, изучение износа стальных конструкций, подверженных усталости (Luque and Straub, 2016), включая военные корабли (Liang et al., 2017); внедрение мониторинга надежности подводных труб в режиме реального времени (Cai et al., 2015) и промышленных систем с ЧПУ (Tobon-Mejia et al., 2012); обнаружение неисправностей, идентификация и восстановление в автономных космических аппаратах (Codetta-Raiteri and Portinale, 2014); и прогнозирование динамики движения с помощью зондов в Сан-Франциско (Hoheitner et al., 2012) В каждом из этих приложений, DBNs фиксируйте взаимодействие компонентов механических систем, которые они моделируют, и в то же время учитывайте воздействие окружающей среды, что позволяет операторам (или самой системе) скорректировать режим ее работы или полностью остановить ее, прежде чем вызвать какие-либо повреждения или проблемы безопасности. Аналогичным образом моделировались программные системы, например, для изучения взаимосвязи между уязвимостями системы безопасности и векторами атак в сетевой безопасности (Frigault et al., 2008).

Что касается неполных данных, то они являются настолько широко распространенной проблемой в науке и технике, что трудно определить какую-либо конкретную область, в которой они распространены; следовательно, подходы к обучению BN, описанные в разделе 4, нашли в литературе очень разное применение. Однако мы отмечаем, что анализ клинических данных исторически был одним из ключевых приложений, которые стимулировали методологические разработки в этой области. Ранее мы упоминали изучение истории болезни пациентов с хронической обструктивной болезнью легких (van der Heijden et al., 2014) и процесса выздоровления пациентов с нарушениями, связанными с хлыстовой травмой (Liew et al., 2019); двумя другими примерами являются диагностика заболеваний головного мозга, таких как деменция и болезнь Альцгеймера (Seixas et al., 2014) и апноэ во сне (Ryynänen et al., 2018). Также были исследованы совершенно другие области применения: одним из примеров является классификация радиолокационных изображений (Ma et al., 2016).

6 | КРАТКОЕ СОДЕРЖАНИЕ

В этой статье мы рассмотрели фундаментальные определения и свойства BNs, а также то, как BNs можно использовать для кодирования более сложных вероятностных моделей, чем то, что может быть очевидно из справочных материалов и большей части литературы. В обоих случаях основное внимание уделяется простому случаю, в котором моделируемые данные являются как статическими (то есть без временного измерения), так и полными (то есть без пропущенных значений). Однако динамические и неполные данные занимают центральное место во многих передовых приложениях в различных областях исследований, от генетики до робототехники; BNS может играть важную роль во многих из этих параметров, о чем свидетельствуют примеры, упомянутые в разделах 3 и 4. Учитывая их выразительную силу, BNs также включает в себя несколько классических вероятностных моделей и может дополнять их возможностями автоматического рассуждения с помощью различных видов запросов, которые могут выполняться алгоритмически. Было проведено много исследований, направленных на адаптацию BNs к этим приложениям и превращение их в конкурентоспособный выбор для моделирования сложных данных.