

ИНСТРУМЕНТЫ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ЧИСЛЕННЫХ АЛГОРИТМОВ ОБУЧЕНИЯ СТРУКТУРЫ ДИНАМИЧЕСКИХ БАЙЕСОВСКИХ СЕТЕЙ

П. В. Полухин

Воронежский государственный университет

Поступила в редакцию 08.09.2019 г.

Аннотация. Модели динамических байесовских сетей используются для описания процессов, протекающих в условиях риска и неопределенности, случайный характер имеют не только вершины графа сети, но связи между вершинами. Для определения наличия причинно-следственных связей и их направленности применяют специальные экспертные и статистические методы обучения структуры и параметров сети. При использовании экспертных методов построения байесовских сетей структуру графа задает эксперт на основании своего опыта в исследуемой области, и тогда, обучаются только параметры сети, соответствующие условно-вероятностным распределениям вершин сети. Не всегда эксперт может правильно определить причинно-следственные связи между вершинами сети и их направленность. Достаточно эффективными являются формализованные процедуры обучения структуры и параметров сети. Формализованные методы обучения структуры сети включают этап определения зависимости между вершинами сети и этап определения направленности связей. На этапе определения направленности исследование носит локальный характер и подразумевает решение целого ряда оптимизационных задач. Как правило, в качестве алгоритмов обучения динамических байесовских сетей применяются численные оптимизационные алгоритмы. В связи с большой размерностью решаемых задач, эффективность процедур обучения динамических байесовских сетей зависит от эффективности используемых численных алгоритмов. Достаточно часто применяются численные алгоритмы, построенные на основе Ньютоновского подхода. В данной статье описано применение различных инструментов повышения эффективности Ньютоновских алгоритмов для решения задач обучения структуры динамических байесовских сетей. Применение методов Бройдена, Девидона-Флетчера-Пауэлла и Бройдена-Флетчера-Гольдфарба-Шанно позволяет существенно повысить эффективность алгоритмов, а также дает возможность использовать распараллеливание отдельных блоков.

Ключевые слова: динамическая байесовская сеть, алгоритмы обучения, логарифм правдоподобия, вычисление матрицы Гессе, критерий Шварца, критерий Акаике, численные алгоритмы оптимизации.

ВВЕДЕНИЕ

Динамические байесовские сети (ДБС) являются эффективным инструментом моделирования протекающих во времени стохастических процессов в различных отраслях и сферах деятельности. Динамические байесовские сети можно рассматривать как последовательность статических байесовских сетей (срезов), вершины которых соединяются транзитивными связями. Как правило, фор-

мирование связей внутри среза и между срезами ДБС ограничено Марковским покрытием. Марковским покрытием (МП) узла байесовской сети называется множество родителей, детей и других родителей своих детей. Любой узел байесовской сети условно не зависит при известном значении своего Марковского покрытия от всех остальных узлов сети. Данное предположение дает возможность использовать направленный вывод внутри сети, развернутой на n – временных срезов. При работе с ДБС необходимым этапом является процедура обучения сети, цель

которой заключается в получении: структуры (топологии, направленности сети) и таблиц условных вероятностей для каждого из узлов структуры. Классические алгоритмы обучения сети, используемые для статических байесовских сетей, не всегда могут использоваться для динамических сетей в связи с высоким уровнем сложности. Применение классических алгоритмов накладывает ограничения на размер сети и число входящих в нее параметров. Необходимыми становятся гибридные алгоритмы, направленные на снижение сложности классических алгоритмов. Гибридные алгоритмы представляют собой комбинацию методов, реализующих тесты на условную независимость переменных ДБС и методов поиска максимума мер, построенных на основе различных статистических критериев, в частности логарифма правдоподобия, критерия Акаике и Шварца. В данной статье рассматриваются гибридные алгоритмы, реализующие Ньютоновские и Квази-Ньютоновские подходы для определения топологической направленности ДБС и позволяющие использовать стратегию распараллеливания алгоритмов.

МАТЕРИАЛЫ И МЕТОДЫ

Задача обучения структуры ДБС может быть разделена на два основных этапа. Первый этап подразумевает построение направленного графа сети за счет проведения тестов на условную независимость вершин на основе обучающего набора данных. На втором этапе решается последовательность локальных задач, связанных с определением направленности для отдельного узла (узлов). Рассматриваются варианты направленности, для каждого варианта подбираются оптимальные параметры, вычисляется критерий оптимальности и выбирается вариант, максимизирующий данный критерий. В качестве критериев оптимальности, могут выступать различные функции качества. Широко используемыми критериями являются функции, сформированные на основе логарифма правдоподобия, в частности критерии Шварца и Акаике.

Гипотезы, соответствующие байесовской сети G , накладывают ограничения на множество параметров распределения G . В связи с этим, определим специальное параметрическое множество, соответствующее гипотезе G следующим образом: для каждой i переменной пронумеруем множество возможных значений ее родителей $\{1, 2, \dots, q_i\}$, $q_i = \prod_{s \in Pa(X_i)} r_s$ и введем множество $\sum_{i=1}^n r_i q_i$ – мерных векторов [1]:

$$\Theta^G = \left\{ \theta^G = (\theta_{i,j,k}^G) \in \mathbb{R}^{\sum_{i=1}^n r_i q_i} \right\}.$$

В соответствии с гипотезой о факторизации, задаваемой байесовской сетью G , распределение переменных, принадлежащих множеству $X = \{X_1, \dots, X_n\}$, может быть представлено в виде следующего выражения [7]

$$P_\theta(X_1, \dots, X_n) = \prod_{i=1}^n \theta_{i,j,k}^G, \forall \theta \in \Theta_G \quad (1)$$

$$\theta_{i,j,k}^G = P_\theta(X_i = k | P(X_i) = j). \quad (2)$$

Предполагается, что дана база данных $D = \{X^l = (X_1^l, \dots, X_n^l) | l = 1, \dots, N\}$, которую можно представить виде набора N независимых случайных величин, имеющих одинаковое многомерное распределение вероятностей для всех значений параметров из множества θ^G . Вероятности для каждой случайной величины из $D = \{X^l = (X_1^l, \dots, X_n^l) | l = 1, \dots, N\}$ вычисляются по формуле [3]:

$$P_{\theta^G}(X_1^l, X_2^l, \dots, X_n^l) = \prod_{i=1}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} (\theta_{i,j,k}^G)^{1_{i,j,k}} \quad (3)$$

$$1_{i,j,k} = \begin{cases} 1, & X_i = k, Pa(X_i) = j \\ 0, & \text{в противном случае} \end{cases} \quad (4)$$

Учитывая тот факт, что все элементы множества $D = \{X^l = (X_1^l, \dots, X_n^l) | l = 1, \dots, N\}$ являются условно независимыми, вероятностное распределение принимает следующий вид:

$$P_{\theta^G}(D) = \prod_{i=1}^n \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} (\theta_{i,j,k}^G)^{\sum_{l=1}^N 1_{i,j,k}}, \quad (5)$$

где величина $\sum_{l=1}^N 1_{i,j,k}$ – равна числу строк в множестве D , для которых переменная X_i принимает значение k , а ее родители j .

В соответствии с изложенными выше фактами, функция правдоподобия для байесовской сети задается выражением [5]:

$$L(G, \theta^G, D) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{i,j,k} \ln \theta_{i,j,k}^G, \quad (6)$$

где $N_{i,j,k} = \sum_{i=1}^N 1_{i,j,k}$.

Критерии Шварца и Акаике формируются на основе логарифма правдоподобия, однако дополнительно используют функцию штрафов $f(N)$. Обобщенная функция для обоих критериев имеет вид [6]:

$$\varphi(X|\theta) = L(G, \theta^G, D) - \sum_{i=1}^{r_i} (r_i - 1) q_i F(N), \quad (7)$$

где N – общий объем обучающей выборки, M – число параметров ДБС, которое включает параметры транзитивных узлов. Функция штрафов $F(N)$ в выражении (7) будет принимать значение $F(N) = 1$ для критерия Акаике и $F(N) = \frac{1}{2} \log N$ для критерия Шварца соответственно [11].

Оптимизация рассмотренных критериев для нахождения структуры, наилучшим образом отвечающей обучающим данным, опирается на применение численных методов оптимизации. Задача нахождения структуры направленности ДБС как задача численной оптимизации характеризуется большой размерностью и высокой сложностью. Решение подобных задач требует применения инструментов повышения эффективности классических методов наискорейшего спуска и Ньютона. В основе градиентных алгоритмов используется разложение функции $\varphi(Z)$ в ряд Тейлора в окрестности точки Z в направлении δ . Данное разложение имеет следующий вид

$$\begin{aligned} \varphi(Z + \delta) &= \varphi(Z) + \delta \times [g(Z)]^T + \mu \\ \mu &= \frac{1}{2} \delta^T \delta \times H(Z) + O, \end{aligned} \quad (8)$$

где O – локальная погрешность, $g(Z)$ – градиент функции $\varphi(Z)$ и $H(Z)$ – матрица Гессе, которые определяются в соответствие со следующими выражениями [2]

$$g(Z) = \nabla \varphi = \left[\frac{d\varphi}{dZ_1}, \frac{d\varphi}{dZ_2}, \frac{d\varphi}{dZ_3}, \dots, \frac{d\varphi}{dZ_n} \right]^T$$

$$H(Z) = \begin{bmatrix} \frac{d^2\varphi}{dZ_1 dZ_1} & \dots & \frac{d^2\varphi}{dZ_m dZ_1} \\ \frac{d^2\varphi}{dZ_2 dZ_2} & \dots & \frac{d^2\varphi}{dZ_m dZ_2} \\ \dots & \dots & \dots \\ \frac{d^2\varphi}{dZ_1 dZ_n} & \dots & \frac{d^2\varphi}{dZ_m dZ_n} \end{bmatrix} \quad (9)$$

В уравнении (8) основную роль играет выбор направления δ , зависящего от текущей точки Z . Данное направление выбирается таким образом, чтобы для каждого следующего шага выполнялось условие $\varphi(Z_{k+1}) < \varphi(Z_k)$.

В методе наискорейшего спуска в качестве направления используется градиент $\delta = g(Z) = \nabla \varphi(Z): Z_{k+1} = Z_k + \beta_k \nabla \varphi(Z_k)$. Критерием останова является условие: $\nabla \varphi(Z) < \varepsilon$. При большой размерности задачи алгоритмы наискорейшего спуска работают достаточно медленно, особенно на финальной части работы. В методе Ньютона, также широко используемом при решении задач безусловной оптимизации, в качестве направления движения используется $\delta = H^{-1} \nabla \varphi(Z)$. Метод Ньютона имеет хорошую скорость сходимости, в частности на финальной части работы, однако является слишком ресурсоемким, поскольку на каждой итерации метода нужно: вычислять матрицу Гессе H ; решать систему линейных уравнений $H\delta = \nabla \varphi(Z)$ (сложность алгоритма $O(n^3)$). Метод Ньютона невозможно применять для многих задач большой размерности. Необходимы методы, которые сочетали-бы высокую скорость сходимости метода Ньютона с менее ресурсоемкими итерациями.

Для решения системы уравнений с положительно определенной матрицей можно применять метод Холесского, который основывается на представлении матрицы H в виде следующего произведения [8]

$$H = LL^T, (Hx, x) \geq 0, \quad (10)$$

где L – нижняя треугольная матрица, состоящая из положительных элементов на диагонали и являющаяся коэффициентом Холесского для матрицы H . Процедуру вычисления значений для элементов матрицы H

можно описать на основе следующего выражения

$$H_{ij} = \sum_{k=1}^n L_{ik} L_{kj}^T, (Hx, x) \geq 0. \quad (11)$$

С учетом того, что матрица H является симметричной и положительно определенной, достаточно рассмотреть случай, где $i \leq j$. Исходя из этого выражение (6) может быть преобразовано к следующему виду

$$H_{ij} = \sum_{k=1}^{i-1} L_{ik} L_{kj}^T + L_{ii} L_{ij}. \quad (12)$$

Из формулы можно получить значения всех элементов матрицы L на основе следующего выражения

$$L_{ii} = \sqrt{H_{ii} - \sum_{k=1}^{i-1} L_{ik}^2}, i = j$$

$$L_{ji} = \frac{(H_{ii} - \sum_{k=1}^{i-1} L_{ik} L_{jk})}{L_{ii}}, i < j \quad (13)$$

Алгоритм вычисления значений симметричной, матрицы L размерностью $n \times n$ с использованием разложения Холецкого можно представить в виде совокупности следующих равенств

$$H_{i,k} = H_{i,k} - L_{i,k} L_{jk},$$

$$L_{i,j} = \sqrt{H_{jj}}, i = j \quad (14)$$

$$L_{i,j} = \frac{H_{k,j}}{L_{j,j}}, i \neq j$$

где $j = 1, 2, \dots, n$; $i = 1, 2, \dots, j$; $k = 1, 2, \dots, j-1$ для нижнего треугольного разложения и $j = m, m-1, \dots, 1$; $i = j, j-1, \dots, 1$; $k = j+1, \dots, m$.

Рассмотрим алгоритм Гаусса-Ньютона. Выбор направления δ для каждого Z в данном алгоритме осуществляется путем поиска минимума оценочной функции. В частности, минимум оценочной функции может быть получен методом наименьших квадратов:

$$f(w) = \min F(w)^2 = \sum_{k=1}^n (g_k(w) - a_k)^2 \quad (15)$$

где $g_k(w)$ – градиент, $F(x) = [g_k(w) - a_k]_{k=1}^n$ представляет собой разностное выражение для градиента $g_k(w)$.

Для нахождения минимума функции $f(w)$, необходимо определить Якобиан $J(w)$. Якобиан представляет собой производную от

непрерывной дифференцируемой функции $F: R^n \rightarrow R^m$ в точке w , а транспонированную матрицу для $J(w)$ называют градиентом F в точке w [2]

$$F'(w) \in R^{m \times n}, F'(w)_{ij} = \frac{dF_i}{dw_j}(w) \quad (16)$$

$$F'(w) = J(w) = \nabla F(w)^T$$

С учетом введенных обозначений, матрица Якоби может быть представлена в следующем виде [4]:

$$J(w) = \begin{pmatrix} \frac{dg_1(w)}{dw_1} & \dots & \frac{dg_1(w)}{dw_n} \\ \frac{dg_2(w)}{dw_1} & \dots & \frac{dg_2(w)}{dw_n} \\ \dots & \dots & \dots \\ \frac{dg_m(w)}{dw_1} & \dots & \frac{dg_m(w)}{dw_n} \end{pmatrix}, w \in R^m \quad (17)$$

Последовательные приближения для некоторого шага w_{j+1} , при заданном начальном значении w_0 можно определить следующим образом:

$$w_{j+1} = w_j - \frac{J^T(w_j)F(w_j)}{J^T(w_j)J(w_j)}. \quad (18)$$

Выражение для w_{j+1} может быть преобразовано к виду:

$$w_{j+1} = w_j - \frac{g[f](w')}{H(w)}, \quad (19)$$

где $H(w)$ – матрица Гессе, $f(w): R^m \rightarrow R$.

Применение метода наименьших квадратов для минимизации функции $f(w)$ позволяет переопределить значение градиента и матрицы Гессе [12]

$$g[f](w) = J^T F(w), \quad (20)$$

$$H(w) = J^T(w)J(w) + Q(w).$$

Приближение на шаге $j+1$ описывается следующим выражением:

$$w_{j+1} = w_j - \Omega \times J^T(w)F(w)$$

$$\Omega = \left[J^T(w)J(w) + \sum_{i=1}^m F(w)H_i(w) \right]^{-1}. \quad (21)$$

Основная идея, заложенная в алгоритме Гаусса-Ньютона, заключается в выполнении неравенства

$$[J^T(w)J(w)] > Q(w). \quad (22)$$

Если значение $Q(w) \rightarrow 0$, то им можно пренебречь. На основе этого предположения конечную формулу обновления на основе подхода Гаусса-Ньютона можно записать в следующем виде

$$w_{j+1} = w_j - [J^T(w)J(w)]^{-1} J^T(w)F(w). \quad (23)$$

Между тем алгоритм Гаусса-Ньютона обладает существенным недостатком в случаях, когда $Q(w)$ может принимать достаточно большие значения, что ведет к проблемам со сходимостью и снижению быстродействия алгоритма.

В процессе реализации Ньютоновских подходов предполагается перерасчет матрицы Гессе для каждой итерации алгоритма. Для оптимизации процедуры расчета матрицы Гессе $H(Z)$ наиболее применимы квазиньютоновские методы, в частности алгоритм Бройдена – Флетчера – Гольдфарба – Шанно (БФГШ). Алгоритм БФГШ является итерационным методом и использует разложение оценочной функции в полином для состояния Z_k

$$m_k(\lambda) = \alpha + \frac{1}{2} \lambda^T H_k(Z_k) \lambda. \quad (24)$$

В алгоритме БФГШ используется приближенная матрица Гессе $B(Z_k)$. Прямое вычисление Гессиана требуется лишь для первой итерации алгоритма. Значения λ могут быть определены на основе следующего выражения

$$\lambda_k = -B^{-1}(Z_k) \nabla \varphi(Z_k). \quad (25)$$

Для получения приближенного значения матрицы Гессе $B(Z_k)$ необходимо определить условие Вольфе. Данное условие определяется в виде неравенств для оценочной функции и градиента оценочной функции для каждой итерации Z_{k+1} :

$$\varphi(Z_k + \alpha_k \lambda_k) \leq \varphi(z_k) + c_1 \alpha_k \nabla \varphi_k^T(Z_k) \lambda_k \quad (26)$$

$$\nabla \varphi(Z_k + \alpha_k \lambda_k)^T \geq c_2 \nabla \varphi_k^T(Z_k) \lambda_k,$$

где c_1 и c_2 – константы, удовлетворяющие условию $0 < c_1 < c_2$.

Константы c_1 и c_2 выбираются из достаточно ограниченной окрестности, что позволяет получить наименьшее значение градиента или изменить его знак.

Исходя из неравенства (26) можно определить строгое условие Вольфе [9]. Данное условие имеет место, если приближение оценочной функции будет в пределах локального минимума.

$$\begin{aligned} \varphi(Z_k + \alpha_k \lambda_k) &\leq \varphi(Z_k) + c_1 \alpha_k \nabla \varphi_k^T(Z_k) \lambda_k \\ |\nabla \varphi(Z_k + \alpha_k \lambda_k)^T| &\leq c_2 |\nabla \varphi_k^T(Z_k) \lambda_k| \end{aligned} \quad (27)$$

Как правило, выбор констант c_1 и c_2 для формул (26) и (27) осуществляется из небольшой окрестности нуля – это дает возможность увеличить (уменьшить) значение градиента или изменить его направление. Для алгоритма БФГШ мы будем использовать нестрогое условие Вольфе (26), так как нам важен факт исключения попадания функции в локальный минимум. Значение Z_{k+n} для итерации $k+n$ в соответствии с нестрогим критерием Вольфе определяется как:

$$Z_{k+n} = Z_k + \alpha_k \lambda_k, \quad (28)$$

где α_k – длина шага итерации, выбираемая в соответствии с нестрогим критерием Вольфе

Для получения искомого выражения для приближенной матрицы $B(X_k)$ в соответствии с методом БФГШ заменим исходную матрицу Гессе на ее приближенный эквивалент.

$$m_{k+n}(\lambda) = \alpha + \frac{1}{2} \lambda^T B_{Z_{k+n}} \lambda \quad (29)$$

$$\alpha = \varphi(z_{k+n}) + \nabla \varphi^T(z_{k+n}) \lambda$$

Приближенные значения матрицы Гессе $B_{Z_{k+n}}$ для каждого следующего шага $k+n$ будет формироваться с учетом матрицы $B_{Z_{k+n-1}}$ для предыдущей итерации. При этом за начальное приближение матрицы Гессе B_{X_0} принято брать единичную матрицу $B_{Z_0} = I$. Рассмотрим процедуру получения приближенной матрицы $B_{Z_{k+n}}$. Для этого необходимо выполняться ряд условий. Первое условие накладывает ограничение на то, что для получения $B_{Z_{k+n}}$ необходимо, чтобы на итерациях, предшествующих состоянию $k+n$ значение $m_{k+n}(\lambda)$ совпадало со значением градиента как минимум для двух связанных состояниях $\nabla \varphi(Z_{k+n})$ и $\nabla \varphi(Z_{k+n-1})$. Второе условие накладывается в соответствии с нестрогим условием Вольфе. Исходя из этого формулу для расчета $m_{k+n}(\lambda)$ можно записать в следующем виде

$$\nabla m_{k+n}(-\alpha_{k+n-1}\lambda_{k+n-1}) = \nabla \varphi(x_{k+n}). \quad (30)$$

Выражение (30) можно представить через разность градиентов двух смежных состояний Z_{k+n-1} и Z_{k+n}

$$B_{Z_{k+n}} \alpha_{k+n-1} \lambda_{k+n-1} = \nabla \varphi(Z_{k+n}) - \nabla \varphi(Z_{k+n-1}). \quad (31)$$

Выражение (31) является уравнением секущих. Обозначим разность значений $s_{k+n-1} = Z_{k+n} - Z_{k+n-1}$ и разность градиентов $y_{k+n-1} = \nabla \varphi(Z_{k+n}) - \nabla \varphi(Z_{k+n-1})$. В таком случае можно преобразовать уравнение секущих:

$$B_{Z_{k+n}} s_{k+n-1} = y_{k+n-1}, \quad (32)$$

Для подтверждения условия того, что градиент функции будет изменять свое значение при осуществлении сдвига s_{k+n-1} относительно точек Z_{k+n-1} и Z_{k+n} необходимо, чтобы оценочная функция $\varphi(Z_{k+n})$, была выпуклой.

В процессе решения задачи нахождения приближенной матрицы $B_{Z_{k+n}}$ мы предполагаем, что для каждого шага процесса формирования матрицы $B_{Z_{k+n}}$ выполняется условие секущих (32). Сформулируем задачу следующим образом:

$$\min_B |B_{Z_{k+n}} - B_{Z_{k+n-1}}|. \quad (33)$$

Алгоритм БФГШ опирается на алгоритм Девидона – Флетчера – Пауэлла (ДФП). В соответствии с введенными обозначениями можно получить выражения для вычисления $B_{Z_{k+n}}$, предложенное в алгоритме ДФП:

$$B_{Z_{k+n}} = (I - \delta_{k+n-1} s_{k+n-1} y_{k+n-1}^T) \times \beta \quad (34)$$

$$\beta = B_{Z_{k+n-1}} (I - \delta_{k+n-1} y_{k+n-1} s_{k+n-1}^T) + \delta_{k+n-1} s',$$

где $\delta_{k+n-1} = 1 / y_{k+n-1} s_{k+n-1}^T$, $s' = s_{k+n-1} s_{k+n-1}^T$.

Вводится обратная матрица $H_{Z_{k+n}} = B_{Z_{k+n}}^{-1}$. В данном случае можно получить выражение для расчета приближенной обратной матрицы Гессе [10] для некоторого следующего шага $k+n$

$$H_{k+n} = H_{k+n-1} - \frac{H'}{y_{k+n-1}^T H_{k+n-1} y_{k+n-1}} + \sigma \quad (35)$$

$$\sigma = \frac{s'}{y_{k+n-1}^T s_{k+n-1}}$$

$$H' = H_{k+n-1} y_{k+n-1} y_{k+n-1}^T H_{k+n-1}$$

Основная идея алгоритма БФГШ базируется на использовании выражения для расчета обратной матрицы $B_{x_{k+n}}^{-1}$. В свою очередь,

на матрицу $B_{x_{k+n}}^{-1}$ накладывается ряд ограничений, связанных с тем, что она должна быть симметричной положительно определенной, а также удовлетворять уравнению секущих. Тогда выражение (34) может быть переписано в соответствии с методом БФГШ

$$H_{Z_{k+n}} = I' \times H_{Z_{k+n-1}} I'' + \xi_{k+n-1} s' \quad (36)$$

$$I' = (I - \xi_{k+n-1} s_{k+n-1} y_{k+n-1}^T)$$

$$I'' = (I - \xi_{k+n-1} s_{k+n-1} y_{k+n-1}^T),$$

где $\xi_{k+n-1} = 1 / y_{k+n-1} s_{k+n-1}^T$.

Как правило, в процессе вычисления обратных матриц $H_{Z_{k+n}}$ возникает ряд ограничений, связанных с необходимостью хранения матрицы с предшествующей итерации $H_{Z_{k+n}}$. В таком случае для снижения ресурсных ограничений алгоритма БФГШ можно использовать модификацию алгоритма БФГШ с ограниченной памятью. Тогда история для $n+k+l$ итераций будет сохраняться как $\rho_{n+k-1} = ((s_{k+n-i}, y_{k+n-i}))_{i=1}^{n+k+l}$. При использовании алгоритма БФГШ с ограниченной памятью возможно неточное восстановление матрицы H_{k+n} из ρ_{n+k-1} . Однако в некоторых ситуациях, когда ресурсы вычислительной системы ограничены, а требуется получить первичное приближения для матрицы Гессе, то для этого достаточно использование алгоритма БФГШ с ограниченной памятью. В рамках исследования предполагается применение классического алгоритма БФГШ.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

С практической точки зрения, рассматривая метод БФГШ применительно к оценочным функциям на основе логарифма правдоподобия, критериев Шварца и Акаике, можно сделать вывод, значение матрицы $H_{Z_{k+n}}$ не всегда будет сходиться к истинному значению матрицы Гессе. Данный факт связан с тем, что не всегда можно определить соответствие с $H_{Z_{k+n}}$ после выполнения $n+k$ шагов алгоритма, особенно при использовании схемы метода с ограниченной памятью. Алгоритм БФГШ хорошо адаптируется для поиска экстремумов оценочных функции в процессе определения направленности связей между

Сравнение производительности алгоритмов

№ ДБС	Название ДБС	Алгоритм восхождения	Алгоритм Гаусса – Ньютона	Алгоритм наискорейшего спуска	Алгоритм БФГШ
1.	Alarm	120 сек	100 сек.	110 сек.	80 сек.
2.	Water	600 сек.	700 сек.	800 сек.	300 сек.

узлами ДБС, а использование приближенной матрицы Гессе H_{k+n} позволяет применять алгоритм для ДБС со сложной топологией. Для получения практических результатов, было разработано программное обеспечение для описанных гибридных алгоритмов обучения структуры ДБС. Вычислительный эксперимент проводился на примере ДБС Alarm (37 узлов, 46 дуг, 509 параметров), Water (32 узла, 66 дуг, 10082 параметра).

Приведем сравнительную таблицу производительности рассмотренных алгоритмов. В табл. 1 представлены временные затраты, необходимые для обучения структуры ДБС Alarm и Water, состоящие из трех временных срезов.

Из анализа табл. 1 видно, что в процессе обучения сети Alarm, прирост эффективности алгоритма БФГШ не имеет ярко выраженного представления, так как сеть содержит всего 509 параметров, в то время как для обучения ДБС Water явно наблюдается прирост производительности относительно классических алгоритмов поиска. В результате проведенного исследования получено, что в классических алгоритмах доля точек попавших в локальный минимум при большом числе параметров ДБС достигает порядка 15–20 % от всего объема обучающей выборки. Это ведет к неправильному определению направления между некоторыми узлами ДБС и, в частности, узлами с транзитивными связями. Таким образом, применение такого подхода усложняет дальнейшее использование стохастических алгоритмов вероятностного вывода на основе цепи Маркова, так как точно не известно корректное направление связи между узлами ДБС в двух соседних временных срезах. Применение алгоритма БФГШ позволяет существенно минимизировать риски неправильного определения топологических связей и направленности внутри ДБС.

ЗАКЛЮЧЕНИЕ

Применение численных методов на основе квазиньютоновского подхода и, в частности, метода БФГШ в процессе обучения ДБС дает возможность оптимизировать процедуру определения направленности сети. Использование нестроого критерия Вольфе позволяет исключить попадание точки в локальный экстремум и оптимизировать алгоритм Бройдена – Флетчера – Гольдфарба – Шанно. В результате проведенного исследования: предложены инструменты повышения эффективности классических алгоритмов обучения структуры ДБС; проведен вычислительный эксперимент и анализ результатов применения рассмотренных инструментальных средств. Также проанализированы структуры ДБС, полученных экспертным методом и на основе применения разработанного алгоритма. В результате выполнения алгоритма получена более адекватная структура по отношению к экспертной. Связи некоторых узлов в экспертном варианте были определены некорректно, на этапе обучения параметров в результате применения рассмотренных инструментальных средств получены более правдоподобные распределения для таблиц условных вероятностей для каждого из узлов ДБС, включая узлы, имеющие транзитивные связи.

Для проведения вычислительно эксперимента было разработано программное обеспечение, реализующих представленные алгоритмы. Программное обеспечение может быть использовано для обучения структуры ДБС для различных областей моделирования и позволяет осуществлять подстройку сети под решения конкретной научно-практической задачи.

СПИСОК ЛИТЕРАТУРЫ

1. Азарнова, Т. В. Расширение функциональных возможностей фаззинга веб-приложений на основе динамических сетей Байеса / Т. В. Азарнова, П. В. Полухин // Научно-техническая информация. Серия 2. Информ. процессы и системы. – 2014. – № 9. – С. 12–19.
2. Вержбицкий, В. М. Численные методы (линейная алгебра и нелинейные уравнения) / В. М. Вержбицкий. – М. : Оникс 21 век, 2005. – 432 с.
3. Грот, М. Д. Оптимальные статистические решения / М. Д. Грот. – М. : Мир, 1974. – 491 с.
4. Магнус, Я. Р. Матричное дифференциальное исчисление с приложениями к статистике и эконометрике / Я. Р. Магнус, Х. Нейдеккер. – М. : Физматлит., 2002. – 496 с.
5. Фишер Р. А. Статистические методы для исследователей / Р. А. Фишер. – М. : Гостатизд, 1958. – 267 с.
6. Рао, С. Р. Линейные статистические методы и их применение / С. Р. Рао – М. : Наука, 1968. – 548 с.
7. Leman, E. L. Selected works by E. L. Leman / E. L. Leman. – New York : Springer, 2012. – 1109 p.
8. Korb, K. B. Bayesian Artificial Intelligence / K. B. Korb, A. E. Nicholson. – Boca Raton : CRC Press, 2004. – 491 p.
9. Турчак, Л. И. Основы численных методов / Л. И. Турчак, П. В. Плотников. – М. : Физматлит, 2003. – 304 с.
10. Wolfe, P. Convergence conditions for ascent methods / P. Wolfe // Siam Review, 1969. – Vol. 11, №. 2. – P. 226–235.
11. Nocedal, J. Numerical optimization / J. Nocedal, S.J. Wright. – N.Y. : Springer, 1999. – 634 p.
12. Кендалл, М. Статистические выводы и связи / М. Кендалл, А. Стюарт. – М. : Наука, 1973. – 878 с.
13. Турчак, Л. И. Основы численных методов / Л. И. Турчак, П. В. Плотников. – М. : Физматлит, 2003. – 304 с.

Полухин Павел Валерьевич – кандидат техн. наук, кафедра математических методов исследования операций факультета прикладной математики, информатики и механики Воронежского государственного университета. E-mail: alfa_force@bk.ru

П. В. Полухин

APPLIANCE FOR IMPROVING THE EFFICIENCY OF NUMERICAL METHODS OF STRUCTURE LEARNING ALGORITHMS FOR DYNAMIC BAYESIAN NETWORKS

P. V. Polukhin

Voronezh State University

Annotation. Dynamic Bayesian network models are used to describe processes in the conditions of risk and uncertainty, the random nature have not only the vertices of the network, but communication between nodes. To determine the presence of causal relationships and their orientation, special expert and statistical methods of the network structure and learning are used. When using expert methods for constructing Bayesian networks, the graph structure is set by the expert on the basis of his experience in the research area, and then only the network parameters corresponding to the conditional probability distributions of the network vertices are trained. It is not always possible for the expert to correctly determine causal relationships between the vertices of the network and their direction. Formalized procedures for the network structure and parameters learning are quite effective. Formalized methods of learning network structure include the stage of determining the relationship between the vertices of the network and the stage of determining the direction of relations. The stage of determining the direction of learning is local in nature and involves the solution of a number of optimization problems. As a rule, numerical optimization algorithms are used as learning algorithms for dynamic Bayesian networks. Due to the large dimension of solving problems, the efficiency of dynamic Bayesian network learning procedures depends on the efficiency of the numerical algorithms. Numerical algorithms based on the Newtonian approach are often used. This article describes the use of various tools to improve the efficiency of Newtonian algorithms for solving problems of learning the structure of dynamic Bayesian networks. The use of the methods of Broyden, Davidon-Fletcher-Powell and Broyden-Fletcher-Goldfarb-Shanno can significantly improve the efficiency of the algorithms, as well as makes it possible to use the parallelization of individual blocks.

Keywords: dynamic Bayesian network, learning algorithms, likelihood logarithm, Hessian matrix calculation, Schwartz criterion, Akaike criterion, numerical optimization algorithms.

Polukhin P. V. – PhD in Technical Sciences, Department of Mathematical Methods Operations Research, Faculty of Applied mathematics, Informatics and mechanics, Voronezh State University. E-mail: alfa_force@bk.ru