

АНАЛИЗ ЭКВИВАЛЕНТНОСТИ БАЙЕСОВСКИХ СЕТЕЙ НА ОСНОВЕ АСИМПТОТИЧЕСКИХ ОЦЕНОК БАЙЕСА — ДИРИХЛЕ

© 2022 П. В. Полухин✉

*Воронежский государственный университет
Университетская пл., 1, 394018 Воронеж, Российская Федерация*

Аннотация. Динамические и статические байесовские сети являются эффективным инструментом моделирования стохастических процессов. Области практического внедрения данных моделей существенно расширились за последнее время. Качество их применения при решении практических задач во многом определяется возможностями алгоритмов обучения структуры и вероятностных параметров моделей, позволяющих произвести настройку сети для решения рассматриваемого круга прикладных задач. В алгоритмах определения оптимальной структуры байесовских сетей и верификация алгоритмов настройки их параметров важную роль играют инструменты, основанные на принципах эквивалентности. На базе принципа эквивалентности формируются асимптотические оценки преобразований, получаемых в процессе добавления, изменения или удаления отдельных узлов графа байесовской сети и создается аппарат получения локального априорного распределения для каждого из параметров сети. В данной работе исследуются инструменты оценивания эквивалентности байесовских сетей на основе метрики Байеса — Дирихле, структурного расстояния Хэмминга и Кульбака — Лейблера. Данные инструменты можно применить и к динамическим байесовским сетям, для работы с которыми дополнительно нужно определить структуру модели перехода между временными срезами. В рамках исследования рассматриваются также вопросы эквивалентности априорных распределений вероятностей, формируемых в процессе обучения параметров. В заключительной части работы приведен вычислительный эксперимент, отражающий эффективность применения различных алгоритмов обучения с позиции сравнения их результатов с эквивалентными эталонными байесовскими сетями. Предложенные в работе инструментальные средства позволяют адаптировать статические и динамические байесовские модели для решения практических задач, оптимизируя процессы обучения данных моделей за счет использования принципа эквивалентности графических вероятностных моделей.

Ключевые слова: эквивалентность байесовских сетей, асимптотические оценки, эквивалентная оценка Байеса — Дирихле, структурное расстояние Хэмминга.

ВВЕДЕНИЕ

Актуальным направлением исследований в области развития теории и алгоритмической базы аппарата байесовских сетей, является разработка инструментов повышения эффективности и качества обучения байесовских сетей. Процессы обучения включают обучение структуры и параметров графиче-

ской вероятностной модели. При обучении структуры вначале строится ненаправленный граф, а затем осуществляется процесс формирования направленности связей на основе решения оценочных задач. Одной из эффективных метрик, существенно используемых при обучении структуры и направленности связей является эквивалентная метрика Байеса — Дирихле (БДЭ), которая впервые приведена в работах Хекермана. Метрика Байеса — Дирихле используется для формирования локальных оценок каждой вершины,

✉ Полухин Павел Валерьевич
e-mail: alfa_force@bk.ru



Контент доступен под лицензией Creative Commons Attribution 4.0 License.

The content is available under Creative Commons Attribution 4.0 License.

максимизация данной метрики направлена на получение оптимальной структуры байесовской сети за счет изменения топологических связей. Формирование структуры осуществляется на основе обучающей выборки, оценивается правдоподобие структуры и эквивалентность байесовских сетей. Для оценки эквивалентности используется вычисление структурного расстояния Хэмминга. Структурное расстояние Хэмминга определяет разницу числа ребер, а также направленность ребер для каждой их сетей. Исследования по оценке эквивалентности байесовских сетей в общем виде описаны в работах [1], однако затрагивают лишь применение изложенного подхода исключительно к статическим байесовским сетям. В рамках данной работы исследуются преимущества асимптотических оценок для определения оптимальной структуры байесовской сети и рассматриваются оценки эквивалентности байесовских сетей на основе вычисления энтропии распределения параметров сети.

МАТЕРИАЛЫ И МЕТОДЫ

Байесовская сеть (БС) $B = \{X, G\}$ представляет собой вероятностную модель с множеством $X = \{X_1, X_2, \dots, X_n\}$ переменных, являющихся вершинами направленного ациклического графа (НАГ) и имеющих совместное распределение $P(X)$ с выделенным марковским покрытием. Марковское покрытие является мерой условной независимости вершины X при условии, что заданы ее родительские вершины Y , от других вершин сети. В предположении условной независимости априорное распределение вероятностей для всех переменных X , соответствующих байесовской сети B , имеет вид:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^N P(x_i | Y_i), \quad (1)$$

где N — общее число переменных байесовской сети, $Y = Parents(X_i)$ — множество родительских вершин, соответствующих переменной X_i .

В процессе работы с байесовской сетью при решении конкретной задачи или класса

задач важным этапом является этап обучения сети, который функционально делится на обучение структуры и параметров. Обобщенное решение задачи обучения можно сформулировать в виде следующего выражения [2]:

$$P(B|G) = P(G, X|D) = P(G|D)P(X|G, D), \quad (2)$$

где $P(G|D)$ и $P(X, G, D)$ — распределения вероятностей, соответствующие процедуре обучения структуры и параметров, D — обучающая выборка для переменных X .

Для обучения структуры БС рассмотрим применение оценочных алгоритмов, в основу которых заложен принцип, что каждая операция добавления, удаления или изменение направленности дуги графа БС будет приводить к изменению специальной оценки $S(G, D)$, причем общая оценка БС может быть представлена в виде суммы условных (в соответствии с родительскими вершинами) оценок для каждой из вершин:

$$S(G, D) = \sum_{i=1}^n s(x_i | Y_i). \quad (3)$$

Критерий эквивалентности двух графов G и G' представим в виде условия $S(G, D) = S(G', D)$. В качестве оценки $P(G|D) = S(G, D)$ можно рассмотреть метрику Байеса — Дирихле. В соответствии с введенной метрикой, задачу поиска оптимальной структуры БС можно представить в виде задачи поиска сети, отвечающей максимуму апостериорной вероятности (МАВ). В качестве оценочной функции для вычисления $\operatorname{argmax} P(G|D)$ в соответствии с МАВ рассмотрим эквивалентный критерий Байеса — Дирихле [3]. Предполагается, для каждой переменной X при фиксированном числе родительских вершин Y задано распределение Дирихле:

$$P(\theta|G) = \frac{\Gamma\left(\sum_{k=1}^r \alpha_k\right)}{\prod_{k=1}^r \Gamma(\alpha_k)} \prod_{k=1}^r (\theta_k)^{\alpha_k - 1}. \quad (4)$$

где α_k — экспоненциальный параметра распределения Дирихле, r — число возможных состояний для X_k , θ — вектор параметров БС, соответствующий переменным X , $\Gamma(\cdot)$ — гамма-функция, такая что

$$\frac{\Gamma(n + \alpha)}{\Gamma(\alpha)} = (n - 1 + \alpha)(n - 2 + \alpha) \times \dots \times \alpha, \quad (5)$$

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \Gamma(1) = 1.$$

Математическое ожидание для параметров θ в соответствии с распределением Дирихле (4) имеет вид:

$$\mathbb{E}(\theta) = \frac{\alpha_k}{\sum_{i=1}^r \alpha_{i,k}}. \quad (6)$$

Для вычисления МАВ определим распределения вероятностей, соответствующие выполнению условий глобальной и локальной независимости для каждой из переменных X_k байесовской сети:

$$P(\theta|G) = \prod_{k=1}^n P(\theta_k | G),$$

$$P(\theta_i|G) = \prod_{j=1}^q P(\theta_{k,j} | G), \quad (7)$$

$$q = \prod_{x_j \in Y} r_j.$$

Перепишем выражение (4) с учетом условий глобальной и локальной независимости (7):

$$P(\theta|G) = \prod_{i=1}^n \prod_{j=1}^q \frac{\Gamma(\sum_{k=1}^r \alpha_{i,j,k})}{\prod_{k=1}^r \Gamma(\alpha_{i,j,k})} \pi(\theta_{i,j,k}), \quad (8)$$

$$\pi(\theta_{i,j,k}) = \prod (\theta_{i,j,k})^{\alpha_{i,j,k} - 1}.$$

В соответствии с полученным выражением (8) определим распределение структуры БС [4]

$$P(D | G) = \pi \prod_{i=1}^n \prod_{j=1}^q \frac{\Gamma(\sum_{k=1}^r a_{i,j,k})}{\Gamma(\sum_{k=1}^r N_{i,j,k} + a_{i,j,k})}, \quad (9)$$

$$\pi = \prod_{k=1}^r \frac{\Gamma(N_{i,j,k} + a_{i,j,k})}{\Gamma(a_{i,j,k})},$$

где N_{ijk} — число случаев в k -й выборке, когда переменная $X = i$, а родительские переменные принимают значение $Y = j$, $N_{i,j}$ — число случаев для всей совокупности выборок D , в которых переменная $X = i$, а родительские переменные принимают значение $Y = j$.

Логарифмируя выражение (9), получим метрику Байеса — Дирихле (БД):

$$BD = \ln P(\theta|G) = \sum_{i=1}^n \sum_{j=1}^q \left(\omega_1 + \sum_{k=1}^r \omega_2 \right),$$

$$\omega_1 = \ln \frac{\Gamma(\sum_{k=1}^r \alpha_{i,j,k})}{\Gamma(\sum_{k=1}^r (N_{i,j,k} + \alpha_{i,j,k}))}, \quad (10)$$

$$\omega_2 = \ln \frac{\Gamma(N_{i,j,k} + \alpha_{i,j,k})}{\Gamma(\alpha_{i,j,k})}.$$

Тогда из формулы (11) получим оценку максимума апостериорной вероятности

$$\hat{G} = \arg \max_G P(\theta|G). \quad (11)$$

Для определения эквивалентной метрики Байеса — Дирихле, введем понятие эквивалентности байесовских сетей, сформулированной Перлом и Верма [5, 6]. Две байесовские сети будут считаться эквивалентными, если им соответствует одинаковая совокупность высказываний о разделении. Эквивалентные байесовские сети имеют одинаковый набор высказываний об условной независимости, и следовательно, одинаковые ограничения на параметры распределения. Справедливо также утверждение о том, что две сети эквивалентны тогда и только тогда, когда они имеют одинаковый скелет и V-структуры. Эквивалентная метрика Байеса — Дирихле (БДЭ) будет соответствовать БД-метрике, но только при оценке графов имеющих эквивалентный набор классов. В таком случае мера БДЭ позволяет преобразовать распределение $P(X | G)$ к $P(X | Y, G)$. Распределение $P(X | Y, G)$ будет формироваться из классов эквивалентности, соответствующих вершинам X и множеству их родителей Y . В качестве процедуры оценки эквивалентности графов G и G' , формируемых в соответствии с классами эквивалентности будем использовать условие эквивалентности правдоподобия (ЭП). Оно заключается в том: если существует два графа G и G' , такие что $P(G) > 0$ и $P(G') > 0$ эквивалентны, то справедливо следующее равенство [7]:

$$P(\theta|G) = P(\theta|G'). \quad (12)$$

С учетом условия ЭП, экспоненциальный параметр распределения можно выразить в следующем виде:

$$\alpha_{i,j,k} = N'P(X_i = k, Y_j = j | G), \quad (13)$$

где N' — размер эквивалентной выборки, соответствующий распределению $P(\theta | G)$.

С учетом условия ЭП и выражения (13) получим метрику БДЭ за счет упрощения выражения (11):

$$BDe = \ln P(\theta | G) = \sum_{i=1}^n \sum_{j=1}^q \left(\gamma_1 + \sum_{k=1}^r \gamma_2 \right),$$

$$\gamma_1 = \ln \frac{\Gamma\left(\frac{N'}{q}\right)}{\Gamma\left(N_{ij} + \frac{N'}{q}\right)}, \quad (14)$$

$$\gamma_2 = \ln \frac{\Gamma\left(N_{i,j,k} + \frac{N'}{rq}\right)}{\Gamma\left(\frac{N'}{rq}\right)},$$

где N' — размер эквивалентной выборки

Для оценки метрики БДЭ рассмотрим неравенство, характеризующее взаимную информацию между X_i , Y_i и Z , условную независимость X_i и Y_i при наличии множества переменных Z :

$$BDe(X \cup Z) BDe(Y \cup Z) \geq BDe(Z), \quad (15)$$

$$BDe^{x,y,z} = BDe(X \cup Y \cup Z).$$

Взаимную информацию $I(X, Y)$ между переменными X и Y определим в следующем формализованном виде:

$$I(X, Y) = \sum_x \sum_y P_{X,Y}(x, y) \ln \rho(x, y),$$

$$\rho(x, y) = \frac{P_{X,Y}(x, y)}{P_X(x) P_Y(y)}. \quad (16)$$

а условную взаимную информацию X и Y при наличии Z :

$$I(X, Y) = \sum_x \sum_y \sum_z P_{X,Y,Z}(x, y, z) \ln \rho_{x,y,z},$$

$$\rho_{x,y,z} = \frac{P_{X,Y,Z}(x, y, z) P_Z(z)}{P_{X,Y}(x, z) P_{Y,Z}(y, z)}. \quad (17)$$

Из выражения (17) следует, что метрика Байеса — Дирихле \hat{G} будет является эквивалентной, если она принимает одно и тоже значение для графов, имеющих один и тот же набор эквивалентных классов размерностью N' . Эквивалентный класс может включать в

себя направленные и ненаправленные дуги. Покажем, что метрика БДЭ является эквивалентной с точки зрения правдоподобия. Для этого введем следующую функцию, зависящую от переменной X :

$$L(X) = \prod_k \frac{\Gamma(N'P(X = k|G) + N_k)}{\Gamma(N'P(X = k|G))}, \quad (18)$$

где k — число возможных состояний, которые может принимать X , $N_k \subset D$ число выборок, в которых $X = k$.

Правдоподобие для выборки D можно записать в следующем виде:

$$P(D | G) = \prod_{i=1}^n L(X, Y) = \prod_{i=1}^n \frac{L(X_i \cup Y_i)}{L(Y_i)}. \quad (19)$$

Воспользуемся теоремой эквивалентности Чикерина [8], обобщающей определение эквивалентности байесовских сетей, введенное Перлом и Верма. Пусть заданы два графа структуры G и G' и множество дуг $A_{G,G'}$, характеризующих различие в направленности дуг графов G и G' . Тогда два графа БС G и G' будут являться эквивалентными, когда существует множество противоположно направленных дуг $A'_{G,G'}$, которые могут быть добавлены к графу G , чтобы получить G' и удовлетворяющие следующим условиям: после добавления всех дуг $a \in A'_{G,G'}$ результирующая структура БС не содержит циклов и эквивалента G' ; если $x \rightarrow y$ есть каждая следующая дуга, для которой происходит изменение направления, то x и y , будет иметь одних и тех же родителей в G и G' .

Следствие теоремы Чикерина заключается в том, что любая структура БС может быть преобразована в эквивалентную форму за счет последовательных изменений направленностей дуг $A'_{G,G'}$. Несложно доказать, что метрика БДЭ является эквивалентной с точки зрения правдоподобия. Пусть G и G' два графа, описывающих структуры БС и отличающиеся лишь направленностью дуги $X_i \rightarrow X_j$ в G , Y_i — множество родительских верши для X_i . Согласно теореме Чикерина $Y_j = X_i \cup Y_i$ — родительские вершины для X_j в графе G и $Y_i = X_j \cup Y_j$ для G' . С учетом того, что G и G' отличаются лишь направленностью дуги $X_i \rightarrow X_j$, правдоподобие выборки для G имеет следующий вид:

$$L(X, Y) = \frac{L(X_i \cup Y_j)}{L(Y_j)} \frac{L(\{X_i, X_j\} \cup Y_j)}{L(X_i \cup Y_j)}. \quad (20)$$

Для графа G' имеем:

$$L'(X, Y) = \frac{L(X_j \cup Y_i)}{L(Y_i)} \frac{L(\{X_i, X_j\} \cup Y_i)}{L(X_j \cup Y_i)}. \quad (21)$$

Так как в обоих графа X_i и X_j имеют одни и тех же родителей $Y_i = Y_j$, $L(X, Y)$ и $L'(X, Y)$ будет эквивалентными. Следовательно, $P(D|G) = P(D|G')$, и условие эквивалентности правдоподобия для БДЭ показано. В процессе обучения структуры сети могут быть использованы различные алгоритмы поиска, основной задачей которых является нахождение $\arg \max_G P(\theta|G)$ на основе вычисления метрики БДЭ. Наиболее оптимальными являются алгоритмы: минимаксного восхождения [9], Питера — Кларка (РС-алгоритм) [10], марковского покрытия с применением статистических оценок. Первые два алгоритма базируются на методе поиска с восхождением, последний на методе Левенберга — Марквардта (ЛМ). Отметим, что применение метода ЛМ в значительной степени позволяет повысить правдоподобие структуры сети, так как метрика БДЭ является неразложимой, следовательно, она будет зависеть только от множества родительских узлов Y , связанных с текущей переменной Y . Операциями, влияющими на изменение метрики БДЭ являются: добавление, удаление и изменение направленности дуг. Неразложимость метрики БДЭ дает возможность ее одновременного вычисления сразу для нескольких вершин. В таком случае при изменении дуг, связанных с текущей вершиной X_i требуется лишь перерасчет слагаемых БДЭ, отвечающих за вклад, вносимый родителями Y_i . Следовательно, алгоритмы обучения на основе вычисления метрики БДЭ могут быть распараллелены, что дает возможность повысить интенсивность обучения и снизить временные затраты. В результате выполнения алгоритмов обучения подразумевается получение наиболее правдоподобной структуры сети, отвечающей текущей предметной области и формируемой в соответствии с полученной

обучающей выборкой D . Для верификации алгоритмов обучения рассмотрим структурное расстояние Хэмминга (СРХ), а также оценку метрик БДЭ на основе принципа максимальной энтропии (МПЭ). В качестве расстояния СРХ $H^*(G, G')$ понимается число направленных дуг, на которые будут отличаться графы G и G' , чем больше отличий в множестве дуг G и G' , тем больше значение структурного расстояния Хэмминга.

Далее рассмотрим подход на основе МПЭ. Для этого введем классическое определение энтропии Шеннона применительно к семантике БС с учетом выражения (1)

$$H(X|\theta) = \sum_{i=1}^N H(X_i|\theta_{X_i}), \quad (22)$$

где $H(X_i|\theta_{X_i})$ — энтропия X_i при наличии множества родительских вершин $Y = Parents(X_i)$.

В таком случае, апостериорное математическое ожидание энтропии $H(X_i)$ при заданной обучающей выборке D запишем в следующем виде

$$\begin{aligned} \mathbb{E}(H(X_i)|D) &= \int \rho(X, \theta) d\theta_{X_i}, \\ \rho(X, \theta) &= H(X_i|\theta_{X_i}) P(\theta_{X_i}|D), \end{aligned} \quad (23)$$

где $P(\theta_{X_i}|D)$ — распределение Дирихле, имеющее следующий вид

$$\begin{aligned} P(\theta_{X_i}|D) &= \int \rho(\theta, \alpha, D) d\alpha_{i,j,k}, \\ \rho(\theta, \alpha, D) &= P(\theta_{X_i}|D, \alpha_{i,j,k}) P(\alpha_{i,j,k}|D). \end{aligned} \quad (24)$$

Тогда перепишем выражение (24) для вычисления $P(H(X_i)|D)$ с учетом метрики БД

$$\begin{aligned} \mathbb{E}(H(X_i)|D) &= \mathbb{E}_{\alpha_{i,j,k}} BD(X_i|Y, \alpha_{i,j,k}), \\ \mathbb{E}_{\alpha_{i,j,k}} &= \mathbb{E}(H(X_i)|D, \alpha_{i,j,k}). \end{aligned} \quad (25)$$

Выражение для вычисления математического ожидания $\mathbb{E}(H(X_i)|D, \alpha_{i,j,k})$ получим с использованием подхода, предложенного Андерсоном [11]

$$\begin{aligned} P(H(X_i)|D, \alpha_{i,j,k}) &= H_{X_i} = \psi - \varphi_q, \\ \varphi_q &= \sum_{j=1}^q \frac{r-1}{2(a_{ij} + N_{ij})}, \\ H_{X_i} &= H(X_i|D, \alpha_{i,j,k}) - \sum_{j=1}^q \frac{r-1}{2(a_{ij} + N_{ij})}, \end{aligned} \quad (26)$$

$$\psi = - \sum_{j=1}^q \sum_{k=1}^r \frac{a_{ijk} + N_{ijk}}{a_{ij} + N_{ij}} \ln \left(\frac{a_{ijk} + N_{ijk}}{a_{ij} + N_{ij}} \right).$$

Сформулируем принцип ПМП для оценки энтропии графов G' и G'' , отличающихся добавлением дополнительного родительской вершины в G'' :

$$P(H'(X) | D) \leq P(H''(X) | D), \quad (27)$$

где H' и H'' значения энтропий, соответствующие графам G' и G'' соответственно.

Рассматривая МПЭ применительно к семантике БС, Сузуки [12] установил зависимость энтропии и метрики БД. Следствие данной зависимости сформулируем в виде следующих неравенств:

$$\begin{aligned} H(X_i | Y', p_{i,jk}) &\leq H(X_i | Y'', p_{i,jk}), \\ Y' &\subset Y'', \end{aligned} \quad (28)$$

$$\begin{aligned} p_{i,jk} &= \frac{\alpha_{ijk} + N_{ijk}}{\alpha_{ij} + N_{ij}}, \\ BD(X_i | Y', \alpha_i) &\geq BD(X_i | Y'', \alpha_i). \end{aligned} \quad (29)$$

При увеличении числа обучающих выборок D наблюдаем сходимость вероятностей $p_{i,jk}$ к значениям α_{ijk} . Следовательно, асимптотически (29) будет эквивалентно следующему неравенству:

$$\begin{aligned} H(X_i | Y', \alpha_{ijk}) &\leq H(X_i | Y'', \alpha_{ijk}), \\ Y' &\subset Y''. \end{aligned} \quad (30)$$

Из неравенств (29) и (30) вытекает обратная зависимость между энтропией и метрикой Байеса — Дирихле.

Применительно к эквивалентной метрике Байеса — Дирихле, с учетом того, что $\alpha_{ijk} = \alpha_i / rq$ выражений (36) и (30) сформулируем МПЭ:

$$\begin{aligned} \mathbb{E}' BDe(X_i | Y', \alpha_i) &\leq \mathbb{E}'' BDe(X_i | Y'', \alpha_i), \\ \mathbb{E}' &= \mathbb{E}(H'(X_i) | D, \alpha_{i,j,k}), \\ \mathbb{E}'' &= \mathbb{E}(H''(X_i) | D, \alpha_{i,j,k}). \end{aligned} \quad (31)$$

В том случае, если выборка D содержит пропуски данных для некоторого множества переменных $X_h = X_h^1, X_h^2, \dots, X_h^n$, области выборок, соответствующие графам G' и G'' могут отличаться. Как следствие получаем случай обучения с неполными данным, а оценка X_i на основе метрики БДЭ для случая G' и

G'' могут отличаться, так как значение родительских вершины Y'' соответствующие графу G'' в соответствии с обучающей выборкой не определены. В таком случае в выражении (31) уточняется параметр α_i в правой части неравенства:

$$\begin{aligned} \mathbb{E}' BDe(X_i | Y', \alpha_i) &\leq \mathbb{E}'' BDe'', \\ \mathbb{E}' &= \mathbb{E}(H'(X_i) | D, \alpha_{i,j,k}), \\ \mathbb{E}'' &= \mathbb{E}(H''(X_i) | D, \alpha_{i,j,k}) \\ BDe'' &= BDe(X_i | Y'', \alpha_i (\tilde{q} / q)), \end{aligned} \quad (32)$$

где \tilde{q} — число возможных состояний, для которых $N_{i,j,k} > 0$.

Применение МПЭ позволяет произвести верификацию значений оценок $BDe(X_i | Y', \alpha_i)$ и $BDe(X_i | Y'', \alpha_i)$, формируемых в результате выполнения алгоритмов обучения для двух графов G' и G'' и характеризующих полученную по результатам обучения БС и эталонную БС, заданную экспертным путем. В рамках оценки эквивалентности G' и G'' будем использовать комбинацию методов ПМП и вычисления расстояния СРХ.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

В экспериментальной части исследования проведем оценку правдоподобия существующих алгоритмов обучения структуры БС с учетом СРХ и ПМП. В качестве эталонных БС будем использовать общедоступные сети с заранее определенным набором обучающих данных и известной структурой. В качестве таких сетей будем использовать: Alarm (37 узлов, 46 дуг, 509 параметров), Water (32 узла, 66 дуг, 10082 параметра) и Mildew (34 узла, 46 дуг, 540150 параметров), Andes (223 узла, 338 дуг, 1157 параметров), Link (724 узла, 1125 дуг, 14211 параметров).

Рассмотрим оценку расстояния СРХ для каждой из сети Mildew, Andes и Link в зависимости от размерности обучающей выборки. Чем меньше значение СРХ, тем выше правдоподобие структуры на основе алгоритма МПСО. В процессе проведения вычислений предполагается, что формирование оценок БДЭ для каждой переменной БС X_i производится независимо друг от друга. После фор-

мирования оценок производится поиск оптимального значения критерия БДЭ и выполнение операций добавления, удаление или изменение направленности дуг БС.

На рис. 1 приведем численную оценку СРХ от объема обучающей выборки каждой их сетей Mildew, Andes и Link для алгоритмов МПСО, минимаксного восхождения (ММВ), возрастания-сокращения (ВС) и алгоритм Питера —Кларка (РС).

Рис. 1, характеризующий зависимость СРХ от размера выборок, а также от числа параметров БС, наглядно показывает эффективность применения описанных в статье комбинаций оценочных алгоритмов. Алгоритмы МПСО и ММВ позволяют получить

минимальное значение расхождение обученной структуры и эталонной, и, как следствие, оптимальную модель БС в соответствии с обучающей выборкой, формируемой в соответствии с решением задач заданной предметной области.

ЗАКЛЮЧЕНИЕ

Решение задачи оценки эквивалентности и правдоподобия структуры графа БС является важным аспектом проверки корректности работы алгоритмов обучения. Для решения данной задачи могут быть использованы различные алгоритмы на основе принципа максимальной энтропии, вычисления структурного

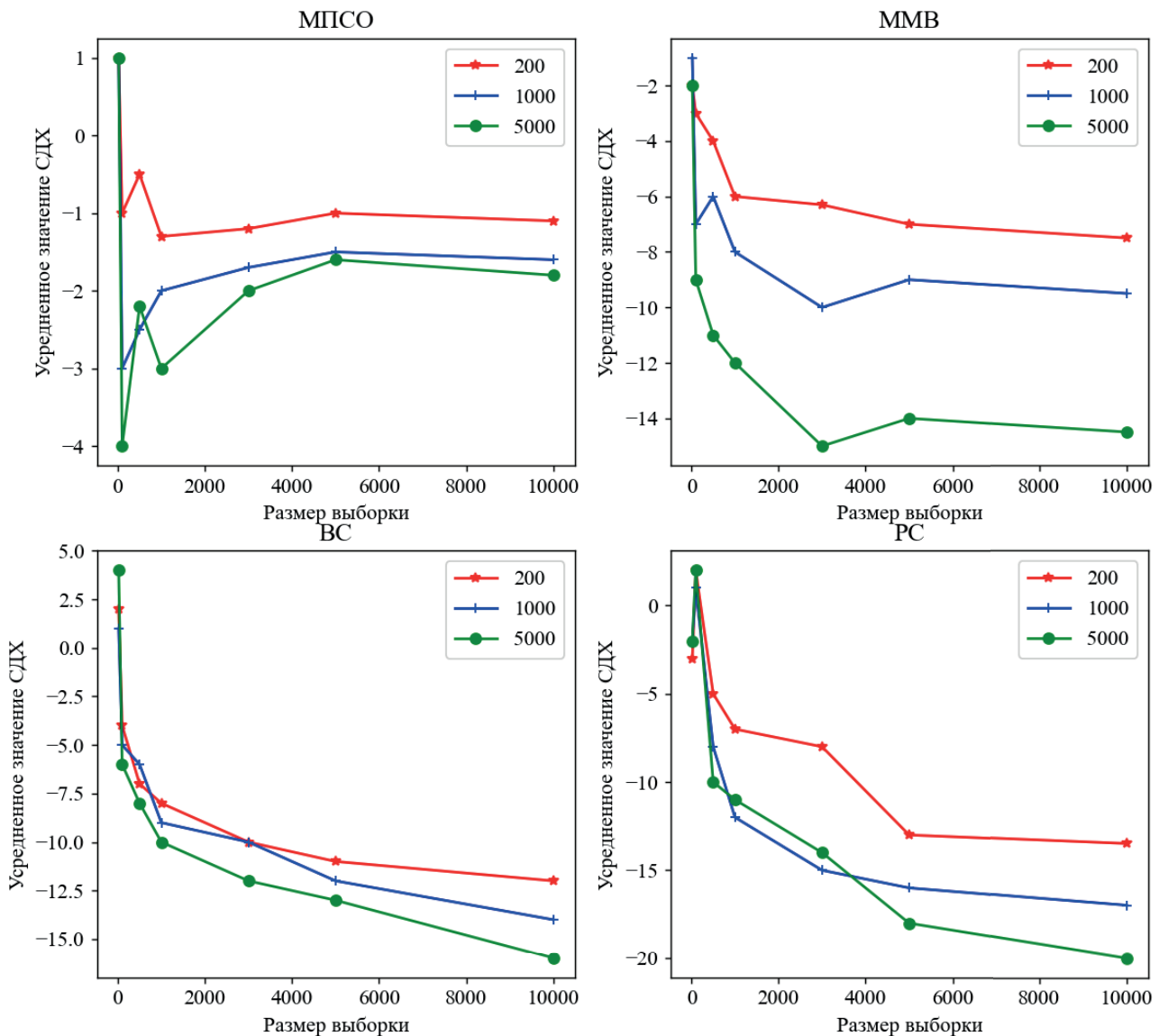


Рис. 1. Усредненное значение расстояния СРХ от объема обучающей выборки ДБС [Fig. 1. Average value of SHD distance from DBN training sample volume]

расстояния Хэмминга. Особую роль при формировании оценок Байеса — Дирихле и эквивалентной оценки Байеса — Дирихле является оценка их состоятельности с учетом определения перекрестной энтропии в соответствии с критерием условной независимости. Проведенный анализ существующих алгоритмов обучения БС на основе предложенного подхода, позволяет оценить эффективность каждого из них, а также определить алгоритмы, в большей степени подходящие для решения конкретных задач из заданной предметной области. Особое внимание в работе уделено вопросам оценки эквивалентности структуры БС с точки зрения правдоподобия за счет обобщения теоремы Чикерина и определение условий для $L(X, Y)$ и $L'(X, Y)$. Применение данного подхода позволят показать, что при выполнении одних и тех же правил условной независимости для X и Y при заданном множестве X получаем эквивалентные структуры графом G и G' .

Практические результаты показывают эффективность рассматриваемых комбинаций оценочных алгоритмов. Использование МПЭ позволяет определить энтропию H' и H'' с учетом эталонной структуры графа, а также описать связи между X_i и Y_i в соответствии с неравенством (31) и (32) для случая полного и частичного наблюдения родительских переменных Y_i . Рассмотренные подходы являются научно-обоснованными и имеют высокую практическую значимость при решении задач верификации моделей на основе БС и алгоритмов их обучения.

КОНФЛИКТ ИНТЕРЕСОВ

Автор декларирует отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Scutari, M.* An Empirical-Bayes Score for Discrete Bayesian Networks / *M. Scutari // Machine Learning*, 2016. – Vol. 52. – P. 438–448.

Полухин Павел Валерьевич — канд. техн. наук, кафедра математических методов исследования операций факультета прикладной математики, информатики и механики Воронежского государственного университета. E-mail: alfa_force@bk.ru
ORCID iD: <https://orcid.org/0000-0001-5645-6312>

2. *Рассел, С.* Искусственный интеллект: современный подход, 2 издание: Пер. с англ. / *С. Рассел, П. Норвиг.* – М. : Вильямс, 2006. – 1408 с.

3. *Heckerman, D.* Learning discrete Bayesian networks / *D. Heckerman, D. Geiger, D. Chickering // Machine Learning*, 1995. – Vol. 20. – P. 197–243.

4. *Азарнова, Т. В.* Формирование структуры байесовской сети процесса тестирования надежности информационных систем / *Т. В. Азарнова, Н. Г. Аснина, Д. К. Проскурин, П. В. Полухин // Вестник Воронежского государственного технического университета*, 2017. – Т. 13. – № 6. – 156 с.

5. *Pearl, J.* Causality: Models, Reasoning and Inference / *J. Pearl.* – N.Y. : Cambridge University Press, 2009. – 484 p.

6. *Pearl, J.* Equivalence and Synthesis of Causal Models / *J. Pearl, T. Verma // Proc. UAI.* – NY : Morgan Kaufman, 1990. – P. 220–227.

7. *Buntine, W.* Theory refinement on Bayesian networks / *W. Buntine // Proc. UAI.* – NY: Morgan Kaufman, 1991. – P. 52–60.

8. *Chickering, D. M.* A Transformational Characterization of Equivalent Bayesian Network Structures / *D. M. Chickering // Proc. UAI.* – NY : Morgan Kaufman, 1995. – P. 87–98.

9. *Tsamardinos, I.* The max-min hill-climbing Bayesian network structure learning algorithm / *I. Tsamardinos, L. E. Brown, C. F. Aliferis // Machine Learning*, 2006. – No 65. – P. 31–78.

10. *Spirtes, P.* Causation, Prediction and Search / *P. Spirtes, C. Glymour, R. Sheines.* – Cambridge: MIT Press, 2000. – 568 p.

11. *Anderson, G.* A Monotonicity Property of the Gamma Function / *G. Anderson, S. Qiu // Proc. AMS*, 1997. – Vol. 125(11). – P. 3355–3362.

12. *Suzuki, J.* A Theoretical Analysis of the BDeu Scores in Bayesian Network Structure Learning / *J. Suzuki // Behaviormetrika*, 1995. – Vol. 1(1). – P. 1–20.

BAYESIAN NETWORK EQUIVALENCE ANALYSIS VIA BAYESIAN — DIRICHLET ASYMPTOTIC SCORE

© 2022 P. V. Polukhin✉

Voronezh State University

1, Universitetskaya Square, 394018 Voronezh, Russian Federation

Annotation. Dynamic and static Bayesian networks are an effective tool for modeling stochastic processes. The areas of practical implementation of these models have significantly expanded recently. The quality of their application in solving practical problems is largely determined by the capabilities of algorithms for learning the structure and probabilistic parameters of models that allow the network to be configured to solve the range of applied problems under consideration. Tools based on equivalence principles play an important role in the algorithms for determining the optimal structure of Bayesian networks and the modification of algorithms for setting their parameters. On the basis of the equivalence principle, asymptotic estimates of transformations obtained in the process of adding, changing or deleting individual nodes of the Bayesian network graph are formed and an apparatus for obtaining a local a priori distribution for each of the network parameters is created. In this paper, we study tools for estimating the equivalence of Bayesian networks based on the Bayes-Dirichlet metric, the Hamming and Kullbak-Leibler structural distance. These tools can also be applied to dynamic Bayesian networks, for working with which it is additionally necessary to determine the structure of the transition model between time slices. In the framework of the study, the issues of equivalence of a priori probability distributions formed in the process of parameter learning are also considered. In the final part of the paper, a computational experiment is presented that reflects the effectiveness of using various learning algorithms from the point of view of comparing their results with equivalent reference Bayesian networks. The tools proposed in the work allow adapting static and dynamic Bayesian models to solve practical problems, optimizing the learning processes of these models by equivalence principle of graphical probabilistic models.

Keywords: Bayesian network equivalence, asymptotic score, Bayesian — Dirichlet equivalence score, structural Hamming distance.

CONFLICT OF INTEREST

The author declare the absence of obvious and potential conflicts of interest related to the publication of this article.

REFERENCES

1. Scutari M. (2016). An Empirical-Bayes Score for Discrete Bayesian Networks. *Machine Learning*. (52). P. 438–448.

2. Russel S. and Norvig P. (2006) Artificial intelligence a modern approach, second edition. Moscow : Williams (in Russian).

3. Heckerman D. and Geiger D. (1995) Learning discrete Bayesian networks. *Machine Learning*. (20). P. 197–243.

4. Azarnova T. V., Asnina N. G., Proskurin D. K. and Polukhin P. V. (2017) Formirovanie struktury Bajesovskoj seti processa testirovaniya nadezhnosti informacionnyh system. *Vestnik Voronezhskogo gosudarstvennogo tehničeskogo universiteta*. 13 (6). P. 45–51 (in Russian).

5. Pearl J. (2009) Causality: Models, Reasoning and Inference. New York : Cambridge University Press.

6. Pearl J. and Verma T. (1990) Equivalence and Synthesis of Causal. *Proc. UAI*. P. 220–227.

7. Buntine W. (1991) Theory refinement on Bayesian network. *Proc. UAI*. P. 52–60.

✉ Polukhin Pavel Valeryevich
e-mail: alfa_force@bk.ru

8. Chickering D. M. (1995) A Transformational Characterization of Equivalent Bayesian Network Structures. *Proc. UAI*. P. 87–98.
9. Tsamardinos I., Brown L. E. and Aliferis C. F. (2006) The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*. (65). P. 31–78.
10. Spirtes P., Glymour C. and Scheines R. (2000) Causation, Prediction and Search. Cambridge: MIT Press.
11. Anderson G. and Qiu S. (1997) A Monotonicity Property of the Gamma Function. *Proc. AMS*. 125(11). P. 3355–3362.
12. Suzuki J. (1995) A Theoretical Analysis of the BDeu Scores in Bayesian Network Structure Learning. *Behaviormetrika*. 1(1). P. 1–20.

Polukhin Pavel Valeryevich — 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

ORCID iD: <https://orcid.org/0000-0001-5645-6312>