

П.В. Полухин

ФОРМИРОВАНИЕ ОЦЕНОЧНЫХ ФУНКЦИЙ
ДЛЯ РЕШЕНИЯ ЗАДАЧИ ПОСТРОЕНИЯ НАПРАВЛЕННОЙ
ДИНАМИЧЕСКОЙ БАЙЕСОВСКОЙ СЕТИ

Аннотация. Формирование и описание оценочных функций является важной задачей определения необходимых и достаточных условий существования направленности связей между узлами байесовской сети. Применение оценочных функций различного типа позволяет оптимизировать процедуру обучения сети и определить ту функцию, которая наиболее адаптирована для конкретной сети с фиксированным набором обучающей выборки. В работе рассматриваются оценочные функции, построенные на основе логарифма правдоподобия, взаимной энтропии, а также асимптотической метрики Байеса – Дирихле. Данные функции широко применяются в процессе решения задач обучения структуры вероятностных моделей и могут быть адаптированы для определения направленности связей между узлами временных моделей, построенных на основе динамических байесовских сетей.

Ключевые слова: взаимная информация, направленный ациклический граф, байесовская сеть, метрика Байеса – Дирихле, гамма-функция.

P.V. Polukhin

FORMATION OF EVALUATION FUNCTIONS FOR SOLVING THE TASK
OF CONSTRUCTING DIRECTIONAL DYNAMIC BAYESIAN NETWORK

Abstract. The formation and description of evaluation functions is an important task for determining the necessary and sufficient conditions for the existence of directional links between Bayesian network nodes. The use of evaluation functions of various types allows you to evaluate and optimize the network training procedure and determine the function that is most adapted for a particular network with a fixed set of training samples. The work examines evaluation functions based on the logarithm of likelihood, mutual entropy, as well as the Bayes – Dirichlet asymptotic metric. These functions are widely used in the process of solving problems of training the structure of probabilistic models and can be adapted to determine the direction of connections between nodes of temporary models built on the basis of dynamic Bayesian networks.

Keywords: mutual information, directed acyclic graph, Bayesian network, Bayes – Dirichlet metric, gamma-function.

Введение

Вероятностные модели являются достаточно хорошо апробированным инструментом для моделирования сложных стохастических процессов, в том числе имеющих переходные состояния во времени. Наибольший интерес представляют вероятностные модели байесовских сетей (далее – БС), позволяющие определить вероятностные связи между узлами путем задания таблиц условных вероятностей.

Процедура получения направленной структуры требует формализации существующих подходов к обучению и определению тех оценочных функций, которые наиболее применимы для решения определенной практической задачи. Применение оценочных функций в процессе обучения тесно связано с операциями добавления, удаления и изменения направленности связей между ребрами. Для каждой из операций рассчитывается значение оценки, характеризующей вес каждого ребра, после чего выбирается значение

Полухин Павел Валерьевич

кандидат технических наук, преподаватель кафедры математических методов исследования операций, Воронежский государственный университет, город Воронеж. Сфера научных интересов: машинное обучение, статистический и системный анализ данных, математическое и компьютерное моделирование, программные комплексы. Автор более 30 опубликованных научных работ.

Электронный адрес: alfa_force@bk.ru

оценки, имеющей максимальный вес. Процедура поиска экстремума оценочной функции направлена на получение корректной структуры графа БС с максимальными весами и на установление зависимости весов ребер графа от обучающей выборки, поступающей на вход данного алгоритма. Применение различных алгоритмов поиска минимаксного значения оценок обусловлено реализацией задачи повышения точности за счет исключения образования локальных оптимумов и как следствие получения некорректной структуры вероятностной сети.

Процедура формирования оценочных функций

Для задания оценочных функций предварительно определим семантику БС, представляющей собой граф $G=(X,E)$, где $X=(X_1, X_2, \dots, X_n)$ – переменные, характеризующие узлы байесовской сети; $D_i=(D_1^i, D_2^i, \dots, D_n^i)$ – домен значений, который могут принимать переменные X_i . Каждой из переменных X соответствует условное распределение $P(X|Parents(X))$, формируемое с учетом присутствия родительских вершин $Parents(X)$. Если для вершины X отсутствуют родительские вершины, распределение $P(X)=P(X|Parents(X))$ будет являться безусловным. Полное совместное распределение БС будет иметь вид [1; 2]

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | Parents(X_i)). \quad (1)$$

Из общей формулировки БС следует, что процедура обучения структурно делится на два основных этапа: обучение структуры и параметров. Рассмотрим более детально обучение структуры сети и связанную с ней процедуру формирования определенных типов оценочных функций, участвующих в получении результирующей оценки каждой из дуг графа, и формирование БС с максимальным значением правдоподобия структуры. В этом случае задача поиска оптимальной структуры сводится к решению следующей экстремальной задачи:

$$\hat{G} = \arg \max f(G, D). \quad (2)$$

Определим основные оценочные функции, применяемые в процессе решения задачи обучения структуры БС. Для формирования оценочных функций введем понятие локальной и глобальной условной независимости родительских вершин переменной X :

$$P(\Theta_i^G | G) = \prod_{j=1}^{q_i} P(\Theta_{i,j}^G | G), \quad (3)$$

Формирование оценочных функций для решения задачи построения направленной ...

где $\Theta_i^G = \bigcup_{j=1}^{q_i} \Theta_{i,j}^G$ – значения параметров для переменной X_i с учетом родительских вершин $Parents(X_i)$ для графа G .

На основании свойства условной независимости Θ_i^G можно определить распределение Дирихле следующего вида [3]:

$$P(\Theta^G | G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\bar{A}(\sum_{k=1}^{r_i} \alpha_{i,j,k})}{\prod_{k=1}^{r_i} \bar{A}(\alpha_{i,j,k})} \prod_{k=1}^{r_i} (\Theta_{i,j,k}^G)^{\alpha_{i,j,k}-1}. \tag{4}$$

С учетом распределения Дирихле можно определить метрику Байеса – Дирихле следующего вида [4]:

$$BD = \sum_{i=1}^n \sum_{j=1}^{q_i} \ln \frac{\Gamma(\sum_{k=1}^{r_i} \alpha_{i,j,s})}{\Gamma(\sum_{s=1}^{r_i} N_{i,j,s} + \alpha_{i,j,s})} + \sum_{s=1}^{r_i} \frac{\Gamma(N_{i,j,s} + \alpha_{i,j,s})}{\Gamma(\alpha_{i,j,s})}, \tag{5}$$

$$\alpha_{i,j,k} = \alpha(x_i = k, P(x_i) = j) = \frac{1}{r_i q_i},$$

где $\alpha_{i,j,k}$ – параметр распределения Дирихле; r_i – домен всех возможных значений для родительских вершин $Parents(X_i)$; q_i – произведение по всем r_i ; $\Gamma(\cdot)$ – гамма-функция.

Частным случаем оценки Байса – Дирихле является метрика $K2$, формируемая из выражения (5) при $\alpha_{i,j,k} = 1$ [5]:

$$K2 = \ln P(G) \sum_{i=1}^n \sum_{j=1}^{q_i} \ln \frac{(r_i - 1)!}{(N_{i,j} + r_i - 1)!} + \sum_{s=1}^{r_i} \ln(N_{i,j,s}!). \tag{6}$$

В реальных условиях вычисление параметра $\alpha_{i,j,k}$ сопряжено со сложностями, в связи с этим необходимо установить его зависимость от размера эквивалентной выборки, определяемой для БС с учетом ее семантики:

$$\alpha_{i,j,k} = N\alpha(x_1, \dots, x_n), \tag{7}$$

где $\alpha(x_1, \dots, x_n)$ – вероятность, соответствующая априорной структуре БС.

Вероятность $P(X|G)$ с учетом положительной меры $\{1, \dots, r_1\} \times \dots \times \{1, \dots, r_n\}$ можно задать в следующем виде:

$$\alpha(x_1, \dots, x_n) = \frac{1}{\prod_{i=1}^n r_i}. \tag{8}$$

Перепишем параметр Дирихле (8) в виде

$$\alpha_{i,j,k} = \alpha(x_i = k, P(x_i) = j) = \frac{1}{r_i q_i}. \tag{9}$$

Из выражения (5) с учетом (9) получим математическое описание эквивалентной метрики Байеса – Дирихле [6]:

$$BDe = \sum_{i=1}^n \sum_{j=1}^{q_i} \ln \frac{\Gamma\left(\frac{1}{q_i}\right)}{\Gamma\left(\sum_{s=1}^{r_i} N_{i,j,s} + \frac{1}{q_i}\right)} + \sum_{s=1}^{r_i} \frac{\Gamma\left(N_{i,j,s} + \frac{1}{q_i r_i}\right)}{\Gamma\left(\frac{1}{q_i r_i}\right)}. \quad (10)$$

Наряду с метриками, построенными на основе метрики Байеса – Дирихле, рассмотрим формирование оценочных функций, реализованных на основе логарифма правдоподобия и взаимной информации. Оценочную функцию на основе логарифма правдоподобия можно записать в виде

$$L(G, \theta^G, D) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{i,j,k} \ln \frac{N_{i,j,k}}{N_{i,j}}. \quad (11)$$

Чтобы определить критерии Шварца и Акаике, получаемые из выражения логарифма правдоподобия, найдем общее число параметров БС по формуле [7]

$$M = \sum_{i=1}^n (r_i - 1) q_i. \quad (12)$$

Метрику минимальной длины описания (далее – МДО) получим из выражений (11) и (12):

$$f(M) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{i,j,k} \ln \frac{N_{i,j,k}}{N_{i,j}} - \frac{1}{2} M \ln(N). \quad (13)$$

Альтернативным подходом для формирования оценок структуры БС является использование теории информации. Данный подход позволяет получить наиболее качественную оценку структуры с учетом имеющейся обучающей выборки. Обобщая выражение (13), можно получить выражение для оценок Шварца и Акаике следующего вида:

$$Q(M) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{i,j,k} \ln \frac{N_{i,j,k}}{N_{i,j}} - MF(N). \quad (14)$$

Из выражения (14) при $F(N)=1$ имеем метрику Акаике, $F(N)=\ln(N)/2$ – критерий Шварца.

Формулировку логарифма правдоподобия можно определить также на основе условной энтропии $H(X)$, устанавливающей связь между дочерней вершиной X_i и ее родительскими вершинами [8]:

$$L(G, \theta^G, D) = -N \sum_{i=1}^n H(X_i | Parents(X_i)) = -NH(G), \quad (15)$$

$$H(G) = \sum_x \left(\prod_{i=1}^n P(x_i | Parents(X_i)) \times \ln \left(\prod_{i=1}^n P(x_i | Parents(X_i)) \right) \right),$$

где $H(G)$ – значение энтропии графа байесовской сети G .

Из выражения (15) следует, что оценки, построенные на основе логарифма правдоподобия, минимизируют значение условной энтропии для каждого из параметров БС с учетом всех родительских вершин. Это дает возможность найти те значения родителей, которые вносят больше информации относительно интересующей вершины X_i . Наряду с условной энтропией сформулируем оценку на основе взаимной информации. Для этого

Формирование оценочных функций для решения задачи построения направленной ...

определим распределение для всех переменных сети графа $P_g(X)$ и распределение, соответствующее обучающей выборке D для параметров θ^G с учетом семантики БС:

$$P_g(X) = \prod_{i=1}^n P(x_i | Parents(X_i)). \tag{16}$$

Тогда для распределений $P_d(X)$ и $P_g(X)$ можно определить выражение для расчета дистанции Кульбака – Лейблера как [9]

$$D_{KL} = \sum_X P_g(X) \ln \frac{P_g(X)}{P_d(X)}. \tag{17}$$

Используя формулировку энтропии $H(X)$, выражение (17) для расстояния Кульбака – Лейблера перепишем в следующем виде:

$$D_{KL} = -H_d(X) + \sum_{i=1}^n H_d(X_i) + \sum_{i=1}^n H_d(Y_i) - H_d(X_i), \tag{18}$$

$$Y_i = X_i \cup Parents(X_i),$$

где $H_d(X)$ – энтропия для параметров X с учетом имеющейся обучающей выборки $d \subset D$.

Определим взаимную информацию для двух распределений параметров X и $Z \in Parents(X)$:

$$MI(X, Z) = \sum_X P(X) \ln \frac{P(X|Z)}{P(X)P(Z)}. \tag{19}$$

Используя энтропию $H(X)$ и $H(Z)$, можно переписать выражение (19) в терминах энтропии Шеннона:

$$MI(X, Z) = H(X) + H(X|Z) - H(X). \tag{20}$$

Отметим, что последовательное добавление переменных в оценку на основе взаимной информации приводит к ее постепенному росту в соответствии с неравенством

$$MI(X, X \cup Parent_1(X), Parent_2(X)) \geq MI(X, X \cup Parent_1(X)). \tag{21}$$

Установим связь между взаимной информацией $MI(X, Z)$ и оценкой правдоподобия (11):

$$\sum_{i=1}^n MI(X_i, Parent(X_i)) = \frac{1}{N} L(G, \theta^G, D) + \sum_{i=1}^n H(X_i). \tag{22}$$

Тогда, максимизируя оценку логарифма правдоподобия, получим следующее выражение:

$$\sum_{i=1}^n MI(X_i, Parent(X_i)) = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{i,j,k} \ln \frac{N N_{i,j,k}}{N_{i,k} N_{i,j}}. \tag{23}$$

Отметим, что аналогичным образом можно получить выражение для оценки условной взаимной информации при наличии переменных X, Z и V :

$$MI(X, Z|V) = \sum_X P(Z) \sum_{X,Z} P(X, Z|V) \ln \frac{P(X, Z|V)}{P(X|V)P(Z|V)}. \tag{24}$$

Рассмотрим процедуру применения оценки на основе взаимной информации применительно к структуре динамической байесовской сети (далее – ДБС). Рассмотрим струк-

туру двухслойной ДБС с множеством обучающих данных для двух временных срезов D_0 и D_1 . Тогда значение логарифма правдоподобия для графа, характеризующего начальный срез ДБС, можно определить с учетом выражения (11) следующим образом:

$$L_0(G_0, \theta^{G_0}, D_0) = \sum_{i=1}^n \sum_{k=1}^{r_i} N_{ik} \ln \frac{N_{ik}}{N}. \quad (25)$$

Аналогично получим оценку логарифма правдоподобия $L_1(G_1, \theta^{G_1}, D_1)$ для графа G_1 с учетом обучающей выборки D_1 . Тогда искомая оценка взаимной информации, соответствующая графам из двух смежных временных срезов, будет иметь вид

$$\sum_{i=1}^n MI(X_i, Parent(X_i)) = \frac{\Delta L}{N}, \quad (26)$$

$$\Delta L = L_1(G_1, \theta^{G_1}, D_1) - L_0(G_0, \theta^{G_0}, D_0).$$

Выражая разность оценок для двух срезов на основе взаимной информации, получим выражение для расчета $\Delta Q(M)$ [10]:

$$\Delta Q(M) = \left((L_1(G_1, \theta^{G_1}, D_1) - M_1 F(N)) - (L_0(G_0, \theta^{G_0}, D_0) - M_0 F(N)) \right) =$$

$$= \frac{1}{2} \sum_{i=1}^n 2NMI(X_i, Parent(X_i)) - \sum_{i=1}^n (r_i - 1)(q_i - 1). \quad (27)$$

Из выражения (27) следует, что мы можем получить универсальную оценку для метрик Шварца, Акаике и логарифм правдоподобия для двух смежных временных срезов с учетом взаимной информации между данными срезами. Отметим, что взаимная информация между вершинами, имеющими связи между срезами, формируется с учетом родительских вершин $Parent(X_i)$ из начального среза на основе анализа обучающих выборок D_0 и D_1 . Чем больше значение метрики, тем больше вес ребра между родительской и дочерней вершинами двухслойной ДБС.

Для оценки достоверности получаемой структуры на основе вычисления метрик Байеса – Дирихле, логарифма правдоподобия, Шварца и Акаике воспользуемся вычислением максимума перекрестной энтропии (далее – МПЭ). Сформулируем МПЭ для двух графов ДБС, соответствующих срезам:

$$P(H_0(X_0, Parents(X_1)) | D_0) \leq P(H_1(X_1, Parents(X_1)) | D_1). \quad (28)$$

Рассмотрим МПЭ применительно к семантике БС на основе подхода, предложенного ранее в работах Сузуки:

$$H_0(X_i | p_{i,jk}) \leq H_1(X_i | Y', p_{i,jk}), Y' \subset Y'' \subset Parents(X_i),$$

$$p_{i,jk} = \frac{\alpha_{ijk} + N_{ijk}}{\alpha_{ij} + N_{ij}}, \quad (29)$$

$$BD(X_i | Y, \alpha_i) \geq BD(X_i | Y', \alpha_i).$$

С ростом размерности обучающих выборок D_0 и D_1 устанавливаем сходимость $p_{i,jk}$ к α_{ijk} . Тогда имеем

$$H_0(X_i | Y, \alpha_{ijk}) \leq H_1(X_i | Y', \alpha_{ijk}), Y \subset Y'. \quad (30)$$

Заключение

Рассмотренные в работе процедуры формирования оценочных функций напрямую влияют на качество обучения структуры БС и ДБС. Применение различных оценочных метрик позволяет адаптировать процедуру обучения для решения конкретной практической задачи и определить ту метрику, которая позволяет получить наиболее правдоподобную структуру. Метрики Шварца и Акаике являются развитием логарифма правдоподобия за счет внесения дополнительной информации относительно родительских вершин, а также корректировки оценки с учетом размера данных N . В то же время метрики на основе условной и взаимной информации доказывают свою состоятельность применительно к семантике двухслойных моделей ДБС и достаточно просто могут быть выражены через оценку логарифма правдоподобия $L(G, \theta^G, D)$. Применение оценок для ДБС позволяет установить факт изменения оценок для каждого из временных срезов, что особенно проявляется для переменных, имеющих транзитивные связи между вершинами. В этом случае общая оценка представляет собой разностную меру оценок для каждого из графов G_0 и G_1 . В процессе обучения БС и ДБС данные метрики могут быть использованы в алгоритмах восхождения к вершине, имитации отжига для поиска максимума оценки в результате процедуры добавления, удаления или изменения направленности графа сети. В этом случае вычисление метрики может выполняться рекурсивно для каждой переменной X сети после формирования узлов-кандидатов в родительские вершины $Parent(X)$. В случае ДБС в состав узлов-кандидатов попадают родительские вершины из предыдущего временного среза $Parents(X_0)$, а формирование оценок происходит с учетом обоих наборов обучающих данных D_0 и D_1 .

Другой затронутой темой исследования является оценка достоверности структуры БС и ДБС. Применение оценки МПЭ позволяет определить тот граф сети, который в наибольшей степени отражает все параметры обучающих выборок D_0 и D_1 для двух срезов ДБС. Двухслойная модель ДБС может быть представлена в виде набора статических БС, связанных между собой ребрами. В результате МПЭ позволяет оценить наиболее согласованную структуру, где каждая дуга графа имеет правильную направленность, обуславливающую вероятностную семантику БС для каждой из вершин с учетом множества родителей.

Литература

1. Koller D., Friedman N. Probabilistic graphical models: Principles and Techniques. Cambridge : MIT Press, 2009. 1270 p. ISBN 9780262258357.
2. Korb K.B., Nicholson A.E. Bayesian Artificial Intelligence. Boca Raton, London New York, Washington, D.C. : Chapman & Hall/CRC Press LLC, 2004. 491 p.
3. Сироткин А.В. Байесовские сети доверия: дерево сочленений и его вероятностная семантика // Труды СПИИРАН. 2006. Т. 1. № 3. С. 228–239. EDN NCKIFL.
4. Chickering D.M. A Transformational Characterization of Equivalent Bayesian Network Structures. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI1995)*. N.Y. : Morgan Kaufman, 1995. P. 87–98. DOI: <https://doi.org/10.48550/arXiv.1302.4938>
5. Chickering D. Optimal Structure Identification with Greedy Search // *Journal of Machine Learning Research*. 2002. No. 3. P. 507–554. URL: <https://www.jmlr.org/papers/volume3/chickering02b/chickering02b.pdf> (дата обращения: 11.10.2023).

6. Полухин П.В. Разработка параллельных алгоритмов обучения вероятностных моделей тестирования веб-приложений // Интеллектуальные системы в производстве. 2022. Т. 20. № 3. С. 94–103. EDN GPVRDF. DOI: 10.22213/2410-9304-2022-3-94-103
7. Фильченков А.А. Алгоритм построения множества минимальных графов смежности при помощи клик владений // Труды СПИИРАН. 2010. № 2 (13). С. 67–86. EDN NСNРХN.
8. Murphy K.P. Machine learning: A probabilistic perspective. Massachusetts : MIT Press, 2012. 1067 p. ISBN 0262018020.
9. Kullbak S., Leibler R.A. Information and Sufficiency // The Annals of Mathematical Statistics. 1951. Vol. 22. No. 1. P. 79–86. DOI: <http://dx.doi.org/10.1214/aoms/1177729694>
10. Lewis F.L., Lihua Xie, Popa D. (2008) Optimal and robust estimation: With an Introduction to Stochastic Control Theory. 2nd edition. Taylor & Francis. 523 p.

References

1. Koller D., Friedman N. (2009) *Probabilistic graphical models: Principles and Techniques*. Cambridge : MIT Press. 1270 p. ISBN 9780262258357.
2. Korb K.B., Nicholson A.E. (2004) *Bayesian Artificial Intelligence*. Boca Raton, London New York, Washington, D.C. : Chapman & Hall/CRC Press LLC. 491 p.
3. Sirotkin A.V. (2006) Bayesian belief networks: Junction tree and its probabilistic semantics. *SPIIRAS Proceedings*. Vol. 1. No. 3. Pp. 228–239 (In Russian).
4. Chickering D.M. (1995) A Transformational Characterization of Equivalent Bayesian Network Structures. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI1995)*. N.Y. : Morgan Kaufman. Pp. 87–98. DOI: <https://doi.org/10.48550/arXiv.1302.4938>
5. Chickering D. (2002). Optimal Structure Identification with Greedy Search. *Journal of Machine Learning Research*. No. 3. Pp. 507–554. URL: <https://www.jmlr.org/papers/volume3/chickering02b/chickering02b.pdf> (accessed 11.10.2023).
6. Polukhin P.V. (2022). Development of parallel algorithms for probabilistic models learning for web applications testing. *Intelligent Systems in Manufacturing*. Vol. 20. No. 3. Pp. 94–103. DOI: 10.22213/2410-9304-2022-3-94-103 (In Russian).
7. Filchenkov A.A. (2010). Minimal join graph set synthesis possession cliques algorithm. *SPIIRAS Proceedings*. No. 2 (13). Pp. 67–86. (In Russian).
8. Murphy K.P. (2012) *Machine learning: A probabilistic perspective*. Massachusetts : MIT Press, 1067 p. ISBN 0262018020.
9. Kullbak S., Leibler R.A. (1951) Information and Sufficiency. *The Annals of Mathematical Statistics*. Vol. 22. No. 1, Pp. 79–86. DOI: <http://dx.doi.org/10.1214/aoms/1177729694>
10. Lewis F.L., Lihua Xie, Popa D. (2008) *Optimal and robust estimation: With an Introduction to Stochastic Control Theory*. 2nd edition. Taylor & Francis. 523 p.