

**МЕТОДИКА СТРУКТУРНОГО ОБУЧЕНИЯ ДИНАМИЧЕСКИХ БАЙЕСОВСКИХ СЕТЕЙ НА ОСНОВЕ СТАТИСТИЧЕСКИХ ДАННЫХ**

***Анотація.** Розглядаються деякі існуючі підходи й пропонується нова методика статистичного структурного навчання динамічних баєсівських мереж для випадку повних вибірок даних. Запропонована методика структурного навчання базується на спеціальній статистиці. Наведено попередні результати застосування методики, що демонструють можливість поліпшення якості результатів структурного навчання.*

***Ключові слова:** динамічні баєсівські мережі, структурне навчання динамічних баєсівських мереж.*

***Аннотация.** Рассматриваются некоторые существующие подходы и предлагается новая методика статистического структурного обучения динамических байесовских сетей для случая полной наблюдаемости данных. Предложенная методика структурного обучения базируется на специальной статистике. Приведены предварительные результаты применения методики, демонстрирующие возможность улучшения качества результатов структурного обучения.*

***Ключевые слова:** динамические байесовские сети, структурное обучение динамических байесовских сетей.*

***Abstract.** Some known approaches to structural learning of dynamic Bayesian networks are considered and a new methodology is proposed that is based on statistical data. The methodology proposed uses special statistic for structural learning. Preliminary results of application the methodology to real data is given that demonstrate improvement of structural learning quality.*

***Key words:** dynamic bayesian networks, structural learning of dynamic Bayesian networks.*

## 1. Введение

Хорошо известно, что фильтры Калмана и скрытые модели Маркова, являющиеся представителями простейших вероятностных направленных графических моделей, нашли обширное практическое применение и позволяют эффективно решать определенный класс задач [1]. Например, предварительная обработка данных, оптимальное оценивание состояний, краткосрочное прогнозирование, дуальное управление. К множеству вероятностных направленных графических моделей также принадлежат байесовские сети (БС) и динамические байесовские сети (ДБС), которые являются перспективным и эффективным инструментом решения различных задач, таких как интеллектуальный анализ данных, распознавание образов, классификация, прогнозирование и т.д. Описание успешного применения ДБС для распознавания речи приведено в работе [2], результаты прогнозирования состояния пациентов отделения интенсивной терапии описаны в работе [3]. Дальнейшее успешное применение ДБС требует более подробного исследования некоторых теоретических и практических аспектов их функционирования. В частности, более эффективного формирования вероятностного вывода на известной структуре сети, определения оптимальной структуры и параметров сети в различных условиях: полная или частичная наблюдаемость переменных, необходимость получения результатов в реальном времени, возможность влияния на входные данные. БС и ДБС являются на сегодня объектом активных исследований, что находит отражение в соответствующей научной литературе [1–9].

В сущности вероятностные направленные графические модели, такие как БС и ДБС, опираются на понятие условной вероятности и предоставляют средства наглядного и компактного представления вероятностных зависимостей и совместного распределения вероятности (СРВ) случайных величин, описывающих некоторое явление или процесс.

Одним из главных компонентов БС и ДБС является направленный ациклический граф, в котором некоторая дуга от вершины А к вершине В неформально может быть ин-

терпретирована как причинно-следственная связь между А и В по аналогии с классическим представлением о связи между понятиями причины и следствия.

Данная работа посвящена исследованию некоторых существующих подходов и разработке основ новых методик построения ДБС для случая полной наблюдаемости данных. Ниже будут изложены общие теоретические сведения о ДБС, рассмотрены существующие байесовские и предложены альтернативные статистические критерии качества сетевой структуры, приведены примеры использования рассмотренных методов и результаты экспериментальных исследований.

## 2. Постановка задачи

Рассмотреть существующие подходы к определению оптимальной, в смысле выбранного критерия, структуры ДБС на основе выборок данных, порожденных многомерным случайным стационарным марковским процессом первого порядка. Разработать и исследовать основы альтернативной статистической методики.

Формально задачу можно сформулировать следующим образом: известны многомерные временные выборки данных

$$D_v = \{(x_{1v}[0], \dots, x_{nv}[0]), \dots, (x_{1v}[T_v], \dots, x_{nv}[T_v])\},$$

где  $v$  – номер выборки,  $T_v$  – объем выборки с номером  $v$ ; множество всех выборок обозначено через  $D = \{D_1, \dots, D_V\}$ , где  $V$  – количество выборок. Выборки порождены упорядоченным множеством дискретных случайных величин  $\Psi_k = \{X_1[k], \dots, X_n[k]\}$ . Случайные величины  $X_i[k]$ ,  $X_i[m]$  для любых моментов времени  $k, m$  могут принимать значения из одного и того же множества значений  $Val(X_i)$ .

Необходимо найти оценку разложения СРВ, наиболее точно отражающую реальные зависимости, существующие в данных

$$P(\Psi_{k+1}, \Psi_k) = P(X_1[k+1], \dots, X_n[k+1], X_1[k], \dots, X_n[k]),$$

на произведение условных распределений вероятностей (УРВ):

$$P(\Psi_{k+1}, \Psi_k) = P(\Psi_k)P(\Psi_{k+1} / \Psi_k) = P(\Psi_k) \prod_{i=1}^n P(X_i[k+1] / Pa(X_i[k+1])),$$

где  $Pa(X_i[k+1]) \subseteq \Psi_k \cup \Psi_{k+1} \setminus \{X_i[k+1]\}$  – множество случайных величин, от которых условно зависит случайная величина  $X_i[k+1]$ . Предполагается, что для любых моментов времени  $k$  и  $m$  разложение СРВ  $P(\Psi_k, \Psi_{k-1})$ ,  $P(\Psi_m, \Psi_{m-1})$  на произведение УРВ одинаково (процесс стационарен) и подчиняется одним и тем же законам распределения вероятностей, то есть

$$P(\Psi_k / \Psi_{k-1}) = \prod_{i=1}^n P(X_i[k] / Pa(X_i[k])) = \prod_{i=1}^n P(X_i[m] / Pa(X_i[m])) = P(\Psi_m / \Psi_{m-1}).$$

## 3. Динамическая байесовская сеть

ДБС – математический аппарат компактного представления сложных стохастических процессов. ДБС – это расширение обычной (статической) БС, благодаря чему они позволяют исследовать и моделировать конечные последовательности множеств случайных величин  $\Psi_k = \{X_1[k], \dots, X_n[k]\}$ , в отличие от одного множества случайных величин  $\Psi = \{X_1, \dots, X_n\}$ , описываемого БС.

Более формально ДБС представляет собой два направленных ациклических графа  $G_{\rightarrow}, G_0$ , кодирующих разложения распределений вероятности  $P(\Psi_{k+1} / \Psi_k), P(\Psi_0)$  на УРВ

и заданные некоторым способом множества соответствующих УРВ  $\Theta_{\rightarrow}, \Theta_0$ . Каждая вершина любого из указанных ациклических графов соответствует некоторой случайной величине  $X_i[k]$ . Графы  $G_{\rightarrow}, G_0$  состоят из вершин, соответствующих множествам случайных величин:

$$\Psi_{k+1} \cup \Psi_k = \{X_1[k+1], \dots, X_n[k+1], X_1[k], \dots, X_n[k]\},$$

$$\Psi_0 = \{X_1[0], \dots, X_n[0]\}.$$

Если  $X_j[m] \in Pa(X_i[k], \{G_{\rightarrow}, G_0\})$ , то в графе  $G_0$  при  $m = k = 0$  или  $G_{\rightarrow}$  при  $k = m + 1$  имеется направленная дуга  $X_j[m] \rightarrow X_i[k]$ . Так как БС  $B$  представляет собой аналогичным образом сконструированный граф  $G$  и множество  $\Theta$  [1 – 3], то самым компактным определением ДБС, предложенным Мерфи, будет следующее.

**Определение 2.1.** ДБС представляет собой пару  $(B_0, B_{\rightarrow})$ , где  $B_0$  – БС, описывающая СРВ  $P(\Psi_0)$ ,  $B_{\rightarrow}$  – двухслойная БС, описывающая УРВ  $P(\Psi_{k+1}/\Psi_k)$  [1].

Пример возможных графов  $G_0, G_{\rightarrow}$  приведен на рис. 1.

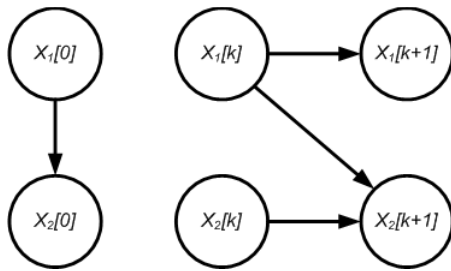


Рис. 1. Графы  $G_0, G_{\rightarrow}$  для ДБС

Для графов  $G_0, G_{\rightarrow}$ , приведенных на рис. 1, разложения СРВ  $P(\Psi_0)$  и УРВ  $P(\Psi_{k+1}/\Psi_k)$  имеют вид

$$P(\Psi_0) = P(X_2[0]/X_1[0])P(X_1[0]),$$

$$P(\Psi_{k+1}/\Psi_k) = P(X_2[k+1]/X_1[k], X_2[k])P(X_1[k+1]/X_1[k]).$$

Таким образом, задачу данной работы кратко можно сформулировать следующим образом: необходимо определить структуру графа ДБС, наиболее точно отражающую фактически существующие зависимости в данных. В литературе эту задачу также называют структурным обучением ДБС [2].

#### 4. Байесовский подход к структурному обучению ДБС

Суть байесовского подхода к структурному обучению ДБС основывается на применении правила Байеса для УРВ структур графов множества  $G = \{G_0, G_{\rightarrow}\}$  в зависимости от типа выборки данных  $D$ :

$$P(G/D) = \frac{P(G, D)}{\sum_G P(G, D)} = \frac{P(G, D)}{P(D)}.$$

Таким образом, при сравнении двух структур  $G_1, G_2$  можно перейти от сравнения условных вероятностей к сравнению безусловных:

$$\frac{P(G_1/D)}{P(G_2/D)} = \frac{P(G_1, D)/P(D)}{P(G_2, D)/P(D)} = \frac{P(G_1, D)}{P(G_2, D)}.$$

Для оценок величины  $P(G, D)$  Купером и Гершковичем в работе [6] предложено аналитическое выражение, но его практическое использование часто приводит к непреодолимым вычислительным трудностям даже для структурного обучения БС с небольшими количествами вершин. В работах [4, 7] предлагается хорошая асимптотическая оценка  $L(G, D)$  величины  $\log(P(G, D))$ , пригодная не только для прикладного структурного обу-

чения БС, но и ДБС.  $L(G, D)$  принято называть байесовским информационным критерием, который определяется выражением

$$L(G, D) = \log(P(D/\hat{\Theta}, G)) - \frac{1}{2} \sigma(G) \log \sum_{v=1}^V T_v,$$

где  $\hat{\Theta}$  – оценки параметров распределений вероятности для графов множества  $G$ , полученные на основе выборки  $D$  методом максимального правдоподобия;  $\sigma(G)$  – функция, характеризующая сложность зависимостей между случайными величинами на графе  $G$ . Для ДБС справедливо равенство:  $\sigma(G) = \sigma_0(G) + \sigma_{\rightarrow}(G)$ , где  $\sigma_0(G)$  и  $\sigma_{\rightarrow}(G)$  характеризуют сложность зависимостей для  $G_0$  и  $G_{\rightarrow}$  соответственно. Функция  $\sigma_0(G)$  имеет вид [4]

$$\sigma_0(G) = \sum_i \sum_{par \in Pa(X_i[0])} Val(par) \cdot (Val(X_i[0]) - 1).$$

Выражение для функции  $\sigma_{\rightarrow}(G)$  имеет аналогичный вид. Для ДБС  $P(D/\hat{\Theta}, G)$  имеет разложение вида

$$P(D/\hat{\Theta}, G) = \prod_{v=1}^V P(D_v/\hat{\Theta}, G) = \prod_{v=1}^V \prod_{i=1}^n P(X_i[0] = x_{iv}[0] / Pa(X_i[0], G) = r_{iv}[0], \hat{\Theta}) \times \\ \times \prod_{k=0}^{T_v-1} P(X_i[k+1] = x_{iv}[k+1] / Pa(X_i[k], G) = r_{iv}[k], \hat{\Theta}),$$

где  $r_{iv}[k]$  – подмножество  $\{x_{1v}[k], \dots, x_{nv}[k]\}$ , соответствующее реализации случайных величин в  $k$ -й момент времени, от которых условно зависит  $X_i[k+1]$  на графе  $G$ . Введем обозначения для упрощения дальнейших записей:

$$\hat{\theta}_{i, x_i, r_i}(G) = P(X_i[0] = x_i / Pa(X_i[0], G) = r_i, \hat{\Theta}), \\ \hat{\theta}_{i, x'_i, r'_i}^{\rightarrow}(G) = P(X_i[k+1] = x'_i / Pa(X_i[k], G) = r'_i, \hat{\Theta}), \\ N_{i, x_i, r_i}(G) = \sum_{v=1}^V I(X_i[0] = x_i, Pa(X_i[0], G) = r_i | D_v), \\ N_{i, x'_i, r'_i}^{\rightarrow}(G) = \sum_{v=1}^V \sum_{k=0}^{T_v-1} I(X_i[k+1] = x'_i, Pa(X_i[k], G) = r'_i | D_v),$$

где  $I(\bullet | D_v)$  – индикатор события « $\bullet$ » в выборке  $D_v$ ; переменные  $x_i, x'_i$  могут принимать все возможные значения случайной величины  $X_i[k]$ ;  $r_i, r'_i$  могут принимать все возможные значения множества случайных величин  $Pa(X_i[k], G)$ . На основании введенных выше обозначений и указанного разложения  $P(D/\hat{\Theta}, G)$  выражение для  $\log(P(D/\hat{\Theta}, G))$  можно переписать в более компактной и удобной для вычислений форме:

$$\log(P(D/\hat{\Theta}, G)) = \sum_{i=1}^n \sum_{x_i} \sum_{r_i} N_{i, x_i, r_i}(G) \hat{\theta}_{i, x_i, r_i}(G) + \sum_{i=1}^n \sum_{x'_i} \sum_{r'_i} N_{i, x'_i, r'_i}^{\rightarrow}(G) \hat{\theta}_{i, x'_i, r'_i}^{\rightarrow}(G).$$

Оценка для  $\hat{\theta}_{i, x_i, r_i}(G)$  вычисляется методом максимального правдоподобия:

$$\hat{\theta}_{i, x_i, r_i}(G) = \frac{N_{i, x_i, r_i}(G)}{\sum_{x_i} N_{i, x_i, r_i}(G)};$$

для  $\hat{\theta}_{i, x'_i, r'_i}^{\rightarrow}(G)$  это выполняется аналогично.

Таким образом, для решения поставленной задачи необходимо найти такие структуры графов множества  $G^*$ , которые максимизируют байесовский информационный критерий:

$$G^* = \arg\{\max_{G \in \Gamma} L(G, D)\},$$

где  $\Gamma$  – множество графов-претендентов. Особенностью байесовского информационного критерия для ДБС является его аддитивность по отношению к свойствам БС, составляющих ДБС [1, 4]. То есть, структурное обучение может быть выполнено отдельно для  $B_0$  и  $B_{\rightarrow}$ .

Алгоритм поиска  $B_{\rightarrow}$ , максимизирующий байесовский информационный критерий для случая, когда множество  $\Gamma$  содержит только такие  $G_{\rightarrow}$ , для которых выполняется условие  $Pa(X_i[k+1]) \subseteq \Psi_k$  (то есть, связи внутри слоев  $B_{\rightarrow}$  отсутствуют), схематически представлен на рис. 2.

for  $i \in \{1, \dots, n\}$

  for  $\Psi^\circ \in 2^{\mathcal{Y}_k}$

$g^{\rightarrow} = \{X_j[k] \rightarrow X_i[k+1] | X_j[k] \in \Psi^\circ\}$

$\{\hat{\theta}_{i,x'_i,r'}^{\rightarrow}(g^{\rightarrow})\} = \left\{ \frac{N_{i,x'_i,r'}^{\rightarrow}(g^{\rightarrow})}{\sum_{x_i} N_{i,x'_i,r'}^{\rightarrow}(g^{\rightarrow})} \right\}$

    if  $L(g^{\rightarrow}, D) > L(g_i^*, D)$  then  $g_i^* = g^{\rightarrow}$

$G_{\rightarrow}^* = \bigcup_i g_i^*$

Рис. 2. Алгоритм поиска наилучшего  $G_{\rightarrow}^*$  на основе  $L(G, D)$

## 5. Статистическая методика структурного обучения ДБС

В рамках классического статистического подхода задача данной работы может быть рассмотрена как задача проверки гипотез  $H_{g_m}$ ,  $g_m \in \Gamma$  о структуре графов  $G$  ДБС. Без существенной потери общности результатов исследования ограничимся проверкой гипотез  $H_{g_m^{\rightarrow}}$  о структуре графа  $G_{\rightarrow}$  ДБС, так как решение этой подзадачи является основным вопросом структурного обучения ДБС. Для проверки вышеуказанных гипотез необходима подходящим образом сконструированная статистика  $S$ .

Рассмотрим свойства выборочной оценки УРВ  $\hat{P}(X_i[k+1]/\Psi_k^\circ)$ . Предположим, что случайная величина  $X_i[k+1]$  находится в вероятностной зависимости от случайных величин  $Pa(X_i[k+1]) = \Psi_k^* \subseteq \Psi_k^\circ$ ; тогда очевидно, что в случае  $\Psi_k^\circ \neq \Psi_k^*, \Psi_k^* \subset \Psi_k^\circ$  оценка  $\hat{P}(X_i[k+1]/\Psi_k^\circ)$  будет содержать повторяющуюся информацию об УРВ  $P(X_i[k+1]/\Psi_k^*)$  в отличие от оценки  $\hat{P}(X_i[k+1]/\Psi_k^*)$ . Можно утверждать, что при любых фиксированных значениях  $q'_i \in Val(\Psi_k \setminus \Psi_k^*)$  для случайных величин  $\Psi_k \setminus \Psi_k^*$ , от которых не зависит  $X_i[k+1]$ , будет иметь место сходимость по вероятности:

$$\forall q'_i, \hat{P}(X_i[k+1]/\Psi_k^*, \{\Psi_k \setminus \Psi_k^*\} = q'_i) \xrightarrow{p} P(X_i[k+1]/\Psi_k^*).$$

Однако, если  $\Psi_k^* \neq Pa(X_i[k+1])$ , сходимости не будет. Отсюда следует, что в результате сравнения выборочных законов распределения вероятностей  $\hat{F}_i(X, \Psi_k^\circ, q'_i)$  и  $\hat{F}_i(X, \Psi_k^\circ)$ , полученных на основе оценок распределений вероятностей  $\hat{P}(X_i[k+1]/\Psi_k^\circ, \{\Psi_k \setminus \Psi_k^\circ\} = q'_i)$  и  $\hat{P}(X_i[k+1]/\Psi_k^\circ)$  для различных  $q'_i$ , можно судить о наличии или отсутствии указанной сходимости для множества кандидатов  $\Psi_k^\circ \subset \Psi_k$ . Проверку гипотез об одинаковости выборочных законов распределения вероятности  $\hat{F}_i(X, \Psi_k^\circ, q'_i)$  и  $\hat{F}_i(X, \Psi_k^\circ)$  можно осуществить на основе критерия согласия Колмогорова.

Исходя из указанных свойств  $\hat{F}_i(X, \Psi_k, q'_i)$  и  $\hat{F}_i^\circ(X, \Psi_k)$ , для определения наилучшего множества  $\Psi_k$  для случайной величины  $X_i[k+1]$  в качестве подходящей статистики, основанной на критерии согласия Колмогорова, может быть рассмотрена величина

$$S_i(g_m^{\rightarrow}) = \sqrt{\frac{1}{2} \sum_v T_v \sum_{\Psi_k(g_m^{\rightarrow})} \sum_{q'_i} \sup_X |\hat{F}_i(X, \Psi_k(g_m^{\rightarrow}), q'_i) - \hat{F}_i^\circ(X, \Psi_k(g_m^{\rightarrow}))|},$$

где  $\Psi_k(g_m^{\rightarrow})$  – множество вершин родителей для  $X_i[k+1]$  в графе  $g_m^{\rightarrow}$ . Очевидно, что чем ближе  $\hat{F}_i(X, \Psi_k, q'_i)$  к  $\hat{F}_i^\circ(X, \Psi_k)$  для любого  $q'_i$  и  $\Psi_k$ , тем меньше будет значение статистики, и наоборот. Следовательно, решение рассматриваемой задачи на основе  $S(g_m^{\rightarrow}) = \sum_i S_i(g_m^{\rightarrow})$  можно записать:

$$G_{\rightarrow}^* = \arg\{\min_{g_m^{\rightarrow} \in \Gamma^{\rightarrow}} S(g_m^{\rightarrow})\},$$

где  $\Gamma^{\rightarrow}$  – множество графов-претендентов, определяющих связи между слоями подсети  $B_{\rightarrow}$  ДБС. Следует отметить, что в случае, когда  $\Gamma^{\rightarrow}$  содержит графы с различной степенью связности вершин, указанная статистика  $S(g_m^{\rightarrow})$  будет принимать минимальные значения для самых «сложных графов», то есть для графов с максимальным количеством связей. Однако, если  $\Gamma^{\rightarrow}$  содержит структуры «одинаковой сложности» с одинаковым количеством связей, использование этой статистики позволит найти самый лучший в смысле постановки задачи граф  $G_{\rightarrow}$ . Чтобы снять указанное ограничение на множество  $\Gamma^{\rightarrow}$ , в решении задачи необходимо учесть величину, аналогичную второй части выражения байесовского информационного критерия, «штрафующую» метод за сложность выбранного графа. Но для данного исследования это ограничение не существенно, так как позволяет в полной мере изучить принципиальные свойства статистики  $S(g_m^{\rightarrow})$ .

```

for  $i \in \{1, \dots, n\}$ 
  for  $\Psi^\circ \in 2^{\Psi_k}$ 
    if  $\text{Card}(\Psi^\circ) \neq R$  continue
     $g^{\rightarrow} = \{X_j[k] \rightarrow X_i[k+1] | X_j[k] \in \Psi^\circ\}$ 
    if  $S_i(g^{\rightarrow}) < S_i(g_i^*)$  then  $g_i^* = g^{\rightarrow}$ 
 $G_{\rightarrow}^* = \bigcup_i g_i^*$ 

```

Рис. 3. Алгоритм поиска наилучшего  $G_{\rightarrow}^*$  на основе  $S(g_m^{\rightarrow})$

Алгоритм поиска сети  $B_{\rightarrow}$ , минимизирующий статистику  $S$  для случая, когда множество  $\Gamma^{\rightarrow}$  содержит только такие графы  $G_{\rightarrow}$ , для которых выполняется условие  $Pa(X_i[k+1]) \subseteq \Psi_k$  и  $\text{Card}(Pa(X_i[k+1])) = R$ , где  $R$  подается на вход алгоритма, схематически представлен на рис. 3.

В качестве примера рассчитаем значение  $S_1(g_{2,1}^{\rightarrow})$  для графа, приведенного на рис. 1 и одной временной выборки данных из табл. 1. Оценки условных распределений приведены в табл. 2 и 3.

Таблица 1. Выборочные данные

$k$	1	2	3	4	5	6
$X_1[k]$	0	0	1	1	0	1
$X_2[k]$	0	1	0	1	1	0

Таблица 2. Оценки условных распределений  $\hat{P}(X_1[k+1]/X_1[k], X_2[k])$

$X_2[k]$	$X_1[k]$	$X_1[k+1]$	$\hat{P}(X_1[k+1]/X_1[k], X_2[k])$	$\hat{F}_1(X, \Psi_k^\circ = X_1[k], q'_i = X_2[k])$
0	0	0	1	1
0	0	1	0	1
0	1	0	0	0
0	1	1	1	1
1	0	0	0	0
1	0	1	1	1
1	1	0	1	1
1	1	1	0	1

Таблица 3. Оценки условных распределений  $\hat{P}(X_1[k+1]/X_1[k])$

$X_1[k]$	$X_1[k+1]$	$\hat{P}(X_1[k+1]/X_1[k])$	$\hat{F}_1(X, \Psi_k^\circ = X_1[k])$
0	0	1/3	1/3
0	1	2/3	1
1	0	1/2	1/2
1	1	1/2	1

Расчет значения  $S_1(g_{2.1}^{\rightarrow})$ :

$$\begin{aligned}
 S_1(g_{2.1}^{\rightarrow}) &= \sqrt{3} \left[ \sup \left( \left| 1 - \frac{1}{3} \right|, |1-1| \right) + \sup \left( \left| 0 - \frac{1}{2} \right|, |1-1| \right) + \sup \left( \left| 0 - \frac{1}{3} \right|, |1-1| \right) + \sup \left( \left| 1 - \frac{1}{2} \right|, |1-1| \right) \right] = \\
 &= \sqrt{3} \left[ \frac{2}{3} + \frac{1}{2} + \frac{1}{3} + \frac{1}{2} \right] = 3,46.
 \end{aligned}$$

## 6. Экспериментальные исследования

Экспериментальные исследования выполнены методом имитационного моделирования. В основу критерия качества структурного обучения положена мера Хэмминга  $H(G_{\rightarrow}, G_{\rightarrow}^*)$ , позволяющая сравнивать граф  $G_{\rightarrow}$  исходной ДБС с графом  $G_{\rightarrow}^*$  обученной сети. Указанная мера для двух графов с одинаковыми множествами вершин будет тем большей, чем больше отличий в соответствующих множествах ветвей.

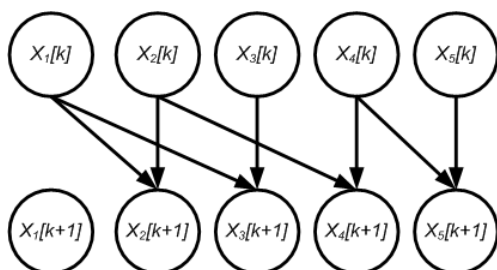


Рис. 4. Граф  $G_{\rightarrow}$  ДБС для экспериментальных исследований

В качестве ДБС для проведения исследований выбрана сеть со структурой, представленной на рис. 4.

Эксперимент организован следующим образом. На вход подавалось количество временных выборок  $V$  и количество элементов множества значений случайных величин  $Z$ ; без ограничения общности полагалось, что  $\forall i: \text{Card}(\text{Val}(X_i)) = Z$ .

Чтобы усреднить полученные значения критерия качества структурного обучения, для каждого выбранного  $V$  и  $Z$  выполнено 5 элементарных экспериментов. Каждый элементарный эксперимент состоял из следующих этапов:

Каждый элементарный эксперимент состоял из следующих этапов:

1. На основе равномерного распределения вероятностей генерировались случайные табличные распределения условных вероятностей, описывающих связи между слоями выбранной ДБС.

2. На основе метода Монте-Карло сгенерированы  $V$  временных выборок данных мощностью  $T_b = 30$  элементов, в каждой из которых любая случайная величина могла принимать не больше, чем  $Z$  различных значений.

3. Выполнялось структурное обучение на основе байесовского информационного критерия и статистического критерия. Для уменьшения вычислительной сложности и учета ограничений статистической методике множество графов претендентов  $\Gamma^{\rightarrow}$  ограничено следующим условием:  $Card(Pa(X_i[k+1])) = 2$  для обоих способов.

4. Вычислялось значение критерия качества  $H(G_{\rightarrow}, G_{\rightarrow}^*)$  структурного обучения для обоих способов.

Экспериментальные исследования направлены на изучение характера поведения эмпирических функций:

$$K^{BIC}(V, Z) = \frac{1}{5} \sum_{E=1}^5 H(G_{\rightarrow}, G_{\rightarrow}^{BIC}(E, V, Z)),$$

$$K^S(V, Z) = \frac{1}{5} \sum_{E=1}^5 H(G_{\rightarrow}, G_{\rightarrow}^S(E, V, Z)),$$

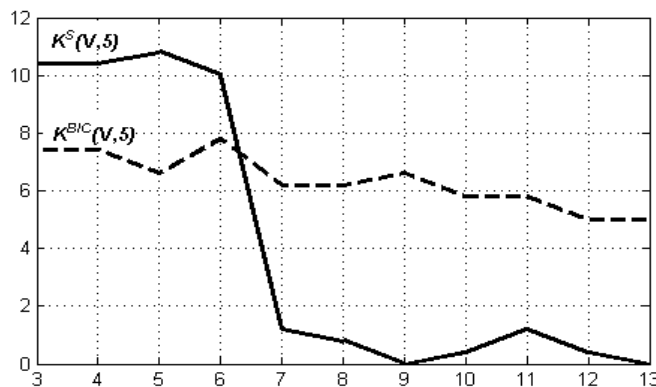


Рис. 5. Графики:  $K^{BIC}(V, 5)$  – штриховая линия и  $K^S(V, 5)$  – сплошная линия

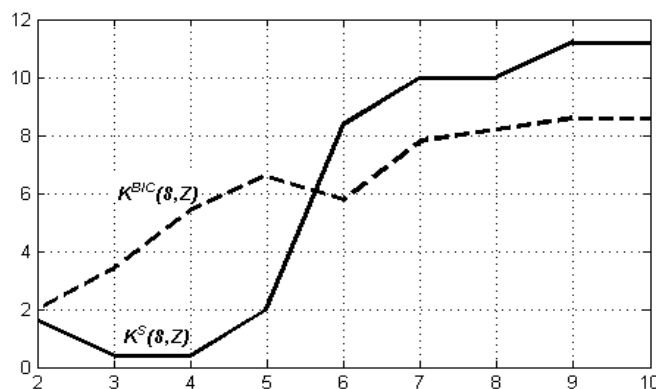


Рис. 6. Графики для  $K^{BIC}(8, Z)$  – штриховая линия и  $K^S(8, Z)$  – сплошная линия

где  $G_{\rightarrow}^{BIC}(E, V, Z), G_{\rightarrow}^S(E, V, Z)$  – графы ДБС, полученные в результате структурного обучения, выполненного по  $V$  временным выборкам с  $Z$  элементами множества значений случайных величин;  $E$  – номер элементарного эксперимента. Эти графы получены на основе байесовского информационного критерия и статистической методике соответственно.

Зависимости  $K^{BIC}(V, Z)$  и  $K^S(V, Z)$  при  $Z=5$  и  $V=\overline{3, 13}$ , полученные в результате проведенных экспериментов, отражены на рис. 5.

Из анализа полученных результатов для  $Z=5$  и  $V=\overline{3, 13}$  следует, что при фиксированном  $Z$  структурное обучение, основанное на предложенной статистической методике, приводит к существенно лучшим результатам при больших  $V$ . Аналогичный характер поведения  $K^{BIC}(V, Z)$  и  $K^S(V, Z)$  получен и для других  $Z$ .

Зависимости  $K^{BIC}(V, Z)$  и  $K^S(V, Z)$  при  $Z=\overline{2, 10}$  и  $V=8$ , полученные в результате выполненных экспериментов, отражены на рис. 6.



Из анализа полученных результатов для  $Z = \overline{2,10}$  и  $V = 8$  следует, что при фиксированном  $V$  структурное обучение, основанное на предложенной статистической методике, приводит к более качественным результатам при меньших  $Z$ . Аналогичный характер поведения  $K^{BIC}(V, Z)$  и  $K^S(V, Z)$  был получен и для других  $V$ .

## 7. Выводы

В работе подробно рассмотрена задача структурного обучения ДБС с позиций байесовского подхода и теории вероятностей. Рассмотрен существующий алгоритм структурного обучения на основе байесовского информационного критерия и разработаны основы методики структурного обучения на базе предложенной статистики.

Выполнен анализ полученных экспериментальных зависимостей, показывающий, что при определенных соотношениях количества временных выборок  $V$  и количества элементов множества значений случайных величин  $Z$  структурное обучение, основанное на предложенной статистической методике, позволяет существенно повысить качество результатов обучения ДБС. Отметим, что, как следует из экспериментальных данных, качество обучения на основе байесовского информационного критерия и предложенной статистической методики обладает противоположными свойствами по отношению к  $Z$  и  $V$ .

В дальнейших исследованиях возможна доработка предложенной методики обучения с целью снятия ограничения на множество графов-претендентов  $\Gamma^{\rightarrow}$  путем учета в методике штрафа за сложность графа. Также необходимо определить в общем виде условия по  $Z$  и  $V$ , при которых для достижения максимального качества структурного обучения необходимо использовать байесовский информационный критерий или предложенную статистику.

## СПИСОК ЛИТЕРАТУРЫ

1. Murphy K.P. Dynamic Bayesian networks: representation, inference and learning: a PhD dissertation / Murphy K.P. – University of California, Berkeley, 2002. – 225 p.
2. Zweig G.G. Speech Recognition with Dynamic Bayesian Networks: a PhD dissertation / Zweig G.G. – University of California, Berkeley, 1998. – 169 p.
3. Kayaalp M.M. Learning Dynamic Bayesian Network Structures from Data: Ph.D. Dissertation / M.M. Kayaalp, G.F. Cooper. – University of Pittsburgh, 2003. – 203 p.
4. Friedman F.N. Learning Structure of Dynamic Probabilistic Networks / F.N. Friedman, K. Murphy, S. Russell // Proc. of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI'98). – 1998. – P. 139 – 147.
5. Тулупьев А.Л. Байесовские сети: логико-вероятностный подход / Тулупьев А.Л., Николенко С.И., Сироткин А.В. – СПб.: Наука, 2006. – 607 с.
6. Cooper G. A Bayesian method for the induction of probabilistic networks from data / G. Cooper, E. Herskovits // Machine Learning. – 1992. – N 9. – P. 309 – 347.
7. Heckerman D. Bayesian Networks for Data Mining / D. Heckerman // Data Mining and Knowledge Discovery. – 1997. – N 1. – P. 79 – 119.
8. Згуровский М.З. Методы построения байесовских сетей на основе оценочных функций / М.З. Згуровский, П.И. Бидюк, А.Н. Терентьев // Кибернетика и системный анализ. – 2008. – № 2. – С. 81 – 88.
9. Терентьев А.Н. Методы построения байесовских сетей / А.Н. Терентьев, П.И. Бидюк // Межведомственный научно-технический сборник „Адаптивные системы автоматического управления”. – Днепропетровск: Системные технологии, 2005. – № 8. – С. 130 – 141.

*Стаття надійшла до редакції 22.01.2010*