



УДК 519.85

© 2022 г. **П.В. Полухин**, канд. техн. наук  
(Воронежский государственный университет)

## РАЗРАБОТКА ОПТИМАЛЬНЫХ СТОХАСТИЧЕСКИХ МЕТОДОВ РЕШЕНИЯ ЗАДАЧ ФИЛЬТРАЦИИ СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ НА ОСНОВЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ

Область применения алгоритмов многочастичных фильтров является достаточно обширной. Они могут быть использованы для решения вероятностных задач в большинстве графических моделей, вершины которых обладают вероятностными характеристиками. Простота реализации и большой потенциал масштабирования данных фильтров делают их универсальным инструментом в процессе решения вероятностных задач. Однако существует ряд недостатков данного алгоритма, связанных с тем, что точность алгоритма пропорциональна объему генерируемых выборок, согласованность которых определяется на этапе повторного взвешивания в соответствии со свидетельствами, поступающими на каждом шаге выполнения алгоритма. Для решения данных проблем в научном исследовании предлагается оптимизация алгоритма многочастичного фильтра за счет использования генетических алгоритмов в процессе формирования выборок. Такой подход позволяет сузить область разброса выборок, а также повысить степень их согласованности.

**Ключевые слова:** многочастичный фильтр, метод Монте-Карло, цепь Маркова, скрытая марковская модель, генетический алгоритм.

DOI: 10.22250/18142400\_2022\_73\_3\_13

### Введение

Генетические алгоритмы представляют собой универсальный инструмент, основанный на представлении естественного процесса эволюции. Применение генетических алгоритмов к математическим моделям было впервые показано в работах Холланда [1]. В его работах рассматривается применение теории естественного отбора Дарвина для решения ряда прикладных математических задач поиска, распознавания образов и статистического вывода. Согласно его предположению, каждой модели может быть

противопоставлено ограниченное число структур  $\alpha$ , представляющих собой некоторое множество хромосом (комбинация генов), изменение которых производит за счет мутации. В таком случае общая сложность такой модели будет определяться общим числом генотипов (структура генов). Генетические алгоритмы с научно-практической точки зрения представляют особый интерес для оптимизации процедур обучения и логического вывода вероятностных моделей. Одним из основных алгоритмов решения данных задач является семейство алгоритмов на основе многочастичного фильтра. Решение задачи нелинейной фильтрации на основе генетических алгоритмов было рассмотрено в исследованиях Морала [2, 3]. Суть предложенного подхода заключается в случайной мутации выборок, формируемых в соответствии с переходными вероятностями марковской цепи. В случае динамических вероятностных моделей данный подход не может быть явно применен, так как выборки, подвергнутые мутации, предварительно должны соответствовать вероятностному распределению по всем свидетельствам. В связи с этим возникает необходимость пошаговой фильтрации с последующим применением генетического алгоритма уже к сформированному весовому распределению, полученному в соответствии с распределениями переходов и свидетельств для момента времени  $t + k$ . Основная особенность предлагаемого подхода заключается в возможности скрещивания выборок, исключительно имеющих общие свидетельства, а следовательно, переменные, не имеющие условных связей, будут исключены из генетического алгоритма. В таком случае можно снизить область разброса выборок и повысить степень их согласования со свидетельствами. Для динамических моделей также допустимо использовать свидетельства всех предыдущих состояний. Следовательно, можно прийти к выводу, что генетические алгоритмы также могут быть использованы в процессе ретроспективного анализа таких моделей, при этом общее число генерируемых выборок для обеспечения требуемой точности может быть сокращено.

### **Материалы и методы**

Задачи вероятностного вывода являются одними из основных решаемых с помощью вероятностных моделей. Наибольший интерес представляют модели, состоящие из нескольких временных состояний. Среди таких моделей выделяют динамические байесовские (ДБС) сети и скрытые марковские модели (СММ). В исследовании рассмотрим СММ, так как в общем случае ДБС можно представить в виде СММ, являющихся графическим представлением марковского процесса, где каждому срезу соответствует модель пе-

рехода  $P(X_1 | X_2)$  и модель восприятия  $P(E_2 | X_2)$ . Каждому узлу СММ ставится в соответствие таблица условных вероятностей, характеризующая связи в рамках как одного, так и нескольких срезов. Типовая структура модели СММ приведена на рис.1.

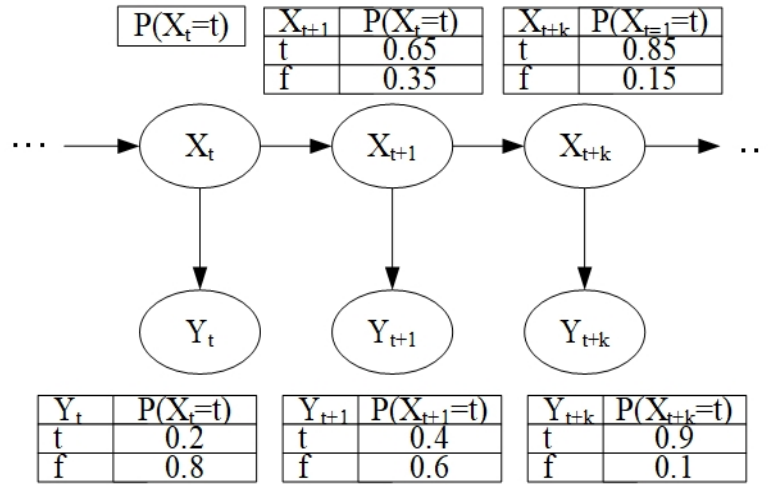


Рис. 1. Структура скрытой марковской модели из трех временных срезов.

Матрица переходов, соответствующая рис. 1, будет иметь следующий вид:

$$T = P(X_{t+1} | X_t) = \begin{pmatrix} T_{11} & T_{12} & \cdots & T_{1k} \\ T_{21} & T_{22} & \cdots & T_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ T_{n1} & T_{n2} & \cdots & T_{nk} \end{pmatrix}, \quad (1)$$

$$T_{ij} = P(X_{t+1} = j | X_t = i).$$

Тогда полное совместное распределение, соответствующее СММ, можно записать в следующем виде [4, 5]:

$$\begin{aligned} P(X, Y) &= P(X_0) \prod_{t=1}^k P(X_{t+1} | X_t) \prod_{j=1}^k P(X_{j+1} | X_{j+1}) = \\ &= P(X_0) \prod_{t=1}^k T_{ij} \prod_{j=1}^k P(X_{j+1} | X_{j+1}), \end{aligned} \quad (2)$$

где  $P(X_0)$  – начальное распределение вероятностей, соответствующее моменту  $t = 0$ .

Определив вероятностную семантику СММ, перейдем к рассмотрению алгоритмов вероятностного вывода. Среди различных подходов к решению задачи вероятностного вывода наибольший интересе представляет многочастичный фильтр (МЧФ). Сущность МЧФ заключается в формировании множества независимых друг от друга выборок с соответствующими им весами  $W(X, Y, E)$ . Отметим, что  $E \subset Y$  – свидетельства, поступающие в виде не-

прерывного потока, вплоть до текущего состояния. В данном случае вес будет определять степень согласованности выборок  $S$ . В таком случае условие согласования выборок  $S$  со свидетельством  $E$  запишем в следующем виде [6]:

$$\begin{aligned}
 P(X_{t+1}|E_{1:t+1}) &= \sum_Y N(X_{t+1}, Y_{1:t}, E_{1:t}) W(X_{t+1}, Y_{1:t}, E_{1:t}) \approx \\
 &\approx \sum_Y S_{ws}(X_{t+1}, Y_{1:t}, E_{1:t}) W(X_{t+1}, Y_{1:t}, E_{1:t}) = \\
 &= \sum_Y P(X_{t+1}, Y_{1:t}, E_{1:t}) = P(X_{t+1}|E_{1:t+1}).
 \end{aligned}
 \tag{3}$$

Существует несколько основных подходов к решению задач логического вывода на основе МЧФ. Наибольший интересе представляют МЧФ на основе выборки по значимости (ВЗ) и взвешивания на основе правдоподобия (ВП). В данном исследовании рассмотрим МЧФ с взвешиванием на основе правдоподобия. Типовая схема фильтра МЧФ с ВП приведена на рис. 2.



Рис. 2. Типовая схема МЧФ-фильтра с ВП.

На рис. 2  $W(X_{t+1} | E_{1:t+1})$  – веса выборок  $S$ ,  $P(E_{t+1}|X_{t+1})$  и  $P(X_{t+1}|X_t)$  – модели перехода и восприятия. Рассматривая метод вероятностной оценки выборок на основе ВП, отметим главное его преимущество по сравнению с ВЗ. ВП исключает формирование приближенного распределения по значимости  $Q(X_t | X_{0:t-1}, E_{1:t})$ . Следовательно, нет необходимости на каждом этапе верификации апостериорного распределения  $P(X_{t+1}|E_{t+1})$  производить вычисление дистанции Кульбака-Лейблера (КЛ) для оценки степени соответствия  $Q(X_{t+1} | E_{t+1})$  и  $P(X_{t+1}|E_{t+1})$ . Вместо этого в ВП каждая выборка формируется в соответствии с весом  $X_{t+1}^i \sim W(X_{t+1} | X_{1:t}, E_{1:t+1})$ , при этом выборки с наименьшим весом исключаются из апостериорного распределения. Следовательно, согласно алгоритму ВП степень правдоподобия выборки будет

определяться в соответствии с ее весом  $W(X_{t+1} | X_{1:t}, E_{1:t+1})$ . С учетом того, что распределение по всем выборкам прямо пропорционально общему числу выборок  $N$ , запишем распределение для выборок в соответствии с распределением  $P(X_t | E_{1:t})$

$$S(X_t | E_{1:t}) = N \times P(X_t | E_{1:t}), N \rightarrow \infty. \quad (4)$$

Используя переходное распределение вероятностей  $P(X_{t+1} | X_t)$  для модели СММ, приведенной на рис. 1, можно определить условное распределение по всем выборкам, достигшим состояния  $t+1$  с учетом распределения (5). Тогда получим следующую сумму [7, 8]

$$S(X_{t+1} | E_{1:t}) = \sum_{X_t} P(X_{t+1} | X_t) S(X_t | E_{1:t}). \quad (5)$$

Далее используем метод ВП для определения весов для каждой из выборок в соответствии со свидетельствами  $E_{1:t}$

$$W(X_{t+1} | E_{1:t+1}) = P(E_{t+1} | X_{t+1}) S(X_{t+1} | E_{1:t}). \quad (6)$$

Для повышения доли согласованных выборок используется процедура повторной генерации выборок в соответствии с распределением  $W(X_{t+1} | E_{1:t+1})$ . Каждая следующая выборка формируется случайным образом. Вероятность того, что будет выбрана конкретная выборка, пропорциональна весу. С учетом этого, распределение по всем выборкам можно записать в следующем окончательном виде [9, 10]:

$$S(X_{t+1} | E_{1:t+1}) = N \times W(X_{t+1} | E_{1:t+1}) = N \times P(X_{t+1} | E_{1:t+1}). \quad (7)$$

МЧФ с ВП, описанный выше, является оптимальным алгоритмом вероятностного вывода, обладает высокой степенью параллелизма в связи с тем, что генерация выборок может производиться независимо. Однако, наряду с преимуществами, у алгоритма имеется недостаток, связанный с тем, что на этапе повторной генерации очередная выборка формируется случайно, тем самым не учитывается предыстория формирования выборок. Для решения данной проблемы в рамках алгоритма МЧФ предлагается использование генетических алгоритмов (ГА). Применительно к МЧФ алгоритм ГА может быть использован на этапе повторного формирования выборок с целью снижения разброса выборок относительно истинного значения и, как следствие, повышения точности определения апостериорного распределения  $P(X_{t+1} | E_{1:t+1})$ . Рассмотрим детально структуру генетического алгоритма, а также возможность его адаптации для решения задач вероятностного вывода в СММ на основе фильтра МЧФ. Основными этапами генетического алгоритма являются: инициализация, отбор, размножение и мутации. На этапе инициализации выбираем множество выборок, веса которых имеют

наибольшие значения. Тогда на этапе отбора доля потомков выборок с высоким значением весов будет преобладающей и вытеснит все другие выборки. Каждая следующая популяция будет формироваться путем изменения генотипа предшествующей выборки за счет выполнения процедуры мутации. Введем обобщенную формулировку ГА, предложенную Бэком [11,12] в соответствии с абстрактной моделью Холланда

$$G(s) = (S_0, n, n_k, \delta, \rho, \Omega, f, t), \quad (8)$$

где  $S_0 = (s_1, s_2, \dots, s_n) \in I^n$ ,  $I = \{0,1\}^{n_k}$  – начальная популяция частиц, полученная на этапе ВП;  $n$  – общее число популяций частиц;  $n_k$  – размер популяции, из которой производим отбор выборок;  $\delta$  – селекция  $I^{n_k} \rightarrow I^n$ ;  $\rho$  – скрещивание  $I^{n_k} \times I^{n_{k+1}} \rightarrow I^n$ ;  $\Omega$  – мутация  $I^n \rightarrow I^n$ ;  $f$  – функция соответствия  $I^n \rightarrow \mathbb{R}$ ;  $t$  – условие останова  $I^n \rightarrow \{0,1\}$ .

Начальная популяция  $S_0$  формируется в соответствии с  $P(E_{t+1}|X_{t+1})$  и  $P(X_{t+1}|X_t)$ , процедура оценки весов для каждой  $S_0^i$  выполняется в соответствии с алгоритмом ВП. В соответствии с формулой (8) процедура селекции может быть выполнена за счет выборки элементов из множества  $S_{t+1} = (s_1, s_2, \dots, s_n)$  в соответствии с равномерным распределением  $P(s): I \rightarrow \{0,1\}$ . Отметим, что в качестве функции соответствия будем использовать весовое распределение  $\omega = W(X_{t+1}|E_{1:t+1})$ , соответствующее множеству выборок  $S_{t+1}$ . В таком случае наилучший генотип, соответствующий выборке  $S_{t+1}$ , можно переделить в виде задача поиска экстремума  $\omega$

$$\hat{S}_{t+1}^i = \operatorname{argmax} W(X_{t+1}|E_{1:t+1}). \quad (9)$$

Распределение  $P(S_{t+1}^i)$  в общем виде можно выразить через функцию соответствия  $f(S_{t+1}^i) = \omega(S_{t+1}^i)$ . Имеем:

$$P(S_{t+1}^i) = \frac{\omega(S)}{\sum_{j=1}^n \omega(S_{t+1}^j)}. \quad (10)$$

Селекция, представленная выражением (10), предложена Холландом и является адаптивно-пропорциональной (АП). АП также называют методом рулетки. Данный подход имеет существенный недостаток, связанный с тем, что в результате выполнения АП-селекции можно получить отдельные выборки с высоким показателем функции соответствия  $\omega(s)$ . В таком случае процедура селекции может быть завершена, так и не достигнув своего оптимального состояния. В связи с этим в работе будем рассматривать метод

Бейкера, позволяющий использовать статистическое распределение для задания  $P(s_i)$ . Такой подход оптимален при решении вероятностных задач на основе МЧФ. Сформулируем селекцию Бейкера (линейное ранжирование) [13] в виде следующего выражения:

$$P(s_i) = \frac{1}{N} \left( a - (a-b) \frac{i-1}{N-1} \right), 1 \leq a \leq 2, b = 2 - a, \quad (11)$$

где  $i$  – порядковый номер особи в популяции частиц;  $N$  – общий размер популяции выборки, задаваемый на этапе инициализации алгоритма.

Отметим, что параметр  $a$  выбирается случайным образом в соответствии с равномерным распределением  $U(1,2)$ . При отборе методом линейного ранжирования (ЛР) производим упорядочивание популяции выборок в соответствии с их функцией соответствия, в этом случае данная функция будет эквивалентна весовому распределению  $\omega$ , соответствующему каждой популяции из выборки  $S$ . В таком случае получаем, что распределение  $P(s_i)$  будет зависеть только от порядкового номера особи в популяции. Совокупность частиц и соответствующие им веса, упорядоченные согласно весам  $\omega$ , запишем в виде следующего множества

$$S = \left\{ \left\{ S_i^1, \omega_i^1 \right\}, \dots, \left\{ S_i^N, \omega_i^N \right\} \right\}, i = 1, 2, \dots, n. \quad (12)$$

Для оптимизации линейного ранжирования Бейкера можно воспользоваться разделением выборки частиц на соответствующие схемы с минимальными и максимальными весами. Впервые применение механизма схем рассмотрено Холландом. Он предполагал, что для формирования схем можно использовать особи со схожим генотипом. В таком случае схему можно определить в следующем виде:

$$H_{S_{t+1}} = \left( \xi = B \mid \xi_{i_1} = h_{i_1}, \xi_{i_2} = h_{i_2}, \dots, \xi_{i_n} = h_{i_n} \right), i_1 < i_2 < \dots < i_n, \quad (13)$$

где  $B$  – множество генотипов, характерное для популяции частиц  $S_{t+1}^i$ .

Учитывая, что один и тот же генотип может быть характерен одновременно для нескольких особей (частиц), можно определить функцию приспособленности популяции частиц  $S_{t+1}$  с учетом схемы  $H_{S_{t+1}}$ . Получим [14]:

$$f(S_{t+1}, \Omega_{t+1}) = \frac{\sum_{\xi_{t+1} \in H} f(\xi_{t+1}^i)}{N(H_{S_{t+1}}, \Omega_{t+1})}, \quad (14)$$

где  $\Omega_{t+1}$  – поколения частиц, соответствующих моменту  $t+1$ ;  $N(H_{S_{t+1}}, \Omega_{t+1})$  – общее число частиц с генотипом  $\xi$  в поколении  $\Omega_{t+1}$ .

В соответствии с выражением (14) вероятность выживания отдельной популяции частиц будет удовлетворять следующему неравенству:

$$P_l(S_{t+1}, \Omega_{t+1}) \geq 1 - P_c(S_{t+1}^i, H_{S_{t+1}}) \frac{\delta(H_{S_{t+1}})}{l-1} (1 - P_l(S_t, \Omega_t)), \quad (15)$$

$$P_c(S_{t+1}^i, H_{S_{t+1}}) = \frac{1 - \delta(H_{S_{t+1}})}{l-1},$$

где  $\delta(H)$  – длина схемы  $H_{S_{t+1}}$ ;  $l$  – число элементов выборки (частиц).

Для анализа разнообразия популяции выборок можно определить вероятность изменения определенного гена в зависимости от параметров ГА  $P(N(H_{S_{t+1}}, \Omega_{t+1}))$  в соответствии с (8). Введем вероятность  $P(S_{t+1}^i, H_{S_{t+1}})$ , определяющую принадлежность выборки  $S_{t+1}^i$  схеме  $H_{S_{t+1}}$ . Запишем данную вероятность с учетом выражения (14):

$$P(S_{t+1}^i, H_{S_{t+1}}) = \frac{f(H_{S_{t+1}}, \Omega_{t+1}) N(H_{S_{t+1}}, \Omega_{t+1})}{N(B, \Omega_{t+1}) N} \quad (16)$$

$$= \frac{\sum_{\xi_{t+1} \in H_{S_{t+1}}} f(\xi_{t+1}^{it})}{\sum_{j=1}^N f(\xi_{t+1}^{jt})}.$$

Тогда выражение для расчета  $P(N(H_{S_{t+1}}, \Omega_{t+1}))$  будет иметь вид:

$$P(N(H_{S_{t+1}}, \Omega_{t+1}) = N) = \theta + P_c(S_{t+1}^i, H_{S_{t+1}}) (1 - 2\theta)^N,$$

$$P(N(H_{S_{t+1}}, \Omega_{t+1}) = 0) = (1 - \theta + P_c(S_{t+1}^i, H_{S_{t+1}}) (1 - 2\theta))^N, \quad (17)$$

$$\theta = P(S_{t+1}^i, H_{S_{t+1}}),$$

где  $P_m(S_{t+1}^i, H_{S_{t+1}})$  – вероятность мутации индивида  $S_{t+1}^i$  в соответствии со схемой  $H_{S_{t+1}}$ .

В соответствии с выражением (12) можно определить две группы, соответствующие области низких и высоких весов частиц  $S_{t+1}$ . Для этого введем критерий разделения. Данный критерий можно получить из нормализации весов  $\omega$ . Запишем данный критерий следующим образом:

$$N_{norm} = \frac{1}{\sum_{j=1}^N (\omega_j^i)^2}. \quad (18)$$

Тогда выборку  $S_{t+1}$ , соответствующую моменту времени  $t+1$ , можно записать в виде совокупности двух групп, соответствующих области высоких и низких весов, разделяемых в соответствии с  $N_{norm}$ . Получим:

$$H_{t+1} = \begin{cases} H'_{t+1} = \{\{\xi_{i_1}, \omega_{i_1}\}, \dots, \{\xi_{i_k}, \omega_{i_k}\}\}, \\ H''_{t+1} = \{\{\xi_{i_{k+1}}, \omega_{i_{k+1}}\}, \dots, \{\xi_{i_N}, \omega_{i_N}\}\}, \end{cases} \quad k \leq N_{norm} < k+1, \quad (19)$$



где  $N'_{t+1}$  и  $N''_{t+1}$  – схемы выборки, соответствующие области наибольших и наименьшим весов выборок  $S_{t+1}$ .

Далее перейдем к операции скрещивания популяции частиц для формирования оптимальной выборки. Основной задачей скрещивания является повышение разнообразия популяции частиц, что позволяет получить предельно неповторяющийся набор генотипов частиц в рамках фиксированного момента времени  $t + 1$ . Для этого воспользуемся равномерным кроссинговером (РК). В классической интерпретации производится равновероятностное копирование генов от нескольких родителей к потомку. Для усиления равномерного кроссинговера вместо равновероятностного выбора родительских вершин будем использовать метод взвешивания с учетом правдоподобия, а именно для каждой новой генерации оцениваются веса возможных родителей. Тогда процесс скрещивания можно представить в виде уравнений

$$\hat{S}_{t+1}^{i,j} = \begin{cases} \hat{S}_{t+1}^i = \alpha S_t^i + (1 - \beta) S_t^j, \\ \hat{S}_{t+1}^j = \beta S_t^i + (1 - \alpha) S_t^j, \end{cases} \quad (20)$$

$$\alpha = \frac{W(S_t^i)}{W(S_t^i) + W(S_t^j)}, \quad \beta = \frac{W(S_t^j)}{W(S_t^i) + W(S_t^j)},$$

где  $\hat{S}_{t+1}^{i,j}$  – новая популяция частиц, образованная путем скрещивания родителей  $S_t^i$  и  $S_t^j$  из выборки  $S_t$ ;  $\alpha, \beta$  – весовые коэффициенты, устанавливающие соответствие между весами родителей  $W(S_t^i)$  и  $W(S_t^j)$ .

Процедура мутации может быть реализована на основе распределения Гаусса. Каждый мутирующий представитель выборки  $S_{t+1} = (s_{t+1}^1, s_{t+1}^2, \dots, s_{t+1}^n)$  будет формироваться в соответствии со следующим выражением:

$$s_{t+1}^i = M(s_t^i), \quad M(s_t^i) = s_t^i + N(0, \sigma_i), \quad (21)$$

где  $N(0, \sigma_i)$  – гауссовское распределение;  $s_{t+1}^i$  – выборка, полученная на фазе  $t + 1$ , на основе которой производится мутация.

### Результаты и их обсуждение

Оптимизация МЧФ на основе ГА является важным алгоритмом оптимизации стохастических процедур вероятностного вывода. По результатам исследования нами решена задача повышения качества формирования выборок на основе МЧФ. Применение генетических алгоритмов доказывает правильность выбора исследования, а также состоятельность предложенного подхода. В рамках практической проверки предложенного метода оптимиза-

ции МЧФ на основе ГА произведено его сравнение с классическим МЧФ за счет оценки согласованности выборок на этапе повторной генерации выборок.

В табл. 1 приведем основные показатели согласованности  $\psi_{S_{t+1}}$  выборок  $S_{t+1}$  и свидетельств  $E_{t+1}$  в зависимости от общего числа свидетельств во всех временных срезах  $N_e$ .

Таблица 1

№	Алгоритм	$N_e = 200$	$N_e = 500$	$N_e = 1000$	$N_e = 5000$
1.	Алгоритм МЧФ	10%	15%	18%	25%
2.	Алгоритм МЧФ с ГА	3%	3%	3%	3%

Из табл. 1 получаем важный практический результат, заключающийся в том, что при использовании алгоритма ГА совместно с МЧФ наблюдаем фиксированную степень согласованности выборок, вне зависимости от числа переменных свидетельств  $N_e$ . Следовательно, каждая результирующая популяция, полученная по итогам выполнения алгоритма ГА, будет иметь наибольшее значение функции приспособленности  $\omega$ , что ведет к повышению точности апостериорного распределения вероятностей  $P(X_{t+1}|E_{t+1})$ . Другая особенность заключается в том, что при использовании ГА можно ограничить общее число первоначально формируемых выборок  $N$ , при этом можно повысить число шагов мутаций на этапе повторного взвешивания. Такой подход позволяет настроить точность алгоритма за минимальное число шагов алгоритма МЧФ. Отметим, что применение алгоритма ГА без МЧФ к СММ не имеет смысла, так как в процессе выполнения генетического алгоритма нет необходимой информации относительно свидетельств, а также вероятностных связей между тиражируемыми между срезами переменными. С помощью ГА можно оптимизировать существующую выборку, полученную по результатам выполнения МЧФ-фильтрации. Следовательно, для повышения эффективности МЧФ можно использовать различные подходы к его распараллеливанию, в таком случае формирования выборок  $S_{t+1}$  будет выполняться независимо друг от друга. Подобный подход позволяет максимально распределить вычислительные ресурсы для поэтапного выполнения МЧФ и ГА алгоритмов. В табл. 2 приведена сравнительная характеристика двух алгоритмов при работе в параллельном и распределенном режимах для различного объема выборок.

Таблица 2

№	Название алгоритма	Однопоточный режим	Параллельный режим на 1 узле	Параллельный режим на 6 узлах
1.	МЧФ	844,28878670758 сек.	648,15711330071 сек.	305,911736891564 сек.
2.	МЧФ с ГА	851,96107700558 сек.	656,059819253189 сек.	321,565701030924 сек.

Из анализа табл. 2 следует, что применение ГА в незначительной степени сказывается на производительности алгоритма. Однако при использовании МЧФ с ГА количество выборок, необходимых для достижения требуемого уровня согласованности  $\psi_{S_{t+1}}$ , может быть существенно сокращено.

### Заключение

Решение задач оптимизации существующих алгоритмов вероятностного вывода является актуальным направлением исследования. В первую очередь это связано с повышением сложности вероятностных моделей, повышением числа скрытых переменных, а также с ростом потока свидетельств на каждом из временных срезов. В работе рассмотрено применение предложенных подходов к скрытым марковским моделям, однако данные алгоритмы могут быть адаптированы для реализации процедуры логического вывода как в статических, так и динамических вероятностных моделях. Использование ГА в процессе МЧФ-фильтрации позволяет решить задачу качества формирования выборок; взамен случайной выборки в процессе повторной генерации используется генетический алгоритм. Такой подход позволяет повысить точность апостериорного распределения условий роста переменных-свидетельств  $E_{t+1}$ , а также в полной мере использовать алгоритм взвешивания с учетом правдоподобия с целью формирования весовых распределений для первичной популяции выборок  $S_0 = (s_1, s_2, \dots, s_n)$ . Применение модели схем Холланда дает возможность разграничить области выборок с разными генотипами в соответствии с распределением  $W(X_{t+1}|E_{1:t+1})$ . Это достигается за счет того, что в рамках ГА весовое распределение выбирается в качестве функции соответствия, устанавливающей соответствия весов и каждой популяции частиц, входящей в состав выборки. Отметим, что предложенный алгоритм обладает схожей с классическим МЧФ производительностью, однако позволяет повысить область согласования выборок в соответствии с потоком свидетельств  $E_{1:t+k} = (e_1, e_2, \dots, e_n)$ , поступающих вплоть до момента времени  $t+k$ . Разработанный алгоритм достаточно хорошо масштабируем, и его можно распараллелить. В этом случае процедура формирования выборок  $S_{t+1}$  может выполняться независимо. Процедуру распараллеливания ГА можно реализовать на этапе скрещивания и мутации, так как отбор родителей особи каждой следующей популяции осуществляется в соответствии с весами потомков, которые уже заранее известны. Практические результаты анализа эффективности предложенного алгоритма позволяют прийти к выводу, что незначительное снижение производительности за счет использова-

ния в процессе выполнения ГА операций селекции, скрещивания и мутации по сравнению с классическим алгоритмом МЧФ дает возможность повысить точность апостериорного распределения и степень согласования выборок в соответствии со свидетельствами, вне зависимости от общего их числа. Такой подход позволяет использовать разработанный алгоритм при решении задач логического вывода в различных динамических вероятностных моделях с неограниченным числом временных состояний. Все это доказывает обоснованность и практическую значимость проведенных исследований.

#### ЛИТЕРАТУРА

1. *Holland J.H.* Adaptation in natural and artificial systems. – Cambridge: The MIT Press, 1992.
2. *Moral P.D.* Nonlinear Filtering: Interacting Particle Resolution // Markov Processing and Related Fields. – 1996. – Vol. 2(4). – P. 555-580.
3. *Moral P.D., Guionnet A.* Large Deviations for Interacting Particle Systems: Applications to nonlinear Filtering Problems // Stochastic Processes and their Applications. – 1998. – P. 69-95.
4. *Николенко С.И., Тулупьев А.Н.* Самообучающиеся системы. – М.: МЦНМО, 2009.
5. *Ross S.M.* Stochastic Processes. Second edition. – New York: Wiley, 1996.
6. *Рассел С., Норвиг П.* Искусственный интеллект: современный подход. – Изд. 2: пер. с англ. – М.: Вильямс, 2006.
7. *Bishop C.M.* Pattern recognition and machine learning. – NY: Springer, 2006.
8. *Liu J.S., Chen R.* Sequential Monte Carlo for Dynamic Systems // Journal of the ASA. – 1998. – Vol. 93(443). – P. 1032-1044.
9. *Pearl J.* Evidential reasoning using stochastic simulation of causal models // J. of Artificial Intelligence. – 1987. – No 32 – P. 247-257.
10. *Azarnova T.V., Polukhin P.V.* Advanced hybrid stochastic dynamic Bayesian network inference algorithm development in the context of the web applications test execution // Journal of Physics: Conf. Series: Materials Science and Engineering 537 052028, 2019.
11. *Back T.* Optimization by means of genetic algorithms // 36th International Scientific Colloquium. – 1991. – P. 163-169.
12. *Back T., Hoffmeister F.* Global Optimization by Means of Evolutionary Algorithms // Random Search as a Method for Adaptation and Optimization of Complex Systems. – 1991. – P. 17-21.
13. *Baker J.E.* Adaptive Selection Methods for Genetic Algorithms // Proceeding of the First Conf. on Genetic Algorithms. – 1985. – P 101-111.
14. *Алтухов Ю.П.* Генетические процессы в популяциях. – М.: Академкнига, 2003.

*Статья представлена к публикации членом редколлегии А.Д. Плутенко.*

*E-mail:*

*Полухин Павел Валерьевич – [alfa\\_force@bk.ru](mailto:alfa_force@bk.ru).*