

Оценивание состояний скрытых узлов контакт-центра¹

Н. А. Кузнецов^{*,a}, М. В. Лебедев^{*,**,b}, К. В. Семенихин^{*,**,c}

^{*}Институт радиотехники и электроники им. В.А. Котельникова РАН, Москва, Россия

^{**}Московский авиационный институт (национальный исследовательский университет),
Москва, Россия

e-mail: ^akuznetsov@cplire.ru, ^bmax_max82@inbox.ru, ^csiemenkv@gmail.com

Поступила в редколлегию 09.12.2024 г. Принята 13.12.2024 г.

Аннотация—Рассматривается частично-наблюдаемая стохастическая сеть, которая моделирует работу контакт-центра. Сеть состоит из основного (наблюдаемого) узла в виде экспоненциальной многоканальной системы конечной емкости с простейшим входящим потоком и двух виртуальных (скрытых) узлов, содержащих соответственно заблокированных и неудовлетворенных абонентов, которые с определенной интенсивностью производят повторные вызовы. Из уравнений оптимальной нелинейной фильтрации синтезирован субоптимальный фильтр, отслеживающий загрузку скрытых узлов. Проведено его сравнение с фильтром, основанным на описании динамики сети в среднем. Кроме того, представлены численные результаты оценивания параметров на основе предложенных схем фильтрации.

КЛЮЧЕВЫЕ СЛОВА: контакт-центр, система массового обслуживания, частично наблюдаемая система, стохастическая сеть, точечный процесс, стохастическая фильтрация, оценка параметров

DOI: 10.53921/18195822_2024_24_4_392

1. ВВЕДЕНИЕ

Контакт-центры обеспечивают важное средство взаимодействия между поставщиками услуг и их потребителями. Для математического описания многоступенчатого характера обработки, которой подвергается поток входящих вызовов, используют современные модели систем массового обслуживания. Основные направления исследований в области математического моделирования контакт-центров приведены в недавнем обзоре [1]. Среди наиболее востребованных задач можно выделить: планирование численности обслуживающего персонала и анализ характеристик обслуживания в стационарном режиме [2], моделирование и прогнозирование количества поступающих звонков [3], статистическое оценивание риска, связанного с увеличением времени ожидания клиентов [4] и их отказом от обслуживания [5].

Последние из указанных направлений относятся к области обратных задач теории массового обслуживания [6], где проблема восстановления неизвестной загрузки канала трактуется как задача стохастической фильтрации на основе частично наблюдаемой динамики системы массового обслуживания [7, 8, 9, 10]. Такие задачи естественным образом возникают при оценивании состояний скрытых узлов стохастических сетей [11] и сетей массового обслуживания [12, 13, 14].

В данной работе рассматривается модель контакт-центра в виде частично-наблюдаемой стохастической сети. Она включает в себя один наблюдаемый узел, обслуживающий простейший поток вызовов как многоканальная система, и два скрытых узла как виртуальные накопители

¹ Работа выполнена по государственному заданию FFWZ-2022-0006.

для заблокированных и неудовлетворенных абонентов, порождающих поток повторных вызовов. Загрузка обоих скрытых узлов потенциально неограничена, поэтому для точного описания апостериорного распределения ненаблюдаемых состояний необходима бесконечная система уравнений [15]. Вместо этого на базе выведенных в [14] уравнений оптимальной нелинейной фильтрации (для условных математических ожиданий, дисперсий и ковариаций) синтезирован субоптимальный алгоритм оценивания загрузки скрытых узлов по текущей наблюдательной информации. В ходе численного эксперимента проведено сравнение субоптимального фильтра с алгоритмом фильтрации, основанным на описании динамики сети в среднем. Кроме того, исследована возможность оценивания неизвестных параметров за счет параметрической настройки фильтров.

2. ОПИСАНИЕ МОДЕЛИ И ПОСТАНОВКА ЗАДАЧИ

Рассмотрим модель контакт-центра в виде сети массового обслуживания, схема которой изображена на рис. 1. Сеть состоит из трех узлов. Основной узел 1 принимает на вход простейший поток вызовов абонентов интенсивности λ . Входящие вызовы обрабатываются m параллельно работающими операторами. Если все они заняты, то вызов становится в очередь, в которой зарезервировано $K - m \geq 0$ мест для ожидания. Если и эти места заняты, то вызов блокируется. Это событие моделируется переходом на узел 2 — виртуальный накопитель абонентов, производящих повторные вызовы. Кроме того, после обслуживания отдельный абонент может с вероятностью r остаться неудовлетворенным: ему нужна дополнительная информация или он забыл что-то уточнить и т.п. Такое событие соответствует переходу на другой виртуальный узел — узел 3. Неудовлетворенный абонент также делает повторный вызов; но если узел 1 занят, его вызов переходит на узел 2.

Время обработки вызова одним оператором определяется случайной величиной с экспоненциальным распределением $\mathcal{E}(\mu_1)$. Максимальное число заявок на основном узле равно K . Повторные вызовы, порождаемые заблокированными и неудовлетворенными абонентами, производятся через время, распределенное по закону $\mathcal{E}(\mu_2)$ и $\mathcal{E}(\mu_3)$, соответственно.

Тем самым узлы 1, 2, 3 представляют собой экспоненциальные системы массового обслуживания: $\cdot/M_{\mu_1}/m/K$, $\cdot/M_{\mu_2}/\infty$ и $\cdot/M_{\mu_3}/\infty$, соответственно.

Начальное состояние сети предполагается нулевым.

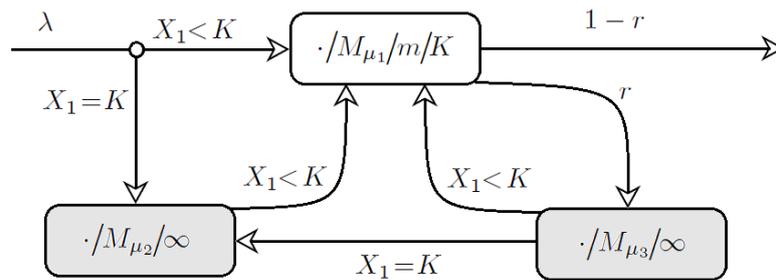


Рис. 1. Схема контакт-центра: светлый прямоугольник — наблюдаемый узел 1 с текущим числом вызовов X_1 , левый и правый темные прямоугольники — скрытые узлы 2 и 3, соответственно.

Узел 1 является *наблюдаемым*, так как его загрузка (т.е. число занятых операторов плюс число вызовов в очереди) известна в любой момент времени. Напротив, узлы 2 и 3 являются *скрытыми*: данные о количестве заблокированных и неудовлетворенных абонентов напрямую не доступны. Поэтому основная цель работы состоит в решении задачи *фильтрации* скрытых

состояний контакт-центра — синтезировать оценки $\hat{X}_2(t), \hat{X}_3(t)$ текущих состояний $X_2(t), X_3(t)$ узлов 2 и 3 по информации $\{X_1(s): s \leq t\}$, доступной к данному моменту t о загрузке узла 1.

Данная постановка весьма востребована на практике, поскольку позволяет оперативно отслеживать предоставляемое качество обслуживания. Если объем необслуженных и неудовлетворенных абонентов оценивается адекватно, то клиенты контакт-центра будут избавлены от избыточных и навязчивых обращений службы обратной связи, которая занимается выяснением мнения абонентов после регистрации их вызовов.

Синтез оценок фильтрации основан на знании параметров модели: $\lambda, \mu_1, \mu_2, \mu_3, r$. Поэтому дополнительно требуется изучить возможность настройки алгоритма фильтрации в случае, когда параметры рассматриваемой сети не заданы.

3. ЧАСТИЧНО НАБЛЮДАЕМАЯ СТОХАСТИЧЕСКАЯ СЕТЬ

Рассматриваемая модель контакт-центра может быть описана в терминах стохастической сети [11]. Допустим, что сеть образована множеством узлов $S = \{1, 2, \dots, d\}$, каждый из которых может принимать заявки от других узлов и извне. Фиктивный узел 0 обозначает одновременно источник входящего трафика и финальный результат — завершение обслуживания в сети. Динамика сети описывается *сетевым процессом* $X(t) = \{X_\alpha(t)\}_{\alpha \in S}$. Это векторный процесс, который определен на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$ и составлен из компонент $X_\alpha(t)$, численно равных текущей загрузке (т.е. числу заявок, вызовов, требований и т.д.) соответствующего узла α в момент $t \geq 0$.

Возможны только три типа событий, когда заявка:

- а) переходит с узла $\alpha \in S$ на другой узел $\beta \in S$;
- б) заканчивает обслуживание в сети на узле $\alpha \in S$;
- в) прибывает в сеть через узел $\beta \in S$.

Эти потоки событий описываются точечными процессами $N_{\alpha,\beta}(t), N_{\alpha,0}(t)$ и $N_{0,\beta}(t)$, которые имеют непрерывные справа траектории и представление

$$N_{\alpha,\beta}(t) = \int_0^t \nu_{\alpha,\beta}(s) ds + M_{\alpha,\beta}(t)$$

через квадратично интегрируемые \mathbb{F} -мартингалы $M_{\alpha,\beta}$ и \mathbb{F} -предсказуемые интенсивности $\nu_{\alpha,\beta} \geq 0$, где $\mathbb{F} = \{\mathcal{F}_t\}_{t \geq 0}$ — пополнение фильтрации, порожденной указанными точечными процессами и начальным состоянием сети $X(0)$ (см., например, [16]).

Состояние узла $m \in S$ принимает вид

$$X_m(t) = X_m(0) + \sum_{\alpha} N_{\alpha,m}(t) - \sum_{\beta} N_{m,\beta}(t),$$

где α, β пробегает $S \cup \{0\}$.

Частично наблюдаемая стохастическая сеть задается разбиением $S = J \sqcup H$ на наблюдаемые узлы $j \in J$ и скрытые узлы $k \in H$.

Далее перечисление $m \notin J$ означает $m \in H \cup \{0\}$, т.е. фиктивный узел 0 не считается наблюдаемым, поскольку для скрытых узлов нет достоверной информации о приходе на них внешних заявок или об уходе с них тех заявок, которые покидают сеть.

Введем фильтрацию $\mathbb{Y} = \{\mathcal{Y}_t\}_{t \geq 0}$, порожденную начальным состоянием всей сети $X(0)$ и векторным процессом $Y(t) = \{X_i(t)\}_{i \in J}$, определяющим загрузку наблюдаемых узлов.

Задача оптимальной фильтрации состояний частичной наблюдаемой стохастической сети состоит в том, чтобы определить условное математическое ожидание $\hat{Z}(t) = \mathbb{E}\{Z(t) | \mathcal{Y}_t\}$

случайного вектора $Z(t) = \{X_k(t)\}_{k \in H}$, описывающего загрузку скрытой части сети, по наблюдениям $\{Y(s): s \leq t\}$, доступным к текущему моменту времени t .

Поскольку эта задача будет решаться без нахождения полного апостериорного распределения $\text{Law}\{Z(t) | \mathcal{Y}_t\}$, для описания точности оценки понадобится условная ковариационная матрица $Q(t) = \text{cov}\{Z(t), Z(t) | \mathcal{Y}_t\}$.

4. ФИЛЬТРАЦИЯ СКРЫТЫХ СОСТОЯНИЙ СТОХАСТИЧЕСКОЙ СЕТИ

4.1. Уравнения оптимального нелинейного фильтра

Оценки состояний скрытых узлов $\hat{X}_k(t) = \mathbf{E}\{X_k(t) | \mathcal{Y}_t\}$, а также условные ковариации $Q_{k,l}(t) = \mathbf{E}\{\varepsilon_k(t)\varepsilon_l(t) | \mathcal{Y}_t\}$, $k, l \in H$ ошибок оценивания $\varepsilon_k(t) = X_k(t) - \hat{X}_k(t)$, будучи согласованы с фильтрацией \mathbb{Y} , выражаются только через начальное состояние сети $X(0)$ и три типа точечных процессов

$$N_{i,j}, \quad N_j^a = \sum_{k \notin J} N_{k,j} \quad \text{и} \quad N_i^d = \sum_{k \notin J} N_{i,k} \quad (i, j \in J).$$

Процессы $\{N_{i,j}\}$ описывают переходы внутри наблюдаемой части сети J , а N_j^a и N_i^d определяют, соответственно, число поступлений на узел $j \in J$ с любого ненаблюдаемого $k \notin J$ и число уходов с узла $i \in J$ на любой ненаблюдаемый $k \notin J$. Соответствующие интенсивности принимают форму $\nu_j^a = \sum_{k \notin J} \nu_{k,j}$ и $\nu_i^d = \sum_{k \notin J} \nu_{i,k}$.

Сделаем несколько предположений о рассматриваемой модели сети:

(i) скачки точечных процессов $N_{\alpha,\beta}$ не происходят одновременно (иначе говоря, моменты всех событий в сети различны);

(ii) для $i \in J$, $\beta \in S \cup \{0\}$ интенсивности $\nu_{i,\beta}$ \mathbb{Y} -предсказуемы, т.е. имеется прямая информация о скорости, с которой заявки уходят с наблюдаемых узлов;

(iii) $\exists C = \text{const}: \sum_{\alpha,\beta} \nu_{\alpha,\beta} \leq C \sum_{\alpha,\beta} N_{\alpha,\beta}$, что означает не более чем линейный рост интенсивностей относительно числа событий в сети.

Теорема. В условиях (i)–(iii) справедливы следующие утверждения:

1) для узла $k \in H$ оценка его состояния $\hat{X}_k(t)$ описывается системой уравнений

$$d\hat{X}_k = \{\hat{\nu}_k^a - \hat{\nu}_k^d - \sum_{j \in J} \hat{c}_{k,j}\} dt + \sum_{i \in J} \xi_{i,k}^d dN_i^d + \sum_{j \in J} \xi_{k,j}^a dN_j^a, \quad (1)$$

$$\hat{\nu}_k^a = \sum_{m \notin J} \hat{\nu}_{m,k}, \quad \hat{\nu}_k^d = \sum_{m \notin J} \hat{\nu}_{k,m}, \quad \xi_{i,k}^d = \frac{\nu_{i,k}}{\nu_i^d}, \quad \xi_{k,j}^a = \frac{\hat{c}_{k,j} - \hat{\nu}_{k,j}}{\hat{\nu}_j^a}, \quad (2)$$

$$\hat{c}_{k,j} = \text{cov}\{X_k(t-), \nu_j^a | \mathcal{Y}_{t-}\} = \widehat{X}_k \nu_j^a - \hat{X}(t-) \hat{\nu}_j^a, \quad (3)$$

2) для узла $k \in H$ условная дисперсия ошибки $Q_{k,k}(t)$ удовлетворяет уравнению

$$dQ_{k,k} = \left(\hat{\nu}_k^a + \hat{\nu}_k^d + 2\hat{b}_{k,k} - \sum_{j \in J} \hat{\tau}_{k,k,j} \right) dt + \sum_{i \in J} (1 - \xi_{i,k}^d) \xi_{i,k}^d dN_i^d + \sum_{j \in J} \left\{ \frac{1}{\hat{\nu}_j^a} (\hat{\tau}_{k,k,j} + \hat{\nu}_{k,j} - 2\hat{\alpha}_{k,k,j}) - (\xi_{k,j}^a)^2 \right\} dN_j^a, \quad (4)$$

3) для пары узлов $k \neq l$ условная ковариация ошибки $Q_{k,l}(t)$ имеет вид

$$dQ_{k,l} = (\hat{b}_{k,l} + \hat{b}_{l,k} - \hat{\nu}_{k,l} - \hat{\nu}_{l,k} - \sum_{j \in J} \hat{\tau}_{k,l,j}) dt - \sum_{i \in J} \xi_{i,k}^d \xi_{i,l}^d dN_i^d + \sum_{j \in J} \left\{ \frac{1}{\hat{\nu}_j^a} (\hat{\tau}_{k,l,j} - \hat{\alpha}_{k,l,j} - \hat{\alpha}_{l,k,j}) - \xi_{k,j}^a \xi_{l,j}^a \right\} dN_j^a, \quad (5)$$

где использованы \mathbb{Y} -предсказуемые версии соответствующих условных ковариаций:

$$\begin{aligned}\hat{\varkappa}_{k,l,j} &= \text{cov}\{X_k(t-), \nu_{l,j} | \mathcal{Y}_{t-}\}, \quad \hat{\tau}_{k,l,j} = \text{cov}\{\varepsilon_k(t-)\varepsilon_l(t-), \nu_j^a | \mathcal{Y}_{t-}\}, \\ \hat{b}_{k,l} &= \text{cov}\{X_k(t-), \nu_l^a - \nu_l^d | \mathcal{Y}_{t-}\} = \sum_{m \notin J} (\hat{\varkappa}_{k,m,l} - \hat{\varkappa}_{k,l,m}).\end{aligned}$$

Доказательство теоремы опубликовано в [14].

Поясним структуру оценки (1). Отношения $\nu_{i,k}/\nu_i^d$ и $\hat{\nu}_{k,j}/\hat{\nu}_j^a$ определяют, соответственно, долю переходов с наблюдаемых узлов на данный скрытый узел k и долю переходов с узла k в наблюдаемую часть сети. Единственное отличие между этими двумя членами — дополнительное корректирующее слагаемое $\hat{c}_{k,j}$, которое добавляется к скачкам dN_j^a и вычитается из соответствующего им сноса.

4.2. Конечномерный фильтр для сети специального типа

Для получения конечномерной реализации фильтра рассмотрим сеть, в которой интенсивность переходов со скрытых узлов линейно зависит от их загрузки:

$$\nu_{0,\beta} = \lambda_\beta \quad \text{и} \quad \nu_{k,\beta} = \mu_{k,\beta} X_k(t-) \quad \text{для} \quad k \in H,$$

где λ_β и $\mu_{k,\beta}$ — \mathbb{Y} -предсказуемые коэффициенты. Тогда коэффициенты фильтра (1)–(5) выражаются через сами оценки $\{\hat{X}_k\}_{k \in H}$ и условные ковариации их ошибок $\{Q_{k,l}\}_{k,l \in H}$

$$\hat{c}_{k,j} = \sum_{m \in H} Q_{k,m}(t-) \mu_{m,j}, \quad \hat{\nu}_{k,\beta} = \mu_{k,\beta} \hat{X}_k(t-) \quad \text{и} \quad \hat{\varkappa}_{k,l,\beta} = Q_{k,l}(t-) \mu_{l,\beta}, \quad (6)$$

но за исключением коэффициентов $\hat{\tau}_{k,l,j}$, определяющих условные моменты третьего порядка.

Для упрощения уравнений (4)–(5) предлагается исключить соответствующие слагаемые, так как они порождают центрированный мартингал относительно фильтрации \mathbb{Y} :

$$M_{k,l}^\tau = \int_0^t \sum_{j \in J} \hat{\tau}_{k,l,j} (dN_j^a / \hat{\nu}_j^a - dt).$$

Тогда такое упрощение имеет смысл операции проектирования.

Теперь система (1)–(5) определяет конечномерный фильтр, который далее называется *субоптимальным*. Его конструкцию можно изложить следующим образом.

1) Согласно постановке начальные значения оценки определяются известным состоянием, поэтому условные ковариации нулевые: $\hat{X}_k(0) = X_k(0)$ и $Q_{k,l}(0) = 0$.

2) Если в момент t заявка уходит с наблюдаемого узла i на какой-либо ненаблюдаемый узел $m \notin J$, то оценка вместе с ее условной дисперсией и условными ковариациями меняется скачкообразно:

$$\begin{aligned}\hat{X}_k(t) &= X_k(t-) + \xi_{i,k}^d, \\ Q_{k,l}(t) &= Q_{k,l}(t-) + (\delta_{k,l} - \xi_{i,k}^d) \xi_{i,l}^d,\end{aligned}$$

где $\xi_{i,k}^d$ обозначено в (2), а $\delta_{k,l}$ — символ Кронекера.

3) Если в момент t заявка приходит на наблюдаемый узла j с какого-либо ненаблюдаемого узла $m \notin J$, то пересчет оценки и ее характеристик происходит по правилу:

$$\begin{aligned}\hat{X}_k(t) &= X_k(t-) + \xi_{k,j}^a, \\ Q_{k,l}(t) &= Q_{k,l}(t-) + \{\hat{\nu}_{k,j} \delta_{k,l} - (\hat{\varkappa}_{k,l,j} + \hat{\varkappa}_{l,k,j})\} / \hat{\nu}_j^a - \xi_{k,j}^a \xi_{l,j}^a,\end{aligned}$$

где коэффициенты $\hat{\nu}_j^a$, $\hat{\nu}_{k,j}$, $\xi_{k,j}^a$ и $\hat{\nu}_{k,l,j}$ выражаются согласно (2) и (6).

4) Между скачками процессов N_i^d , N_j^a субоптимальный фильтр описывается системой линейных обыкновенных дифференциальных уравнений

$$\begin{aligned}\dot{\hat{Z}} &= \Lambda^\top \hat{Z} + \lambda - Q\gamma, \\ \dot{Q} &= (Q - \text{diag}[\hat{Z}])\Lambda + \Lambda^\top(Q - \text{diag}[\hat{Z}]) + \text{diag}[\Lambda^\top \hat{Z} + \lambda]\end{aligned}$$

с учетом следующих векторно-матричных обозначений:

$\hat{Z} = \hat{Z}(t)$ — вектор-столбец, определяющий оценку скрытой части сети;
 $Q = Q(t)$ — матрица условных ковариаций ошибок;
 $\lambda = \{\lambda_k\}_{k \in H}$ — вектор-столбец интенсивностей входящих потоков;
 $\gamma = \{\gamma_k\}_{k \in H}$ — вектор-столбец интенсивностей переходов на скрытые узлы со всех наблюдаемых, т.е. $\gamma_k = \sum_{j \in J} \mu_{k,j}$;
 $\Lambda = \{\Lambda_{k,l}\}_{k,l \in H}$ — матрица интенсивностей переходов между скрытыми узлами, а именно, $\Lambda_{k,l} = \mu_{k,l} - \delta_{k,l} \sum_{m \notin J} \mu_{k,m}$.

5. АЛГОРИТМЫ ФИЛЬТРАЦИИ СКРЫТЫХ СОСТОЯНИЙ КОНТАКТ-ЦЕНТРА

На основе результатов предыдущего раздела опишем алгоритм субоптимальной фильтрации для оценивания состояний скрытых узлов контакт-центра.

Для этого запишем интенсивности переходов между узлами $\{0, 1, 2, 3\}$:

$$\begin{aligned}\nu_{0,1} &= \lambda(1 - \beta), & \nu_{0,2} &= \lambda\beta, \\ \nu_{1,0} &= \mu_1(1 - r)(X_1(t-) \wedge m), & \nu_{1,3} &= \mu_1 r(X_1(t-) \wedge m), \\ \nu_{2,1} &= \mu_2 X_2(t-) (1 - \beta), \\ \nu_{3,1} &= \mu_3 X_3(t-) (1 - \beta), & \nu_{3,2} &= \mu_3 X_3(t-) \beta,\end{aligned}$$

где $\beta(t) = \mathbb{I}\{X_1(t-) = K\}$ представляет собой индикатор того, что все приходящие на основной узел вызовы блокируются.

Динамика сети в терминах точечных процессов может быть записана как

$$\begin{aligned}dX_1 &= dN_{0,1} + dN_{2,1} + dN_{3,1} - (dN_{1,0} + dN_{1,3}), \\ dX_2 &= dN_{0,2} + dN_{3,2} - dN_{2,1}, \\ dX_3 &= dN_{1,3} - (dN_{3,1} + dN_{3,2}),\end{aligned}$$

причем наблюдаемые процессы и их интенсивности имеют вид

$$\begin{aligned}N_1^a &= N_{0,1} + N_{2,1} + N_{3,1}, & N_1^d &= N_{1,0} + N_{1,3}, \\ \nu_1^a &= (\lambda + \mu_2 X_2(t-) + \mu_3 X_3(t-)) (1 - \beta), & \nu_1^d &= \mu_1 (X_1(t-) \wedge m).\end{aligned}$$

Согласно (6) коэффициенты уравнений фильтрации (1)–(5) принимают вид

$$\begin{aligned}\hat{\nu}_3^d &= \hat{\nu}_{3,2} = \mu_3 \hat{X}_3(t-) \beta, & \hat{\nu}_2^d &= \hat{\nu}_3^a = 0, \\ \hat{\nu}_1^a &= (\lambda + \mu_2 \hat{X}_2(t-) + \mu_3 \hat{X}_3(t-)) (1 - \beta), & \hat{\nu}_2^a &= \hat{\nu}_{0,2} + \hat{\nu}_{3,2} = (\lambda + \mu_3 \hat{X}_3(t-)) \beta, \\ \xi_{1,2}^d &= 0, & \xi_{1,3}^d &= r, & \xi_{k,1}^a &= (Q_{k,2}(t-) \mu_2 + Q_{k,3}(t-) \mu_3 - \hat{X}_k(t-) \mu_k) / \hat{\nu}_1^a, \\ \hat{c}_{k,1} &= (Q_{k,2}(t-) \mu_2 + Q_{k,3}(t-) \mu_3) (1 - \beta), & k &= 2, 3, \\ \hat{\nu}_{k,l,1} &= \text{cov}\{X_k(t-), \nu_{l,1} | \mathcal{Y}_{t-}\} = Q_{k,l}(t-) \mu_l (1 - \beta), & k, l &= 2, 3, \\ \hat{b}_{2,2} &= \text{cov}\{X_2(t-), \nu_2^a | \mathcal{Y}_{t-}\} = Q_{2,3}(t-) \mu_3 \beta, \\ \hat{b}_{3,3} &= \text{cov}\{X_3(t-), -\nu_3^d | \mathcal{Y}_{t-}\} = -Q_{3,3}(t-) \mu_3 \beta, \\ \hat{b}_{3,2} + \hat{b}_{2,3} &= (Q_{3,3}(t-) - Q_{2,3}(t-)) \mu_3 \beta.\end{aligned}$$

Теперь субоптимальный фильтр (далее кратко SF) описывается системой из пяти нелинейных дифференциально-разностных уравнений:

$$\begin{aligned}d\hat{X}_2 &= ((\lambda + \mu_3 \hat{X}_3) \beta - (Q_{2,2} \mu_2 + Q_{2,3} \mu_3) (1 - \beta)) dt + \xi_{2,1}^a dN_1^a, \\ d\hat{X}_3 &= (-\mu_3 \hat{X}_3 \beta - (Q_{3,2} \mu_2 + Q_{3,3} \mu_3) (1 - \beta)) dt + \xi_{3,1}^a dN_1^a + r dN_1^d, \\ dQ_{2,2} &= (\lambda + \mu_3 (\hat{X}_3 + 2Q_{2,3})) \beta dt + \left\{ (\hat{\nu}_{2,1} - 2\hat{\nu}_{2,2,1}) / \hat{\nu}_1^a - (\xi_{2,1}^a)^2 \right\} dN_1^a, \\ dQ_{3,3} &= \mu_3 (\hat{X}_3 - 2Q_{3,3}) \beta dt + \left\{ (\hat{\nu}_{3,1} - 2\hat{\nu}_{3,3,1}) / \hat{\nu}_1^a - (\xi_{3,1}^a)^2 \right\} dN_1^a + (1 - r) r dN_1^d, \\ dQ_{2,3} &= \mu_3 (Q_{3,3} - Q_{2,3} - \hat{X}_3) \beta dt - \left\{ (\hat{\nu}_{2,3,1} + \hat{\nu}_{3,2,1}) / \hat{\nu}_1^a + \xi_{2,1}^a \xi_{3,1}^a \right\} dN_1^a.\end{aligned}$$

Помимо SF рассмотрим также фильтр DF (Drift-based Filter): он основан на замене в уравнениях скрытых узлов точечных процессов $dN_{k,l}$ их коэффициентами сноса $\bar{\nu}_{k,l} dt$, в которых вместо состояний $X_2(t), X_3(t)$ взяты их DF-оценки $\bar{X}_2(t), \bar{X}_3(t)$.

В результате фильтр DF описывается системой из двух линейных обыкновенных дифференциальных уравнений:

$$\dot{\bar{X}}_2 = (\lambda + \mu_3 \bar{X}_3) \beta - \mu_2 \bar{X}_2 (1 - \beta), \quad \dot{\bar{X}}_3 = \mu_1 r (X_1(t-) \wedge m) - \mu_3 \bar{X}_3.$$

Поскольку соответствующие траектории оценок представляют собой гладкие кривые (хотя и с изломами в моменты переключения наблюдаемого состояния), фильтр DF будем также называть *сглаженным*.

6. ЧИСЛЕННЫЙ ЭКСПЕРИМЕНТ

6.1. Сравнение алгоритмов фильтрации

Пусть модель контакт-центра, описанная в разделе 2, имеет следующие параметры:

$$m = 20, \quad K = 25, \quad \lambda = 21, \quad \mu_1 = 2, \quad \mu_2 = 10.5, \quad \mu_3 = 4.2, \quad r = 0.45.$$

Здесь важно отметить, что условием устойчивости данной модели выступает неравенство $\lambda < \mu_1(1 - r)m$. Для компьютерного моделирования интенсивность входящего потока $\lambda = 21$ взята меньше указанной границы $\mu_1(1 - r)m = 22$, но близкой к ней. Это с одной стороны гарантирует, что не будет происходить неограниченного роста вызовов на узлах 2 и 3, а с другой

стороны загрузка узла 1 будет чаще достигать предельного уровня, что влечет частое возникновение заблокированных пользователей. Очевидно, что если бы загрузка узла 2 была почти всегда нулевой, то задача оценивания такого состояния не имела бы большого смысла.

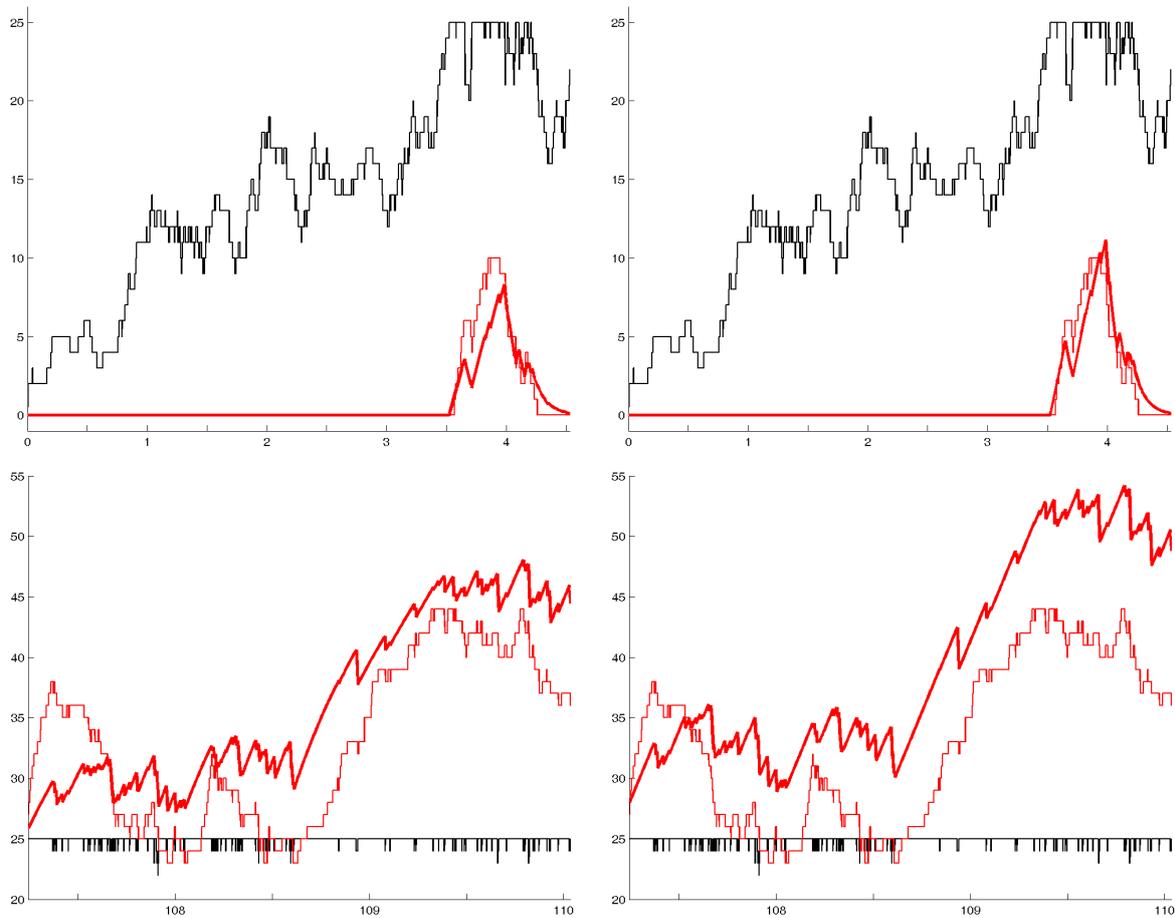


Рис. 2. Траектории состояний узла 1 (тонкая черная линия), узла 2 (тонкая красная линия) и его оценки фильтрации (жирная красная линия: SF — слева, DF — справа) на двух интервалах (верхний и нижний графики).

На рис. 2 представлены траектории загрузки наблюдаемого узла 1 и состояния скрытого узла 2 вместе с его оценками при использовании субоптимального фильтра (SF — левые графики) и сглаженного фильтра (DF — правые графики). Кроме того, для сравнения взяты два интервала: начальный интервал, на котором узел 1 лишь эпизодически оказывается переполнен (см. верхние графики рис. 2), и второй интервал, на котором узел 1 большую часть времени находится в заблокированном состоянии (см. нижние графики рис. 2). На рис. 3 приведено такое же сравнение оценок для узла 3.

По рис. 2 можно сделать вывод о том, что при оценивании состояния узла 2 фильтры SF и DF порождают оценки со схожим характером траекторий: они одинаково быстро отслеживают изменения в состоянии наблюдаемого узла. Но рис. 3 демонстрирует, что оценки состояния узла 3 имеют разную структуру: фильтр SF порождает скачкообразную траекторию, быстро откликающуюся на смену состояний наблюдаемого процесса, а фильтр DF, напротив, приводит к гладкой медленно меняющейся оценке. Последнее особенно явно проявляется на интервале,

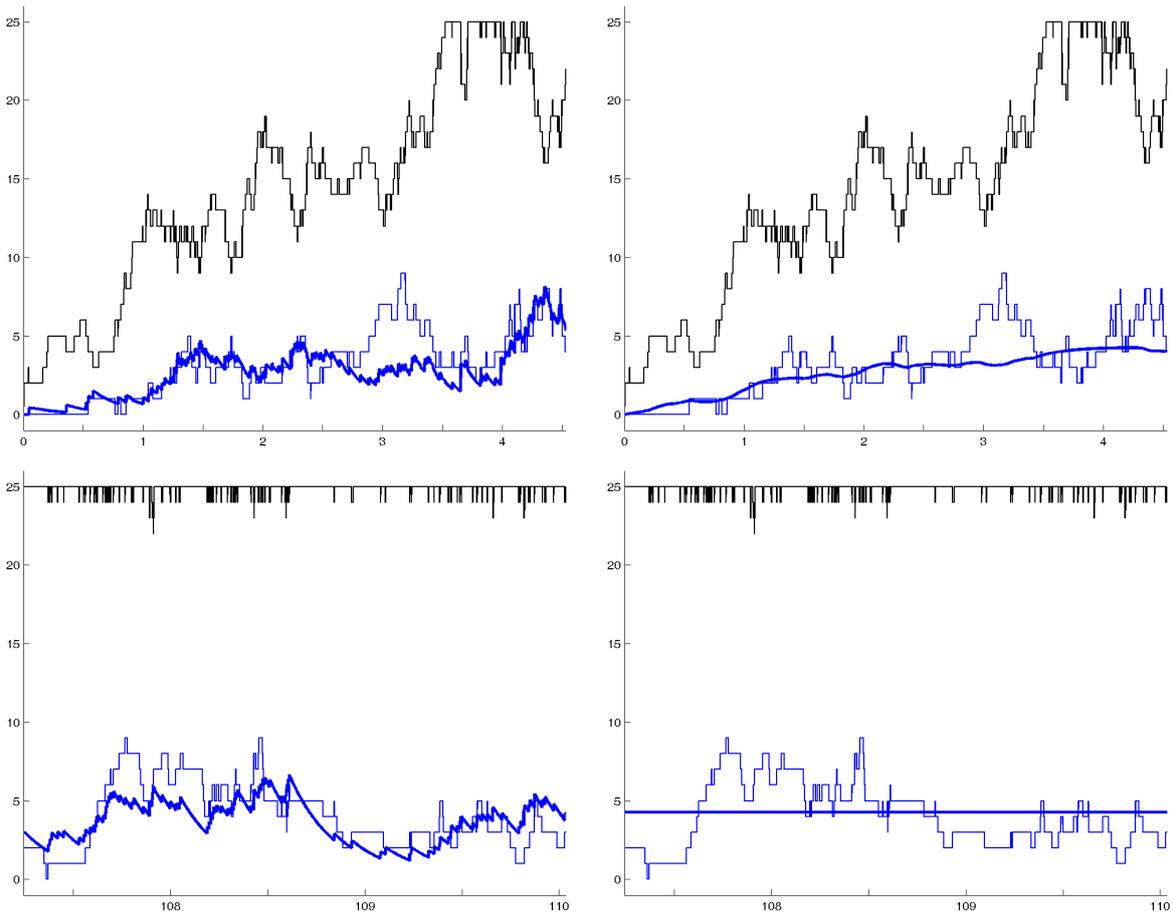


Рис. 3. Траектории состояний узла 1 (тонкая черная линия), узла 3 (тонкая синяя линия) и его оценки фильтрации (жирная синяя линия: SF — слева, DF — справа) на двух интервалах (верхний и нижний графики).

где основной узел большую часть времени проводит в перегруженном состоянии и DF-оценка превращается в константу (см. нижний правый график рис. 3).

Для каждой построенной оценки, взяв соответствующую ошибку фильтрации $\varepsilon(t)$, вычислим среднеквадратичную ошибку (Root Mean Square Error) на N смежных интервалах:

$$\text{RMSE}(t_k) = \left\{ \frac{1}{M} \sum_{l=0}^{M-1} \varepsilon^2(t_k - lh) \right\}^{1/2}, \quad t_k = k \cdot Mh, \quad k = 1, \dots, N,$$

где M — число точек в одном интервале, по которому берется усреднение с шагом h . На рис. 4 представлены графики среднеквадратичной ошибки оценок, полученных с помощью фильтров SF и DF на промежутке, разбитого на $N = 40$ интервалов и содержащего около 10 000 скачков сетевого процесса. Левый график рис. 4 показывает, что для узла 2 качество SF и DF оценок почти одинаково, а правый график демонстрирует превосходство субоптимального фильтра (примерно на 15%) над сглаженным фильтром, использующем только коэффициенты сноса.

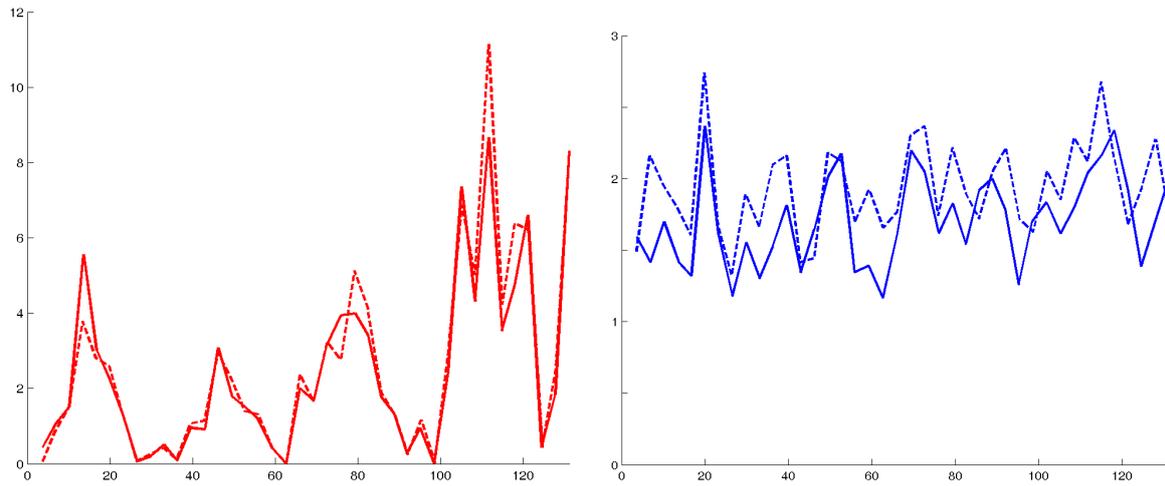


Рис. 4. Среднеквадратичные ошибки SF-фильтра (сплошные линии) и DF-фильтра (штриховые линии) при оценивании состояния узла 2 (слева) и узла 3 (справа).

6.2. Параметрическая идентификация модели контакт-центра

Для оценивания вектора параметров $p = \{\lambda, \mu_1, \mu_2, \mu_3, r\}$ воспользуемся методом наименьших квадратов

$$\text{ASR}_W^{(p)} = \frac{1}{M} \sum_{l=1}^M (\tilde{X}_1^p(t_l) - X_1(t_l))^2 \rightarrow \min_{p \in \mathbb{R}_+^5} \quad (7)$$

с учетом того, что

$\text{ASR}_W^{(p)}$ — среднеквадратичная невязка (Averaged Squared Residuals) на окне $W = [t_0, t_M]$ при использовании вектора параметров p в уравнениях фильтрации;

$\{t_l\}_{l=0,1,\dots,M}$ — разбиение окна, на котором реализуется один шаг обработки данных;

$\tilde{X}_1^p(t)$ — прогноз наблюдаемого состояния $X_1(t)$ на основе оценок фильтрации $\tilde{X}_2^p(t)$ и $\tilde{X}_3^p(t)$ и коэффициентов сноса:

$$\dot{\tilde{X}}_1^p = (\lambda + \mu_2 \tilde{X}_2^p + \mu_3 \tilde{X}_3^p) \mathbb{I}\{X_1 < K\} - \mu_1 (X_1 \wedge m), \quad \tilde{X}_1^p(t_0) = X_1(t_0).$$

Таким образом, прогноз $\tilde{X}_1^p(t)$ играет роль функции регрессии, которая используется для аппроксимации наблюдаемого состояния $X_1(t)$ за счет наилучшего выбора вектора параметров p в смысле минимизации (7).

Для численных расчетов использовался метод Левенберга—Марквардта, реализованной в пакете Octave 9.2 в виде функции `leasqr` [17]. Окно W покрывало 5000 скачков наблюдаемого процесса. На одном окне алгоритм `leasqr` минимизировал $\text{ASR}_W^{(p)}$ до достижения порога чувствительности `stol` = 10^{-3} . Оценки, полученные на одном окне, затем передавались на следующий этап обработки в качестве начальных приближений, и тот же алгоритм применялся к новому окну, сдвинутому относительно прежнего на величину Δt (этот сдвиг соответствовал 50 скачкам наблюдаемого процесса).

На рис. 5 представлены графики, описывающие эволюцию оценок параметров при использовании двух фильтров субоптимального SF и сглаженного DF. В целом, качество восстановления параметров невелико: удается лишь подобрать правильный порядок соответствующих параметров. Сглаженный фильтр DF, который демонстрирует малую чувствительность к вы-

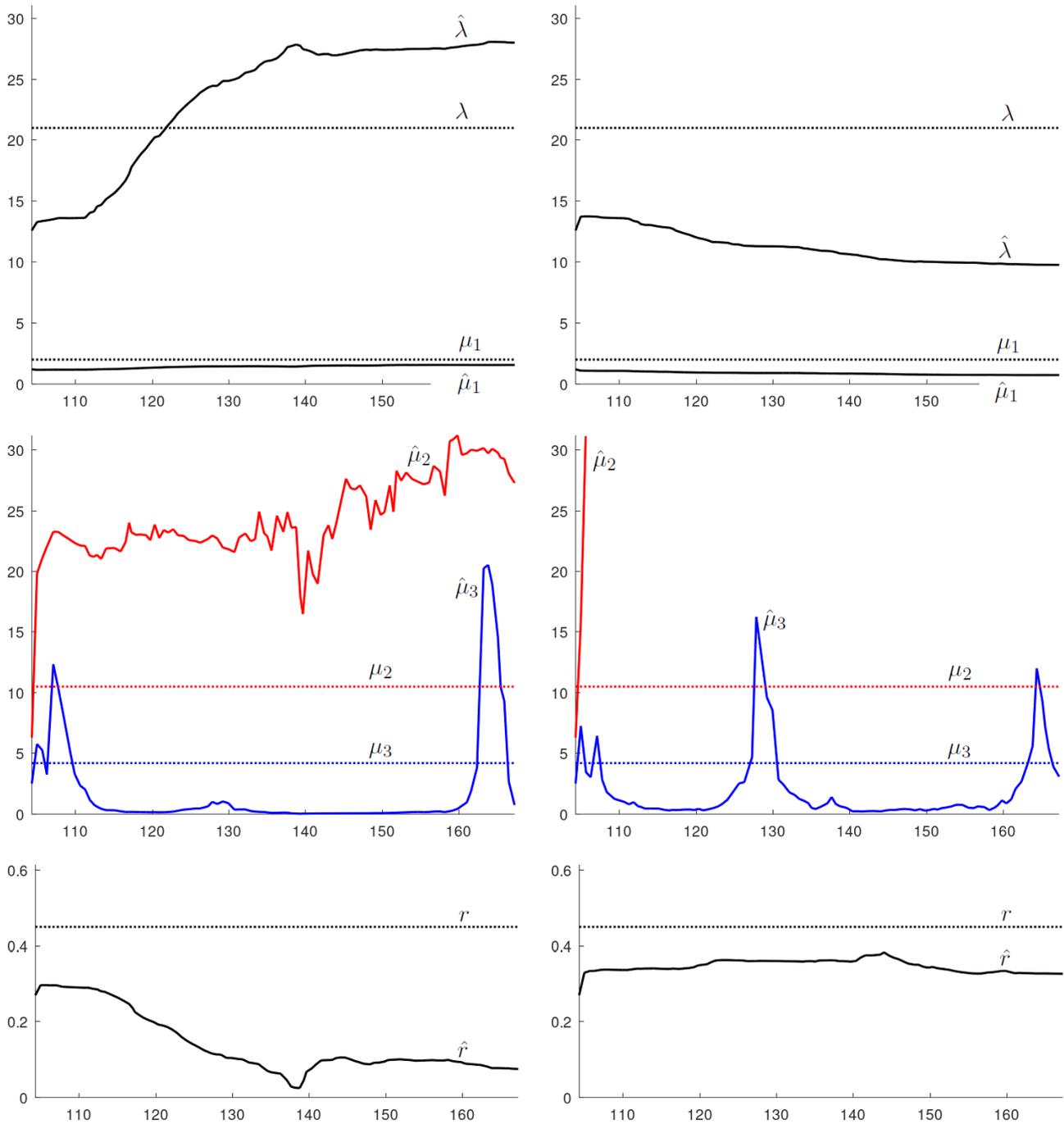


Рис. 5. Оценки параметров (сплошные кривые) и их точные значения (пунктир) при использовании SF-фильтра (слева) и DF-фильтра (справа).

бору параметров, приводит к расходящейся оценке параметра μ_2 , т.е. даже порядок интенсивности повторных вызовов от заблокированных абонентов фильтром DF не определяется. Вместе с тем сложный устроенный субоптимальный фильтр SF, резко реагирующий на изменения в наблюдаемом состоянии и, следовательно, достаточно чувствительный к выбору параметров, оказывается лучше приспособлен к параметрической настройке модели. Хуже всего оценивается параметр μ_3 , отвечающий за интенсивность повторных вызовов неудовлетворенных

абонентов. Это может свидетельствовать о том, что доступная измерительная информация в виде загрузки основного узла недостаточна для состоятельного оценивания всех параметров сети.

7. ЗАКЛЮЧЕНИЕ

В работе предложен алгоритм субоптимальной фильтрации скрытых состояний частично наблюдаемой стохастической сети, которая служит моделью контакт-центра. В сравнении со сглаженным фильтром, основанным на описании динамики сети в среднем, субоптимальный фильтр продемонстрировал более реактивный отклик на изменения наблюдаемого состояния и большую чувствительность к выбору параметров. Эти обстоятельства позволили устойчиво оценивать параметры сети. Вместе с тем, полная параметрическая неопределенность приводит к плохо идентифицируемой модели контакт-центра. Решение этой проблемы основано на введении дополнительной измерительной информации, т.е. организации выборочного опроса. Тогда назначение его интенсивности с учетом баланса интересов абонентов и точности оценивания приводит к новой задаче управления наблюдениями.

СПИСОК ЛИТЕРАТУРЫ

1. Koole G.M., Li S. A practice-oriented overview of call center workforce planning. *Stochastic Systems*, 2023, vol. 13, no. 4, pp. 479–495.
2. Степанов С.Н., Степанов М.С. Алгоритмы оценки показателей пропускной способности обобщенной модели контакт-центра. *Автоматика и телемеханика*, 2016, № 7, стр. 86–102.
3. Ibrahim R., Ye H., L'Ecuyer P., Shen H. Modeling and forecasting call center arrivals: A literature survey and a case study. *International Journal of Forecasting*, 2016, vol. 32, no. 3, pp. 865–874.
4. Ye H., Brown L.D., Shen H. Hazard rate estimation for call center customer patience time. *IJSE Transactions*, 2020, vol. 52, no. 8, pp. 890–903.
5. Aktekin T. Call center service process analysis: Bayesian parametric and semi-parametric mixture modeling. *European Journal of Operational Research*, 2014, vol. 234, no. 3, pp. 709–719.
6. Baccelli F., Kauffmann B., Veitch D. Inverse problems in queueing theory and Internet probing. *Queueing Systems*, 2009, vol. 63, pp. 59–107.
7. Lukashuk L.I., Semenchuk Y.A. Filtering of a semi-Markov queueing system with retrials. *Cybernetics and Systems Analysis*, 1991, vol. 27, no. 4, pp. 627–631.
8. Миллер Б.М., Авраченко К.Е., Степанян К.В., Миллер Г.Б. Задача оптимального стохастического управления потоком данных по неполной информации. *Проблемы передачи информации*, 2005, том 41, № 2, стр. 89–110.
9. Борисов А.В. Применение методов оптимальной фильтрации для оперативного оценивания состояний сетей массового обслуживания. *Автоматика и телемеханика*, 2016, № 2, стр. 115–141.
10. Кузнецов Н.А., Мясников Д.В., Семенихин К.В. Оптимизация управления передачей данных по флуктуирующему каналу связи при неточной информации о его состоянии. *Информационные процессы*, 2018, том 18, № 2, стр. 86–105.
11. Serfozo R. *Introduction to Stochastic Networks*. New York: Springer, 1999.
12. Walrand J., Varaiya P. Flows in queueing networks: A martingale approach. *Mathematics of Operations Research*, 1981, vol. 6, no. 3, pp. 387–404.
13. Солодяников Ю.В. Управление и наблюдение для динамических сетей массового обслуживания. II. *Автоматика и телемеханика*, 2014, № 5, стр. 91–114.

14. Semenikhin K.V. State estimation in partially observed stochastic networks with queueing applications. In: *Modern Trends in Controlled Stochastic Processes: Emergence, Complexity and Computation*, vol. 41. Eds. A. Piunovskiy., Y. Zhang. Cham: Springer, 2021.
15. Elliott R.J., Aggoun L., Moore J.B. *Hidden Markov Models: Estimation and Control*. New York: Springer, 2008.
16. Liptser R.S., Shiriyayev A.N. *Statistics of Random Processes. 3rd ed.* New York: Springer, 2005.
17. <https://octave.sourceforge.io/optim/function/leasqr.html> (дата обращения: 01.12.2024).

State estimation for hidden nodes of a contact center

N.A. Kuznetsov, M.V. Lebedev, K.V. Semenikhin

A partially observable stochastic network is considered for modelling the operation of a contact center. The network consists of three nodes: the main node being observed is an exponential multichannel finite-capacity queueing system with Poisson arrivals; two hidden nodes are virtual stations for blocked and unsatisfied customers who make repeated calls at a certain intensity. Based on the nonlinear optimal filtering equations, a suboptimal filter is derived for load tracking in the hidden nodes. To provide a comparative analysis of the suboptimal filtering scheme, a drift-based filter is used. In addition, the numerical results of parameter estimation based on the proposed filters are presented.

KEYWORDS:

contact center, queueing system, partially observed system, stochastic network, point process, stochastic filtering, parameter estimation