С. А. Амелькин, М. В. Шустова, Байесовская оценка аномальности потока данных, *Искусственный интеллект и принятие решений*, 2016, выпуск 2, 55–59

С.А. Амелькин, М.В. Шустова

# Байесовская оценка аномальности потока данных[1]

**Аннотация.** Рассматривается задача оценки вероятности аномальности данных, ограниченных диапазоном допустимых значений. Предлагается универсальный метод решения задачи, основанный на байесовской оценке. Экспериментальная часть построена на примере анализа потока телеметрических данных с космического аппарата.

**Ключевые слова:** байесовская вероятность, аномалия, прогнозирование, космический аппарат, телеметрия.

## Введение

При работе со сложными системами, такими как бортовые системы космических аппаратов (КА), часто возникает необходимость в прогнозировании будущего состояния. Данные, поступающие из какой-либо системы, нередко ограничены некоторым диапазоном допустимых значений, что дает возможность определить, являются ли они аномальными (т.е. выходящими за пределы диапазона допустимых значений). Задачу прогнозирования аномальности будущего состояния предлагается решить с помощью байесовского подхода, применение которого позволит оценить (на основе имеющихся данных) вероятность перехода в аномальное состояние за заданный период времени, т.е. оценить то, насколько опасно текущее состояние системы.

Универсальность данного подхода позволила применять его для решения задач прогнозирования и моделирования в различных областях: анализ продолжительности жизни объекта [1], моделирование сезонного поведения туристов на основе ограниченных данных [2], прогнозирование продаж [3]. Помимо этого, формула Байеса была использована в работе [4] для анализа вероятности наступления авиационного происшествия, а в работе [5] с помощью байесовского подхода проведено сценарное политическое прогнозирование.

Предлагаемый подход позволяет решить задачу прогнозирования, которая на сегодняшний день является одной из основных проблем при организации корректного функционирования сложных технических систем. Кроме того, байесовский подход обеспечивает высокую точность получаемого прогноза, так как базируется на методах математической статистики, что в совокупности подтверждает актуальность проведенных исследований.

## 1. Постановка и алгоритм решения задачи

Для полноценного функционирования различных сложных систем требуется проводить мониторинг их текущего состояния и прогнозирование будущего. Примером сложных систем могут служить программные комплексы, установленные на борту КА.

Корректное функционирование и жизненный цикл КА напрямую связаны с мониторин-

[1] Работа выполнена в рамках СЧ НИР шифр «Мониторинг-СГ-1.2.5.1» по Программе Союзного государства «Разработка космических и наземных средств обеспечения потребителей России и Беларуси информацией дистанционного зондирования Земли» и при частичной поддержке Российского фонда фундаментальных исследований (проект № 15-29-06945-офи_м «Развитие моделей, методов и программных средств обработки мультиспектральных снимков, видео-потоков и данных телеметрии для задач космического мониторинга арктической зоны»).

гом параметров его бортовой аппаратуры. Различные внешние обстоятельства могут спровоцировать нарушение работы КА, например, изменение космической погоды. В период солнечной активности космическая радиация может вывести из строя всю электронику на борту. На основе данных телеметрии, полученных с этого или однотипных КА в течение достаточно длительного времени, можно спрогнозировать будущее состояние КА, что позволит поддерживать его работоспособность при последующих критических ситуациях [6].

Таким образом, прогнозирование позволяет решить проблему противодействия различным факторам, стремящихся нарушить функционирование сложных систем.

Будем предполагать, что состояние КА определяется данными телеметрии, которые представляют собой непрерывный стационарный марковский процесс, в котором любое состояние КА в будущем зависит только от текущего состояния и не зависит от того, как он пришел в него. При этом условные вероятности перехода не меняются со временем, что позволяет использовать накопленные данные телеметрии для прогноза аномальных (т.е. выходящих за пределы диапазона допустимых значений) состояний.

Задача прогнозирования состояния сложных систем может быть решена в случае, когда известно распределение вектора случайных величин, состоящего из значения и его производной [7]. Тогда по значению функции распределения на границе аномальной зоны можно оценить вероятность перехода в аномальное состояние за заданный период времени. Однако можно повысить точность прогноза, учитывая текущее состояние системы. В этом случае необходимо найти множество условных распределений этого вектора.

Такую задачу предлагается решить с помощью байесовского подхода. Введем некоторые необходимые обозначения:

1) $V_i$ $(i = 1, \ldots, n)$ – множества состояний системы, для которой производится прогнозирование;

2) $x(t)$ – показание датчика в момент времени $t$;

3) $x'(t)$ – рассчитанная производная $x(t)$;

4) $d_1, d_2, \ldots, d_m$ – уровни опасности (УО);

5) $\Delta t_1 < \Delta t_2 < \ldots < \Delta t_m$ – прогнозные границы (временные интервалы), соответствующие уровням опасности;

6) $x^*$ – пороговое значение, превышение которого считается аномалией;

7) $dt$ – шаг дискретизации времени;

8) $T$ – интервал времени для расчетов рядов распределения вероятности появления аномалий;

9) $p^*$ – критическое значение вероятности аварийного состояния.

Допустим, имеется система, которая передает некоторое значение $x(t)$ каждые $dt$ минут. По полученным значениям можно рассчитать вектор показателей $(x(t), x'(t))$, определяющих состояние системы $V_i$ $(i = 1, \ldots, n)$. Требуется определить вероятность того, что при текущем значении $x(t)$, наступит аварийное состояние за назначенные временные интервалы, называемые прогнозными границами и обозначаемые как $\Delta t$. Понятно, что чем больше прогнозная граница, тем выше вероятность того, что за это время хоть раз, но наступит аварийное состояние. Выделим несколько прогнозных границ и зададим $p^*$. Чем меньше $\Delta t$, при котором достигается значение вероятности $P(x(t + \Delta t) > x^*) > p^*$, тем более опасно текущее состояние.

Произвольно определим несколько граничных значений $\Delta t$, разделяющих разные уровни опасности. Тогда состояние системы соответствует $j$-му уровню опасности, если $j = min\ (P(\exists\ \tau \in [t, t + \Delta t_j]: x(t + \tau) > x^*) > p^*)$. В итоге, в ходе реализации байесовского подхода каждому состоянию системы ставится в соответствие уровень опасности и прогнозная граница, в течение которой $x(t)$ превысит пороговое значение $x^*$. Уровни опасности, таким образом, ранжируются по величине $\Delta t_j$ – чем меньше значение $\Delta t_j$, тем выше УО.

Применение байесовского подхода в решении задачи прогнозирования подразумевает вычисление уровня опасности текущего состояния по алгоритму, который можно разделить на две части: обучающую и основную. Вычисление УО требует обучения алгоритма реализации байесовского подхода. Для начала необходимо задать некоторые параметры:

1) уровни опасности $d_1, d_2, \ldots d_m$;

2) прогнозные границы $\Delta t_1 < \Delta t_2 < \ldots < \Delta t_m$, соответствующие уровням опасности;

3) пороговое значение $x^*$;

4) критическое значение вероятности аварийной ситуации $p^*$;

5) интервалы времени $dt$ и $T$.

После определения начальных параметров проводится обучение алгоритма анализа потока данных.

**Обучающая часть алгоритма**

**Шаг 1**. Вычисление количества точек данных:

1) для обучающей выборки на интервале $T$

$$N = T / dt;$$

2) для каждого интервала $\Delta t_j$

$$n_j = \Delta t_j / dt.$$

**Шаг 2.** Вычисление множества состояний $V_i$ на обучающей выборке:

1) определить счетчики $s_i$ и $z_{ij}$ с нулевым начальным значением;

2) в цикле $t = 1, \ldots, N$:

а) рассчитать значение производной $x'(t) = (x(t) - x(t - dt) ) / dt;$

б) определить, к какому состоянию $V_i$ относится вектор $(x(t), x'(t))$, соответствующий этому состоянию счетчик $s_i$ увеличить на 1;

в) во вложенном цикле $\theta = 1, \ldots, n_m$ (т.е. проход по всем точкам на оси времени до максимального значения $n_m$): если $x(t + \theta)$ превышает пороговое значение $x^*$, то для всех $j$, для которых $n_j > \theta$, счетчики $z_{ij}$ увеличить на 1 и выйти из цикла.

**Шаг 3.** Расчет вероятностей:

1) на интервале $T$ рассчитываются вероятности событий:

а) событие $A_i$: вектор $(x(t), x'(t))$ соответствует состоянию $V_i$;

б) событие $A_i \cdot B_j$ (где событие $B_j$: наблюдается $j$-ый уровень опасности):

$$P(A_i) = s_i / N, P(A_i \cdot B_j) = z_{ij} / N;$$

2) расчет вероятности опасности $\rho\ (i, j)$ для каждого значения $i = 1, \ldots, n$ и каждого значения $j = 1, \ldots, m$:

$$\rho\ (i, j) = P(A_i \cdot B_j) / P(A_i).$$

**Шаг 4**. Определение уровня опасности для каждого вектора $(x(t), x'(t))$.

Элементы полученной матрицы $\rho$ сравниваются с заданной вероятностью $p^*$, и для каждого $i$ отмечается уровень опасности. Таким образом, каждому $i$-му значению вектора $(x(t), x'(t))$ ставится в соответствие уровень опасности перехода $x$ за пороговое значение $x^*$.

Результатом обучения является вычисление уровня опасности $d_j$ для каждого состояния $V_i$ и прогнозных границ $\Delta t_j$, определяющих превышение $p^*$. После обучения алгоритм работает с неизвестными данными, т.е. данные поступают в режиме реального времени.

**Основная часть алгоритма**

**Шаг 1.** Получение значения $x(t)$ от источника данных.

**Шаг 2.** Вычисление производной $x'(t)$.

**Шаг 3.** Определение, к какому состоянию множества $V_i$ относится вектор $(x(t), x'(t))$.

**Шаг 4.** В зависимости от результата, полученного на предыдущем шаге, вектору $(x(t), x'(t))$ присваивается соответствующий уровень опасности.

## 2. Экспериментальные исследования. Анализ данных телеметрии

Для проведения экспериментов был использован поток телеметрических данных с КА «Юбилейный» (малый спутник, созданный ОАО «ИСС» им. академика М.Ф. Решетнева совместно с группой российских космических предприятий и высших учебных заведений) [8]. «Юбилейный» каждую минуту передает данные о напряжении, температуре, силе тока и других параметрах. Фрагмент данных телеметрии приведен в Табл. 1.

Табл. 1. Поток телеметрических данных с трех датчиков

| Время | Значения с датчиков КА | | |
|---|---|---|---|
| | Bort_V | Bort_A | Sun_V |
| 12:08:44 | 14,6 | 0,16 | 0,10 |
| 12:09:44 | 14,6 | 0,19 | 0,10 |
| 12:10:44 | 14,5 | 0,19 | 0,29 |
| 12:11:44 | 14,4 | 0,19 | 0,96 |
| 12:12:44 | 13,9 | 0,16 | 14,21 |

Рассмотрим фрагмент телеметрического потока данных на примере датчика Bort_V (Рис. 1). Пороговое значение $x^*$ для данного датчика задано равным 14,8. Значения, которые превышают данный порог, считаются аномальными.

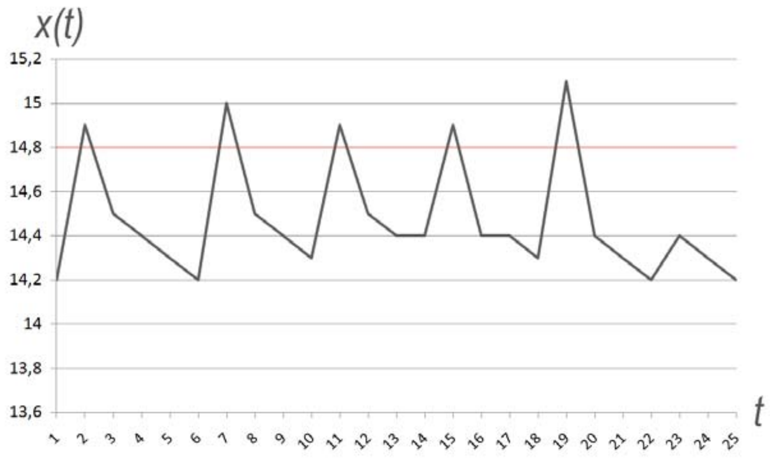


Рис. 1. График отрезка телеметрических данных

Для реализации байесовского подхода требуется задать некоторые параметры:

1) $dt = 1$;

2) $T = 3000$, тогда $N = T / dt = 3000$;

3) пороговое значение $x^* = 14{,}8$;

4) три прогнозные границы (в минутах): $t_1 = 20$, $t_2 = 30$, $t_3 = 50$;

5) заданным прогнозным границам приведем в соответствие уровни опасности: $d_1 = 3$ (высокий УО), $d_2 = 2$ (средний УО), $d_3 = 1$ (низкий УО).

Фрагмент результата работы алгоритма при заданных параметрах приведен в Табл. 2 и на Рис. 2.

Таким образом, получается, что, например, для *x(3)* = 14,5 уровень опасности – высокий, т.е. в течение 20 минут *x(t)* превысит пороговое значение $x^* = 14{,}8$. Действительно, для *x(3)* = 14,5 прогноз подтвердился через 4, 8, 12 и т.д. минут: *x(7)* = 15, *x(11)* = 14,9, *x(15)* = 14,9. Точность прогноза составила около 86%.

Табл.2. Фрагмент результата работы алгоритма

| t | x(t) | $d_j$ |
|---|---|---|
| 1 | 14,2 | 0 |
| 2 | 14,9 | 3 |
| 3 | 14,5 | 3 |
| 4 | 14,4 | 3 |
| 5 | 14,3 | 3 |
| 6 | 14,2 | 3 |
| 7 | 15 | 3 |
| 8 | 14,5 | 3 |
| 9 | 14,4 | 3 |
| 10 | 14,3 | 3 |
| 11 | 14,9 | 1 |
| 12 | 14,5 | 3 |
| 13 | 14,4 | 3 |
| 14 | 14,4 | 2 |
| 15 | 14,9 | 1 |
| 16 | 14,4 | 3 |
| 17 | 14,4 | 2 |
| 18 | 14,3 | 3 |
| 19 | 15,1 | 0 |
| 20 | 14,4 | 0 |
| 21 | 14,3 | 3 |
| 22 | 14,2 | 3 |
| 23 | 14,4 | 0 |
| 24 | 14,3 | 3 |
| 25 | 14,2 | 3 |

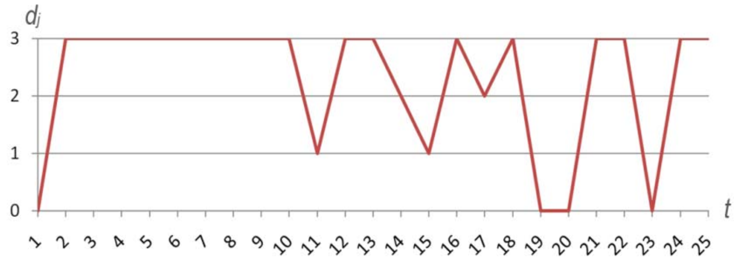


Рис. 2. Уровни опасности для датчика Bort_V

## Заключение

Полученные результаты показали, что байесовский подход может успешно использоваться в системах диагностирования и прогнозирования. Универсальность этого подхода позволяет использовать его в различных областях. Предполагается, что в силу универсальности, предлагаемый подход будет апробирован на ряде других актуальных задач.

## Литература


1. Красоткина О.В., Попов В.А., Нгуен Т.Ч., Моттль В.В. Байесовский подход к оцениванию факторов риска в анализе продолжительности жизни // Известия ТулГУ. Технические науки, №2, 2013. URL: http://cyberleninka.ru/article/n/bayesovskiy-podhod-k-otsenivaniyu-faktorov-riska-v-analize-prodolzhitelnosti-zhizni (дата обращения: 17.03.2016).
2. Enrique de Alba, Manuel Mendoza, Bayesian Forecasting Methods for Short Time Series [Электронный ресурс] // Instituto Technologico Autonomo de Mexico, 2016. URL: http://allman.rhon.itam.mx/~mendoza/Foresight.pdf (дата обращения: 17.03.2016).
3. Phillip M. Yelland, Shinji Kim, Renée Stratulate, A Bayesian Model for Sales Forecasting at Sun Microsystems [Электронный ресурс] // Interfaces 40, 2 (March 2010), 118-129. URL: http://dx.doi.org/10.1287/inte.1090.0477 (дата обращения: 17.03.2016).
4. Шаров В.Д. Применение байесовского подхода для уточнения вероятностей событий в автоматизированной системе прогнозирования и предотвращения авиационных происшествий // УБС, №43, 2013. URL: http://cyberleninka.ru/article/n/primenenie-bayesovskogo-podhoda-dlya-utochneniya-veroyatnostey-sobytiy-v-avtomatizirovannoy-sisteme-prognozirovaniya-i (дата обращения: 17.03.2016).
5. Благовещенский Ю.Н., Кречетова М.Ю., Сатаров Г.А. Экспертно-статистический байесовский подход к сценарному политическому прогнозированию. – Полис, №4, 2012, с. 74-98. URL: http://www.politstudies.ru/files/File/2012/4/7.pdf (дата обращения: 17.03.2016).
6. Баскаков С.И. Радиотехнические цепи и сигналы. – М.: Высшая школа, 2000. – 462 с.
7. Прохоров С.А. Математическое описание и моделирование случайных процессов. – Самарский гос. аэрокосм. ун-т, 2001. – 209 с.
8. Малый космический аппарат «Юбилейный» // Студенческий центр управления полетами СибГАУ. URL: http://sat.sibsau.ru/index.php?option=com_content&view=section&layout=blog&id=3&Itemid=2 (дата обращения: 08.03.2016).


**Амелькин Сергей Анатольевич.** Руководитель Исследовательского центра системного анализа ИПС им. А.К. Айламазяна РАН, первый проректор ЧОУ ВО «УГП им. А.К. Айламазяна». Окончил Московский технологический институт пищевой промышленности в 1991 году. Кандидат технических наук. Область научных интересов: оптимальное управление, системный анализ. E-mail: sergey.a.amelkin@gmail.com

**Шустова Мария Вениаминовна.** Инженер-исследователь Исследовательского центра процессов управления ИПС им. А.К. Айламазяна РАН. Окончила ЧОУ ВО «УГП им. А.К. Айламазяна» в 2015 году. Область научных интересов: информационные технологии, методы искусственного интеллекта, защита компьютерных сетей.
E-mail: m.v.shustova@gmail.com