

УДК 332.14; 338.2; 355

doi: 10.53816/20753608_2025_4_44

МЕТОДИКА АКТУАЛИЗАЦИИ ПРОГНОЗНОЙ ИНФОРМАЦИИ НА ОСНОВЕ БАЙЕСОВСКОГО ПОДХОДА

THE METHODOLOGY OF UPDATING PREDICTIVE INFORMATION BASED ON THE BAYESIAN APPROACH

По представлению академика РАН Р.А. Дурнева

А.Л. Афанасьев

ВНИИ «Центр»

A.L. Afanasyev

Представлена методика актуализации прогнозной информации, в основу которой положено применение формулы Байеса для уточнения апостериорной вероятности прогнозируемых событий по мере поступления новых данных, свидетельствующих об их правдоподобии. Приведено содержание этапов реализации методики с практическими примерами для количественных и качественных оценок. Предложены варианты визуализации результатов прогнозирования и их интерпретация.

Ключевые слова: актуализация, прогнозная информация, формула Байеса, априорная вероятность, апостериорная вероятность, условие правдоподобия.

A methodology for updating predictive information is presented, which is based on the application of the Bayes formula to clarify the a posteriori probability of predicted events as new data becomes available indicating their likelihood. The content of the stages of the methodology implementation with practical examples for quantitative and qualitative assessments is given. Variants of visualization of forecasting results and their interpretation are proposed.

Keywords: updating, predictive information, Bayes formula, prior probability, posterior probability, likelihood condition.

Вводная часть

Актуализация прогнозной информации предполагает ее уточнение (обновление или корректировку) на основе результатов мониторинга, анализа и оценки с учетом вновь поступивших данных о некоторых событиях, поддерживающих вероятность осуществления выдвинутых гипотез (прогнозируемых событий).

В основу методики актуализации прогнозной информации предлагается положить байе-

совский подход (байесовскую статистику) к формализации неопределенностей при прогнозировании будущих событий [1–5].

Байесовский подход является одним из возможных способов формализации неопределенностей, базирующийся на тезисе о том, что степень уверенности в некотором прогнозируемом событии качественного или количественного характера увеличивается либо корректируется по мере пополнения имеющейся информации о данном событии. При этом применяется субъективная

интерпретация вероятности и соотношение априорной (до наступления некоторого события) и апостериорной (при наступлении некоторого события) вероятностей согласно формуле Байеса.

Основная часть

В классическом виде формула Байеса имеет следующий вид:

$$P(H_i/A_j) = \frac{P(A_j/H_i)P(H_i)}{\sum_{i=1}^N P(A_j/H_i)P(H_i)},$$

$$i = \overline{1, N}, j = \overline{1, M} \quad (1)$$

где $P(H_i/A_j)$ — апостериорная вероятность как степень уверенности эксперта в гипотезе H_i с учетом события (свидетельства) A_j ;

$P(H_i)$ — априорная вероятность — убежденность в гипотезе H_i до предъявления данных о событии A_j ;

$P(A_j/H_i)$ — условная вероятность события A_j , если имеет место гипотеза H_i (условие правдоподобия);

N — число прогнозируемых гипотез.

M — число подтверждающих гипотезу событий.

Если предположить, что события A_j независимые и проявляются одновременно, то формула Байеса будет иметь следующий вид:

$$P(H_i/A_1, A_2, \dots, A_M) =$$

$$= \frac{P(A_1/H_i)P(A_2/H_i) \dots P(A_M/H_i)P(H_i)}{\sum_{i=1}^N P(A_1/H_i)P(A_2/H_i) \dots P(A_M/H_i)P(H_i)}. \quad (2)$$

Для оценки апостериорной вероятности гипотезы H_i — $P(H_i/A_j)$ требуются исходные данные, представленные в табл. 1.

Таким образом, байесовский подход предполагает актуализацию прогнозной информации путем уточнения апостериорной вероятности $P(A_j/H_i)$ по мере поступления статистической и эмпирической информации (например, за счет привлечения дополнительных источников информации или экспертов).

Апостериорная вероятность гипотезы H_i будет повышаться в случае ее поддержки

событиями A_j и снижаться в противном случае. По мере поступления новых данных о событиях A_j апостериорные вероятности гипотез применяются в качестве априорных для следующего шага уточнения прогнозной информации.

Методика включает следующие этапы.

Этап 1. Постановка задачи актуализации прогнозной информации.

Данный этап предполагает определение состава прогнозируемых гипотез или значащих переменных. В первом случае речь идет о событийном прогнозировании, когда под гипотезами понимается какое-либо событие либо множество событий. Пример бинарной гипотезы — «Появление человекоподобных роботов на поле боя к 2030 году», значения бинарной переменной: 1 — да, 0 — нет. Пример множественной гипотезы «Срок достижения гиперзвуковыми ракетами скорости более 12 махов», возможные значения переменной: «к 2030 году», «к 2035 году», «к 2040 году».

Во втором случае в качестве значащей переменной прогноза рассматривается какой-либо количественно измеряемый параметр, например, уровень микроминиатюризации, измеряемый плотностью упаковки элементов микросхем (например, на сегодня он составляет 26 млрд транзисторов на 10 нм).

Помимо формирования гипотез на этом этапе определяется множество факторов (данных), которые обуславливают правдоподобие этих гипотез. Примерами таких факторов A_j является рост:

- уровня ежегодного финансирования A_1 ;
- уровня научно-технологического задела A_2 ;
- количества успешных испытаний A_3 ;
- патентной и публикационной активности A_4 ;

Таблица 1

Исходные данные для актуализации прогнозной информации

$P(H_1)$	$P(H_2)$...	$P(H_N)$
$P(A_1/H_1)$	$P(A_1/H_2)$...	$P(A_1/H_N)$
$P(A_2/H_1)$	$P(A_2/H_2)$...	$P(A_2/H_N)$
...
$P(A_M/H_1)$	$P(A_M/H_2)$...	$P(A_M/H_N)$

- уровня кадровой обеспеченности (научных коллективов) A_5 ;
- обеспеченности материалами и веществами A_6 ;
- индекса международной конкуренции A_7 .

Этап 2. Сбор и предварительная обработка данных.

На этом этапе осуществляется сбор данных о ретроспективных, текущих и планируемых показателях факторов, обуславливающих реализацию гипотез.

Предварительная обработка предусматривает нормализацию этих показателей в интервале от 0 до 1, т.е. представление доли показателей событий-факторов относительно требуемых значений. Например, $A_1 = 0,7$ означает уровень финансирования в 70 % относительно требуемого.

В случае невозможности количественной оценки уровней показателей событий-факторов может быть использована модифицированная вербально-числовая шкала Харрингтона (табл. 2), применяемая для приведения в соответствие объективных значений показателей к субъективным предпочтениям экспертов. Например, $A_6 = 0,7$ соответствует высокому уровню патентной и публикационной активности.

Статистический анализ ретроспективных значений показателей позволяет экспертам оценить априорную вероятность гипотез $P(H_i)$ и правдоподобия $P(H_i/A_j)$. Таким образом получаем исходные данные для актуализации прогнозной информации, представляемые в форме табл. 1.

Таблица 2

Модифицированная вербально-числовая шкала Харрингтона

Уровень показателя события-фактора	Числовые значения
Очень высокий	1,00
Высокий	0,80
Средний	0,65
Низкий	0,35
Очень низкий	0,20

Этап 3. Выбор модели и актуализации прогнозной информации.

Наиболее распространенными моделями прогнозирования в байесовской статистике являются байесовское моделирование, модели машинного обучения (регрессия, нейронные сети для анализа временных рядов) и сценарное моделирование с применением байесовских сетей доверия.

В данной методике используется байесовское моделирование, основанное на обновлении прогнозной информации (вероятностей гипотез) на актуальных данных в соответствии с выражением (1) при разновременном возникновении факторов или формула (2) при их одновременной реализации. Такое моделирование имеет высокую наглядность и интерпретируемость результатов.

Пример исходных данных для варианта бинарной гипотезы, приведенного выше, и четырех первых факторов, обуславливающих их правдоподобность, представлен в табл. 3.

Как видно из представленных данных, положительной реализации гипотезы соответствуют достаточно высокие уровни правдоподобий факторов.

Положим, что реализация факторов последовательно рассредоточена во времени. Тогда согласно выражению (1) актуализированная прогнозная информация относительно вероятностей реализации гипотезы «Появление человекоподобных роботов на поле боя к 2030 году» будет рассчитана в соответствии с формулами:

$$P(H_1 = 1 / A_1) = \frac{0,6 \cdot 0,4}{0,6 \cdot 0,4 + 0,4 \cdot 0,4} = 0,60;$$

Таблица 3

Исходные данные для примера актуализации прогнозной информации

$P(H_1 = 1) = 0,4$	$P(H_2 = 0) = 0,6$
$P(A_1 / H_1) = 0,6$	$P(A_1 / H_2) = 0,4$
$P(A_2 / H_1) = 0,7$	$P(A_2 / H_2) = 0,4$
$P(A_3 / H_1) = 0,6$	$P(A_3 / H_2) = 0,3$
$P(A_4 / H_1) = 0,7$	$P(A_4 / H_2) = 0,3$

$$P(H_1 = 1 / A_1, A_2) = \frac{0,7 \cdot 0,6}{0,7 \cdot 0,6 + 0,4 \cdot 0,6} = 0,64;$$

$$P(H_1 = 1 / A_1, A_2, A_3) = \frac{0,6 \cdot 0,64}{0,6 \cdot 0,64 + 0,3 \cdot 0,64} = 0,67;$$

$$P(H_1 = 1 / A_1, A_2, A_3, A_4) = \frac{0,7 \cdot 0,67}{0,7 \cdot 0,67 + 0,3 \cdot 0,67} = 0,70.$$

Таким образом, с учетом вновь поступившей информации вероятность реализации гипотезы повысилась с 0,40 (40 %) до 0,70 (70 %).

Анализ полученных результатов показал, что наибольший вклад в прирост апостериорной вероятности вносит фактор уровня ежегодного финансирования A_1 : с 0,4 до 0,6 (50 %).

Полученные оценки имеют точечный характер, представляют собой среднее значение апостериорного распределения вероятности и в значительной степени не учитывают факторы неопределенностей.

По аналогии с частотным подходом байесовская статистика учитывает разброс возможных значений прогнозных параметров вероятностными (байесовскими) интервалами, которые имеют прямую интерпретацию как вероятность (например, 95 %) нахождения параметра в заданном диапазоне. В отличие от частотных доверительных интервалов байесовские интервалы интерпретируются интуитивно как то, что с вероятностью 95 % истинное значение параметра лежит в этом интервале.

Для нормального распределения границы байесовского интервала $\bar{P}^{\min}(H/D)$, $\bar{P}^{\max}(H/D)$ рассчитываются в соответствии с выражением:

$$\bar{P}^{\min(\max)}(H/D) = P(H/D) \pm z \times \sqrt{\frac{P(H/D) \cdot (1 - P(H/D))}{n}},$$

где $P(H/D)$ — апостериорная вероятность гипотезы H ;

z — квантиль стандартного нормального распределения ($z = 1,96$ для 95 % интервала);

n — объем условных наблюдений (для примера $n = 50$).

Тогда для вышеприведенного примера байесовский интервал составит:

$$\bar{P}(H_1 = 1 / A_1, A_2, A_3, A_4) = [0,57; 0,83].$$

Интервальное представление вероятности гипотез позволяет более корректно учитывать неопределенности при прогнозировании будущих событий.

Этап 4. Сценарный анализ.

На предыдущем этапе были получены точечные (средние) и интервальные оценки апостериорных вероятностей гипотез.

Сценарный анализ предполагает рассмотрение нескольких альтернативных вариантов сценариев, которые могут реализоваться в будущем. Каждый сценарий предусматривает изменение нескольких параметров одновременно.

Если ранее рассмотренный пример оценивать, как базовый, то варианты пессимистичного и оптимистичного сценариев могут быть представлены в виде табл. 4 и 5.

Для пессимистичного сценария среднее значение вероятности реализации гипотезы составит 0,55, байесовский интервал — $[0,41; 0,68]$, для оптимистического соответственно 0,89 и $[0,80; 0,98]$.

Результаты сценарного анализа могут быть положены в основу принимаемых решений, например, по наращиванию объемов финансирования разработок и созданию научно-технического задела.

Таблица 4

Исходные данные для пессимистичного сценария

$P(H_1 = 1) = 0,4$	$P(H_2 = 0) = 0,6$
$P(A_1 / H_1) = 0,5$	$P(A_1 / H_2) = 0,4$
$P(A_2 / H_1) = 0,5$	$P(A_2 / H_2) = 0,3$
$P(A_3 / H_1) = 0,5$	$P(A_3 / H_2) = 0,4$
$P(A_4 / H_1) = 0,6$	$P(A_4 / H_2) = 0,5$

Таблица 5

Исходные данные для оптимистичного сценария

$P(H_1 = 1) = 0,4$	$P(H_2 = 0) = 0,6$
$P(A_1 / H_1) = 0,7$	$P(A_1 / H_2) = 0,3$
$P(A_2 / H_1) = 0,7$	$P(A_2 / H_2) = 0,2$
$P(A_3 / H_1) = 0,8$	$P(A_3 / H_2) = 0,2$
$P(A_4 / H_1) = 0,8$	$P(A_4 / H_2) = 0,1$

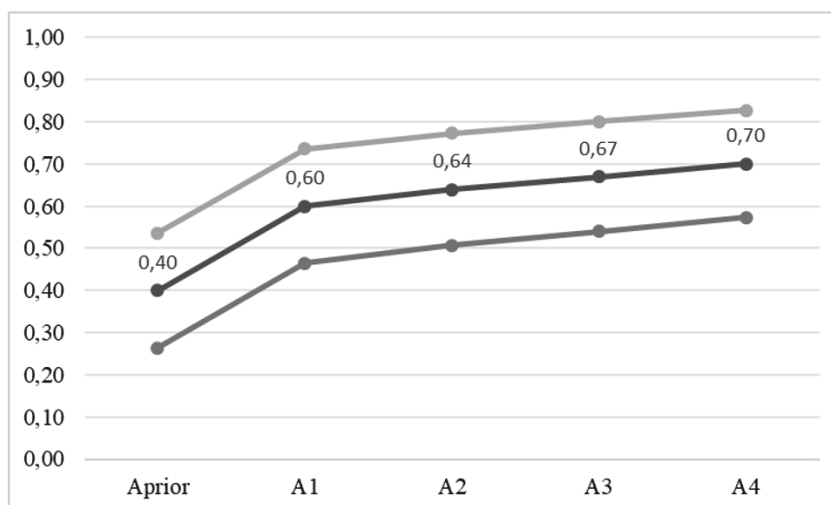


Рис. 1. Точечные оценки и доверительные интервалы вероятности

Этап 5. Анализ устойчивости результатов прогнозирования.

На этом этапе исследуется, каким образом изменения входных параметров оказывают влияние на результаты прогнозирования. Для этого с определенным шагом задается изменение значений исследуемого параметра и фиксируются изменения апостериорной вероятности гипотез.

Проведенная оценка устойчивости результатов прогнозирования при последовательном снижении вероятностей правдоподобия по каждому из факторов на величину 0,1 показала, что падение при этом априорной вероятности не превышает 4 %. Отсюда можно сделать вывод о том, что предложенная модель устойчива к небольшим изменениям параметров правдоподобия.

Этап 6. Визуализация и интерпретация результатов прогнозирования.

На данном этапе осуществляется визуальное представление результатов прогнозирования, позволяющее, в том числе интерпретировать их исходя из динамики поступления обновленных данных и корректировки в этой связи апостериорных вероятностей будущих событий.

В качестве вариантов визуализации результатов прогнозирования целесообразно использовать:

- кривые прогнозируемых точечных оценок и доверительных интервалов вероятности реализации гипотезы (события) в зависимости от вновь поступивших данных о факторах (рис. 1);
- карту вероятностных параметров для разных вариантов сценариев реализации гипотезы (рис. 2).

Интерпретация результатов прогнозирования предполагает формирование на основе их анализа рекомендаций по корректировке значений факторов для реализации наиболее приемлемой гипотезы.

Примеры интерпретации результатов прогнозирования:

- при текущих темпах научно-технологического развития гипотеза H имеет высокую вероятность (72 %) реализации в течение 5 лет;
- основными драйверами реализации гипотезы H являются наращивание научно-технологического задела и увеличение числа испытаний;

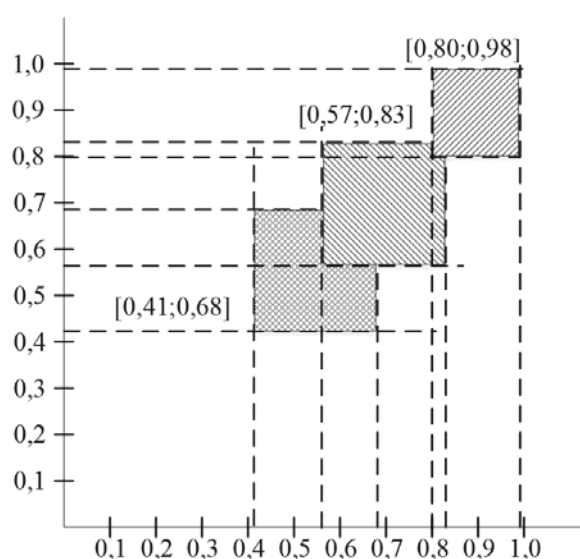


Рис. 2. Карта вероятностей для вариантов сценариев

– оптимистичный вариант сценария реализации гипотезы предполагает повышение уровня ежегодного финансирования разработок на 20–25 %;

– при реализации пессимистичного варианта сценария вероятно смещение срока реализации гипотезы на 2–3 года.

Выводы

Преимущества подхода, реализованного в предложенной методике, являются:

– корректный учет неопределенностей через апостериорные распределения;

– возможность обновления прогноза при поступлении новых данных (например, после очередного испытания);

– гибкое моделирование взаимодействия факторов.

Байесовская модель особенно полезна в условиях неполной информации, она позволяет количественно обновлять прогнозы и реализовывать последовательное их уточнение на основе экспертных знаний и по мере поступления новых данных, что обеспечивает возможность широкого применения технологий искусственного интеллекта в процессе научно-технологического прогнозирования [6–8].

В случае актуализации интервального прогноза с применением байесовского подхода возможны следующие варианты его изменения:

– уменьшение или увеличение доверительного интервала относительно его среднего значения (как симметричное, так и асимметричное);

– изменение значения одной из границ доверительного интервала относительно его среднего значения;

– смещение доверительного интервала относительно оси времени влево или вправо без изменения его ширины;

– смещение доверительного интервала относительно оси времени влево или вправо с одновременным изменением его ширины.

Список источников

1. Моррис У.Т. Наука об управлении. Байесовский подход. М.: Изд-во «Мир», 1971. 304 с.

2. Хей Дж. Введение в методы байесовского статистического вывода. М.: Финансы и статистика, 1987. 336 с.

3. Прокопчина, С.В. Разработка методов и средств байесовской интеллектуализации измерений в задачах мониторинга сложных объектов: дис. ... д-ра техн. наук: 05.11.16. СПб., 1995. 450 с.

4. Акимов В.А., Мишурный А.В., Якимюк О.В. Прогнозно-аналитические решения по природным, техногенным и биологическим угрозам: монография; под ред. А.П. Чуприяна; М.: ФГБУ ВНИИ ГОЧС (ФЦ), 2022. 316 с.

5. Тулупьев А.Л., Николенко С.И., Сироткин А.В. Байесовские сети: Логико-вероятностный подход. СПб.: Наука, 2006. 607 с.

6. Афанасьев А.Л., Курицын А.В., Голубев С.С. Проблемы развития цифровых технологий для Арктической зоны Российской Федерации / Мировая экономика в новых условиях развития: готовность к ответу на вызовы // Материалы Международной научно-практической конференции; под ред. А.А. Ефремова. 2019. С. 201–205.

7. Афанасьев А.Л., Курицын А.В., Голубев С.С. Глобальные тенденции научно-технологического развития России и их использование в интересах обеспечения обороны и безопасности государства; сб. статей XIX Нац. науч. конф. с между. участием «Россия: тенденции и перспективы развития». 2020. № 15. С. 39–46.

8. Афанасьев А.Л., Курицын А.В., Голубев С.С. Применение технологий искусственного интеллекта в научно-технологическом прогнозировании // Прикладная информатика. 2022. Т. 17, № 4 (100). С. 57–74.