

ОБЗОР ОЦЕНОК КАЧЕСТВА МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ

Щербаков М. В.¹,

*(Волгоградский государственный технический универси-
тет, Волгоград, Россия)*

Бребельс А.²

(Высшая католическая школа Кемпена, Гиль, Бельгия)

Щербакова Н.Л.³, Тюков А.П.⁴,

*(Волгоградский государственный технический универси-
тет, Волгоград, Россия)*

В статье дан обзор оценок качества количественных моделей прогнозирования – ошибок прогнозирования. Выделены группы оценок, приведены формулы расчета и описание недостатков. Даны рекомендации по выбору критериев оценки ошибок при решении задачи прогнозирования или задачи управления с прогнозирующими моделями.

Ключевые слова: прогнозирование, качество прогнозирования, ошибка прогнозирования.

¹ Максим Владимирович Щербаков, кандидат технических наук, доцент кафедры САПР и ПК, ВолгГТУ, (Волгоград, пр. Ленина, 28, тел (8442) 248100) maxim.shcherbakov@gmail.com.

² Адриаан Бребельс, доцент, Высшая католическая школа Кемпена (Kleinhoefstraat 4 2440 Geel, Belgium), adriaan.brebels@khk.be.

³ Наталия Львовна Щербакова, кандидат технических наук, доцент кафедры ЭФ, ВолгГТУ (Волгоград, пр. Ленина, 28, тел (8442) 248100) natalia.shchrbakova@gmail.com.

⁴ Антон Павлович Тюков, аспирант, (Волгоград, пр. Ленина, 28, тел (8442) 248100), anton.tyukov@gmail.com.

1. Введение

Для оценки качества прогнозирования используются различные критерии: оценки ошибок прогнозирования, скорость расчета, интерпретируемость и другие [5, 16, 21]. Оценка ошибки прогнозирования (forecast error measures) или точности прогноза (forecast accuracy) являются наиболее важными при решении практических задач [21]. Как правило, в исследованиях используют типовые оценки ошибки прогнозирования, на основе которых формируется либо суждение о качестве предлагаемого механизма прогнозирования, либо осуществляется выбор наилучшего механизма прогнозирования для нескольких объектов. При решении подобных задач в первую очередь используются «традиционные» критерии, свойственные рассматриваемой предметной области, несмотря на существующие в них недостатки.

В данной статье приведены результаты анализа существующих и достаточно распространенных оценок ошибок прогнозирования, используемых в исследованиях по прогнозированию [5, 6, 11, 12, 20]. Для каждой оценки приводятся: формулы ее расчета, описание недостатков, названия оценок на английском языке и соответствие с названиями, принятыми в отечественной литературе.

2. Оценки ошибок моделей прогнозирования

2.1. АБСОЛЮТНЫЕ ОШИБКИ ПРОГНОЗИРОВАНИЯ

Первая группа базируется на расчете абсолютных ошибок и включает в себя оценки, основанные на вычислении величины e_t

$$(1) \quad e_t = (y_t - f_t^{(m)}),$$

где y_t – измеренное значение в момент времени t , $f_t^{(m)}$ – прогнозное значение в момент времени t , полученное в результате

использования исследуемой модели прогнозирования m . Далее по тексту индекс модели (m) будем опускать.

Средняя абсолютная ошибка (Mean Absolute Error, MAE) рассчитывается по формуле:

$$(2) \quad MAE = \frac{1}{n} \sum_{i=1}^n |e_i| = \underset{i=1,n}{mean} |e_i|,$$

где n – размер горизонта прогнозирования (forecast horizon), $mean(\bullet)$ – операция нахождения среднего.

Медиана абсолютной ошибки (Median Absolute Error, MdAE) рассчитывается согласно формуле

$$(3) \quad MdAE = \underset{i=1,n}{median} |e_i|,$$

где $median(\bullet)$ – операция нахождения медианы.

Квадрат средней квадратичной ошибки (Mean Square Error, MSE) рассчитывается по формуле

$$(4) \quad MSE = \frac{1}{n} \sum_{i=1}^n (e_i^2) = \underset{i=1,n}{mean} (e_i^2),$$

следовательно, средняя квадратичная ошибка (Root Mean Square Error, RMSE) рассчитывается согласно формуле:

$$(5) \quad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i^2)} = \sqrt{\underset{i=1,n}{mean} (e_i^2)}.$$

Кроме этого $RMSE$ может носить название среднеквадратичное отклонение прогнозных значений от измеренных, среднее значение квадрата ошибки прогноза. Данные критерии ошибок прогнозирования являются одними из самых распространенных [11, 12].

Рассмотрим недостатки этих оценок.

1. В качестве основного недостатка можно выделить зависимость критериев от шкалы измерения [12]. Следовательно, если стоит задача оценивания моделей прогнозирования на ряде объектов с различными шкалами измерения, то данные оценки неприменимы.

2. Другим недостатком предложенных методов является влияние на оценку выбросов в данных. Так, если в данных присутствует выброс, имеющий максимальное значение, что является распространенным случаем на практике [18], то рассчитанное значение ошибки может быть сильно завышенным. Использование медианы, дает более устойчивую оценку.

3. Оценки $RMSE$, MSE являются ненадежными: применимость данных критериев к различным порциям данных может значительно влиять на результат оценки (см. исследование [5]).

2.2. ПРОЦЕНТНЫЕ (ОТНОСИТЕЛЬНЫЕ) ОШИБКИ ПРОГНОЗИРОВАНИЯ

Процентные ошибки рассчитываются на основании значения p_t

$$(6) \quad p_t = \frac{|e_t|}{y_t}$$

и являются также одними из самых распространенных при решении задач прогнозирования. В отечественной литературе такие оценки называются относительной погрешностью [1, 2, 3], однако в этом случае возникает понятийное противоречие при анализе литературы по прогнозированию на английском языке. В группу процентных ошибок входят следующие оценки.

Средняя относительная ошибка прогноза или (если дословно переводить английское наименование) – средняя абсолютная процентная ошибка (Mean Absolute Percentage Error, MAPE)

$$(7) \quad MAPE = \frac{1}{n} \sum_{i=1}^n 100 \cdot |p_i| = \underset{i=1, n}{mean}(100 \cdot |p_i|),$$

Медиана абсолютной процентной ошибки (Median Absolute Percentage Error, MdAPE) является более устойчивой к выбросам и рассчитывается согласно формуле

$$(8) \quad MdAPE = \underset{i=1, n}{median}(100 \cdot |p_i|).$$

Средняя квадратичная процентная ошибка (Root Mean Square Percentage Error, RMSPE) вычисляется согласно:

$$(9) \text{ RMSPE} = \sqrt{\text{mean}_{i=1,n}(100 \cdot |p_i|)^2},$$

и медиана квадратичной процентной ошибки

$$(10) \text{ RMdSPE} = \sqrt{\text{median}_{i=1,n}(100 \cdot |p_i|)^2},$$

В качестве недостатков критериев данной группы следует отметить следующие.

1. Возникновение ошибки деления на ноль, в случае, когда реальное значение равно нулю.

2. Несимметричность оценок – значение ошибки различно, если прогнозное значение находится выше или ниже реального [10, 13, 17].

3. Значительное влияние на оценку выбросов – если выброс принимает значение, значительно превосходящее максимальное «нормальное» значение [5].

4. В работе [14] показано, что данная ошибка является смещенной. Это может привести к неадекватной оценке моделей прогнозирования.

2.3. СИММЕТРИЧНЫЕ ОШИБКИ ПРОГНОЗИРОВАНИЯ

Критерии, входящие в данную группу, рассчитываются на основании значения s_t , вычисляемого по формуле:

$$(11) s_t = \frac{|e_t|}{(y_t + f_t)}.$$

В группу входят следующие оценки. Симметричная средняя абсолютная процентная ошибка (Symmetric Mean Absolute Percentage Error, sMAPE) вычисляется согласно:

$$(12) \text{ sMAPE} = \frac{1}{n} \sum_{i=1}^n 200 \cdot |s_i| = \text{mean}_{i=1,n}(200 \cdot |s_i|),$$

и медиана средней абсолютной процентной ошибки

$$(13) \text{ sMdAPE} = \underset{i=1,n}{\text{median}}(200 \cdot |s_i|).$$

Во избежание проблем связанных с делением на ноль была предложена модифицированная sMAPE – Modified sMAPE , msMAPE, в знаменателе которой появился дополнительный член:

$$(14) \text{ msMAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - f_i|}{(|y_i| + |f_i|)/2 + S_i},$$

$$\text{где } S_i = \frac{1}{i-1} \sum_{k=1}^{i-1} |y_k - \bar{y}_{i-1}|, \bar{y}_{i-1} = \frac{1}{i-1} \sum_{k=1}^{i-1} y_k.$$

Развивая идею о включении дополнительного члена, были предложены более сложные оценки: KL-N, KL-N1, KL-N2, KL-DE1, KL-DE2, IQR (см. работу [9]).

Отметим следующие недостатки.

1. Несмотря на свое название, данная ошибка также является несимметричной [13].

2. Кроме этого, если реальное значение равно прогнозируемому, но с противоположным знаком, или оба этих значения равны нулю, то возникает ошибка деления на ноль.

3. Аналогично процентным ошибкам данные критерии подвержены влиянию выбросов.

4. При применении более сложных расчетов возникает проблема интерпретируемости вычисления оценок, что тормозит их распространение на практике [5].

2.4. ОШИБКИ, ОСНОВАННЫЕ НА ОЦЕНКЕ ПРОГНОЗА БАЗОВЫХ МОДЕЛЕЙ

Основой для расчета ошибок данной группы является величина r_t , определяемая следующим образом:

$$(15) r_t = \frac{|e_t|}{(y_t - f_t^*)},$$

где f_t^* – прогнозное значение, полученное с использованием базовой модели прогнозирования (benchmark model). Зачастую в качестве такой модели выступает наивная модель, для которой

$$(16) f_t^* = y_{t-l},$$

где l – величина лага и в большинстве исследований $l=1$.

В данную группу входят следующие оценки. Средняя относительная абсолютная ошибка (Mean Relative Absolute Error, MRAE) рассчитывается по формуле:

$$(17) MRAE = \overline{mean} |r_i|,$$

медиана относительной абсолютной ошибки (Median Relative Absolute Error, MdRAE), вычисляемая согласно:

$$(18) MdRAE = \overline{median} |r_i|$$

и геометрическая средняя относительная абсолютная ошибка (Geometric Mean Relative Absolute Error, GMRAE), которая вычисляется аналогично формуле (17), только вместо операции $\overline{mean}(\bullet)$ выполняется операция $\overline{gmean}(\bullet)$ – вычисления геометрического среднего.

Отметим недостатки данных оценок.

1. Как видно из формул (15-18) возможно деление на нуль, в случае если прогнозное значение базовой модели совпадает с реальным значением.

2. В случае, если выбрана наивная модель ошибка деления на нуль возникает при непрерывной последовательности одинаковых значений временного ряда.

2.5. МАСШТАБИРОВАННЫЕ ОШИБКИ ПРОГНОЗИРОВАНИЯ

В качестве базиса для расчета масштабированных ошибок используется величина q_t

$$(19) \quad q_t = \frac{|e_t|}{\frac{1}{n-1} \sum_{i=2}^n |y_i - y_{i-1}|}.$$

В данную группу входит средняя абсолютная масштабированная ошибка (Mean Absolute Scaled Error, MASE), предложенная в [12]. Она рассчитывается согласно:

$$(20) \quad MASE = \mathit{mean}_{i=1,n} |q_i|,$$

Медиана абсолютной масштабированной ошибки находится следующим образом:

$$(21) \quad MdASE = \mathit{median}_{i=1,n} |q_i|.$$

Еще одна оценка входящая в эту группу: среднеквадратическая масштабированная ошибка (Root Mean Square Scaled Error, RMSSE), вычисляемая по формуле [20]:

$$(22) \quad RMSSE = \sqrt{\mathit{mean}_{i=1,n} (q_i^2)}.$$

Данная оценка является симметричной и устойчивой к выбросам. Однако можно выделить 2 недостатка.

1. Если в интервале прогнозирования реальные значения равны друг другу, то возникает ошибка деления на ноль.

2. Кроме этого можно наблюдать слабое смещение оценки, если выполнять эксперименты по аналогии с работой [14].

2.6. ОТНОСИТЕЛЬНЫЕ ОЦЕНКИ

В данную группу входят оценки, рассчитываемые на основе отношения ошибок, рассмотренных выше для анализируемой модели прогнозирования и базовой модели.

Относительная средняя абсолютная ошибка (Relative Mean Absolute Error, RelMAE) вычисляется по формуле.

$$(23) \quad RMAE = \frac{MAE}{MAE^*},$$

где MAE и MAE^* средние абсолютные ошибки для анализируемой модели прогнозирования и базовой модели соответственно, вычисляемые по формуле (2).

Относительное среднеквадратичное отклонение (Relative Root Mean Square Error, RelRMSE) вычисляется аналогично (23), только в правой части вычисляются оценки по формуле (5):

$$(24) \quad RRMSE = \frac{RMSE}{RMSE^*}.$$

В некоторых ситуациях предлагают вычислять логарифм отношения (23), в этом случае оценка называется логарифм относительного среднеквадратичного отклонения (Log mean squared error ratio, LMR)

$$(25) \quad LMR = \log\left(\frac{RMSE}{RMSE^*}\right).$$

В работе [19] предлагается более сложная оценка относительного геометрического среднеквадратичного отклонения Relative Geometric Root Mean Square Error, RGRMSE.

Следующая группа оценок рассчитывает число случаев, когда ошибка исследуемой модели прогнозирования превышает значение ошибки базовой модели. К таким оценкам относятся $PB(MAE)$ – Percentage Better (MAE), вычисляемая по формуле:

$$(26) \quad PB(MAE) = 100\% \cdot \text{mean}\left(I\{MAE < MAE^*\}\right),$$

где $I\{\bullet\}$ – оператор, принимающий значение ноль или единица в соответствии с выражением:

$$(27) \quad I(MAE) = \begin{cases} 0, & \text{если } MAE < MAE^*; \\ 1, & \text{в противном случае.} \end{cases}$$

По аналогии может быть использована оценка квадрата средней квадратичной ошибки $PBMSE$ (Percentage Better (MSE)).

В качестве недостатков следует указать следующие.

1. Возникновение ошибки деления на ноль при условии если ошибка прогноза равна нулю.

2. Кроме этого критерии определяют число случаев, когда исследуемая модель прогнозирования превосходит базовую, но не определяет насколько.

2.7. ДРУГИЕ ГРУППЫ ОЦЕНОК

В данной группе собраны оценки, предложенные в различных исследованиях для устранения недостатков существующих и распространенных оценок.

Так для нивелирования недостатка связанного с зависимостью от шкалы предложена нормализованная среднеквадратическая ошибка (Normalized Root Mean Square Error $nRMSE$), вычисляемая по формуле:

$$(28) \ nRMSE = \frac{1}{\bar{y}} \sqrt{\overline{mean(e_i^2)}},$$

где $\bar{y}$ – фактор нормализации, который, как правило, равен либо максимальному измеренному значению на горизонте прогнозирования, либо разнице между максимальным и минимальным значением. Фактор нормализации может быть рассчитан либо на всем интервале горизонта прогнозирования, либо на заданном заранее коротком интервале наблюдения [4]. Однако такая оценка подвержена влиянию выбросов, т.к. выброс, имеющий значение значительно больше «нормального» максимального, будет занижать оценку. Для снижения влияния выбросов были предложены интегральные оценки Integral Normalized Mean Square Error [7], вычисляемые по формуле:

$$(29) \ inRSE = \frac{1}{\sum_{i=1}^n y_i} \sqrt{\overline{mean(e_i^2)}}.$$

В некоторых источниках оценка $nRMSE$ рассчитывается следующим образом [9]:

$$(30) \text{ inRSE} = \sqrt{\frac{\sum_{i=1}^n (e_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}},$$

где $\bar{y} = \frac{1}{n} \sum_{k=1}^n y_k$.

Другие нормализованные оценки получили название std_APE, std_MAPE [8, 15] и рассчитываются по формулам

$$(31) \text{ Std_AE} = \sqrt{\frac{\sum_{i=1}^n (e_i - MAE)^2}{n-1}}$$

и

$$(32) \text{ Std_APE} = \sqrt{\frac{\sum_{i=1}^n (p_i - MAPE)^2}{n-1}}$$

соответственно.

В качестве недостатка можно указать деление на ноль, в случае равенства нулю фактора нормализации.

3. Рекомендации к выбору оценки качества прогнозирования

Одним из самых сложных вопросов является вопрос выбора наиболее приемлемой оценки ошибки прогнозирования из представленного набора. В связи с тем, что каждый критерий обладает недостатками, которые могут приводить к неточному определению качества прогнозирования, невозможно выделить один универсальный критерий [16].

Предложены следующие рекомендации по выбору оценок.

1. Если оценивается качество прогноза временных рядов с одинаковыми единицами измерения и были выполнены проце-

дуры предварительного анализа данных (очистка данных, выявление аномалий), то целесообразно выбрать оценки MAE , $MdAE$, $RMSE$. В случае если, сравниваются результаты прогнозирования на данных с различными шкалами, оценки из группы абсолютных ошибок неприменимы. Ниже представлены выводы именно для такой ситуации.

2. Процентные ошибки достаточно распространены для оценки качества при решении практических задач, однако в связи с несимметричностью оценки их использование не рекомендуется. Если область изменения значений лежит в положительной полуплоскости, и нет выбросов в данных, то целесообразно применять симметричные оценки.

3. Если данные являются «грязными», т.е. содержат выбросы, то целесообразно применять масштабированные оценки. При этом горизонт прогнозирования должен быть настолько большим и не иметь одинаковых значений, чтобы сумма в знаменателе не была равна нулю.

4. Если прогнозируются данные с сезонными или циклическими паттернами, то целесообразно использовать нормализованные оценки ошибок прогнозирования, при этом фактор нормализации может быть рассчитан на интервале, равном величине цикла или сезона.

5. Если нет предварительного анализа и априорной информации о качестве данных, то требуется включение в расчет оценки качества прогнозирования различных оценок ошибок прогнозирования. После расчета анализируют полученные результаты: возникли ли ситуации деления на ноль, возникают ли противоречия:

- либо на одном и том же наборе данных результаты прогнозирования модели m_1 по одной оценке лучше модели m_2 , а по другой оценке хуже;

- либо на разных наборах результаты прогнозирования модели m_1 лучше только в большинстве случаев, а в некоторых случаях хуже.

4. Заключение

В обзоре представлены оценки ошибок прогнозирования временных рядов. Оценки сгруппированы в семь групп: абсолютные ошибки прогнозирования, процентные (относительные) ошибки прогнозирования, симметричные ошибки прогнозирования, ошибки, основанные на оценке прогноза базовых моделей, масштабированные ошибки прогнозирования, относительные оценки и другие (модифицированные) оценки. Указаны формулы расчета оценок, отражены их недостатки.

Сформулированы рекомендации по выбору тех или иных оценок ошибок прогнозирования в зависимости от задачи и имеющихся данных.

Литература

1. ЗАЙДЕЛЬ А.Н. *Элементарные оценки ошибок измерений*. – 2-е изд. испр. и доп. – Л.: «Наука», Ленинградское отд., 1967. – 88с.
2. МЕЛАМЕД М.А. *Современные методы анализа и прогнозирования режимов электропотребления в электроэнергетических системах* // Итоги науки и техники. – Сер. Энергетические системы и их автоматизация. – 1988. – Т.4. – С. 4–11.
3. САРКИСЯН С. А. *Теория прогнозирования и принятия решений* / С.А. Саркисян, В.И. Каспин, В.А. Лисичкин, Э.С. Минаев, Г.С. Пасечник. – М.: Высшая школа, 1977. – 351 с.
4. ЩЕРБАКОВ М. В., ЩЕРБАКОВА Н. Л., КОЗЛОВ И. П., КАМАЕВ В. А. *Коннективистские модели идентификации динамики систем на коротких интервалах наблюдения с заданным множеством классов поведения* // Известия ЮФУ. Технические науки. – 2012. – №1. – С. 83-91.
5. ARMSTRONG J. S., COLLOPY F. *Error measures for generalizing about forecasting methods: Empirical comparisons* // International Journal of Forecasting. – 1992. – № 8(1). – P. 69-80.

6. ARMSTRONG J. S. *Evaluating forecasting methods* // Chapter 14 in *Principles of forecasting: a handbook for researchers and practitioners*. – Norwell, MA: Kluwer Academic Publishers, 2001.
7. BREBELS A. *Mathematical and statistical framework for comparison of neural network models with other algorithms for prediction of Energy consumption in shopping centres* / Shcherbakov M.V., Kamaev V.A., et al / *Proceedings of the 37 Int. Conf. Information Technology in Science Education Telecommunication and Business*, suppl. to *Journal Open Education*. Yalta – Gurzuf. 2010. – P. 96-97.
8. CASELLA G, BERGER R. *Statistical inference*. 2nd ed. Duxbury Press., Pacific Grove, CA. 1990.
9. CHEN Z., YANG Y. *Assessing Forecast Accuracy Measures*. – 2004. [Электронный ресурс]. – Режим доступа: <http://www.stat.iastate.edu/preprint/articles/2004-10.pdf>
10. GOODWIN P., LAWTON R. *On the asymmetry of the symmetric MAPE* // *International Journal of Forecasting*. – 1999. – № 15(4). – P. 405-408.
11. GOOIJER J. G. D., HYNDMAN R. J. *25 years of time series forecasting* // *International Journal of Forecasting*. – 2006. – № 22(3). – P. 443-473.
12. HYNDMAN R. J., KOEHLER A. B. *Another look at measures of forecast accuracy* // *International Journal of Forecasting*. – 2006. – № 22(4). – P. 679-688.
13. KOEHLER A. B. *The asymmetry of the sAPE measure and other comments on the M3-competition* // *International Journal of Forecasting*. – 2001. – № 17. – P. 570– 574.
14. KOLASSA S., MARTIN R. *Percentage errors can ruin your day (and rolling the dice shows how)* // *Foresight*. – 2011. – Fall. – P. 21-27.
15. KUSIAK A., LI, M., ZHANG, Z. *A data-driven approach for steam load prediction in buildings* // *Applied Energy*. – 2010. – № 87(3). – P. 925-933.

16. MAHMOUD E. *Accuracy in forecasting: A survey* // Journal of Forecasting. – 1984. – № 3(2). – P. 139-159.
17. MAKRIDAKIS S. *Accuracy measures: Theoretical and practical concerns* // International Journal of Forecasting. – 1993. – № 9. – P. 527– 529.
18. SHCHERBAKOV M.V., BREBELS A. *Outliers and anomalies detection based on neural networks forecast procedure*: Proceedings of the 31st Annual International Symposium on Forecasting (ISF 2011). / Prague: International Institute of Forecasters, 2011. – pp. 21-22. [Электронный ресурс]. – Режим доступа:
http://www.forecasters.org/isf/pdfs/ISF11_Proceedings.pdf
19. SYNTETOS A. A., BOYLAN J. E. *The accuracy of intermittent demand estimates* // International Journal of Forecasting. – 2005. – №21(2). – P. 303-314.
20. THEODOSIOU M. *Forecasting monthly and quarterly time series using STL decomposition* // International Journal of Forecasting. – 2011. – №27(4). – P. 1178-1195.
21. YOKUMA J. T., ARMSTRONG J. S. *Beyond accuracy: Comparison of criteria used to select forecasting methods* // International Journal of Forecasting. – 1995. – № 11(4). – P. 591-597.

A SURVEY OF FORECAST ERROR MEASUREMENTS

Maxim Shcherbakov, Volgograd State Technical University, Volgograd, CAD Dept., Cand.Sc. (Engineering), Docent (maxim@shcherbakov.gmail.com).

Adriaan Brebels, Katholieke Hogeschool Kempen, Geel, Belgium, Docent (adriaan.brebels@portacapena.com)

Natalia Shcherbakova, Volgograd State Technical University, Volgograd, Physics Dept., Cand.Sc. (Engineering), Docent (snl@gebeus.ru).

Anton Tyukov, Volgograd State Technical University, Volgograd, CAD Dept., PhD student (anton.tyukov@gmail.com).

Abstract: This article provides an overview of the common used forecast error measurements. All error measurements have been joined in the seven groups. The formulas are presented and drawbacks are discussed for every accuracy measurements. Also the recommendations for selecting the appropriate error measurements are given.

Keywords: forecasting, forecast accuracy, forecast error measurements.