УДК 519.688

СТОХАСТИЧЕСКИЙ КРИТЕРИЙ ОЦЕНКИ ПРИЕМЛЕМОЙ ТОЧНОСТИ ВЕРОЯТНОСТНОГО БИНАРНОГО ПРОГНОЗИРОВАНИЯ ДИНАМИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Ю. М. Краковский*, А. Н. Лузгин**

**Иркутский государственный университет путей сообщения*
***Иркутский государственный университет*

Поступила в редакцию 26.05.2017 г.

Аннотация. Проведено исследование оценки вероятности того, что сделанный прогноз на основе вероятностных бинарных моделей с использованием методов машинного обучения оправдается. Формализован стохастический критерий приемлемой точности вероятностного бинарного прогнозирования для различных объёмов выборок значений динамических показателей. Данный критерий учитывает стохастическую природу точечных оценок вероятности того, что сделанный прогноз на основе вероятностных бинарных моделей оправдается. Предложена аппроксимация стохастического критерия приемлемой точности вероятностного бинарного прогнозирования. При этом отсутствует необходимость дополнительных статистических расчетов и обоснований. Приведены результаты экспериментов и даны соответствующие рекомендации.
Ключевые слова: вероятностное бинарное прогнозирование, интервальное прогнозирование, критерий точности прогнозирования, динамические показатели, машинное обучение.
Annotation. The authors have carried out the study of statistical patterns of a point estimation of the probability, that the made forecast, based on probabilistic binary models using methods of machine learning, will be true. The paper has formalized the stochastic criterion of the acceptable accuracy of probabilistic binary forecasting for various volumes of samples of dynamic indicators values. This criterion takes into account stochastic nature of point probability estimates, that the forecast based on probabilistic binary models will be true. The authors have proposed an approximate stochastic criterion of the acceptable accuracy of probabilistic binary forecasting. Herewith there is no need for additional statistical calculations and justifications. Authors have shown the results of experiments and given the corresponding recommendations.
Keywords: probabilistic binary forecasting, interval forecasting, accuracy forecasting estimating, a criterion of forecasting accuracy, dynamic indicators, machine learning.

ВВЕДЕНИЕ

В настоящее время большое внимание исследователей и разработчиков уделяется методам машинного обучения – обширному подразделу науки об искусственном интеллекте, изучающем интеллектуальные алгоритмы и компьютерные программы способные обучаться по имеющимся данным [1]. Одной из немаловажных задач, решаемых методами машинного обучения, является задача прогнозирования динамических показателей (ДП), с целью повышения эффективности принятия управленческих решений в различных организациях и предприятиях, осуществляющих свою деятельность в условиях неопределённости. Следует также отметить, что в последние годы наблюдается возрастающий интерес исследователей именно к вероятностным методам прогнозирования [2]. Прежде всего, это объясняется тем, что вероятностные прогнозы позволяют получить количественную оценку неопределённости самого прогноза (этой оценкой является оценка вероятности будущего события). Здесь среди методов машинного обучения можно выделить: нейронные и кластерные методы, вероятностные регрессионные методы, Байесовские методы, методы опорных векторов и случайных лесов [2–5].

Частным и достаточно распространённым случаем вероятностного прогнозирования является вероятностное прогнозирование бинарных исходов, когда в будущий момент времени может произойти только два возможных события. Необходимость бинарного прогнозирования часто появляется во многих социально-экономических и технических областях и может быть связанная, например, с такими событиями, как невозврат кредита в банке, появление экономического спада (рецессии) в некоторой отрасли экономики, введение новых законодательных норм, превышение радиационного фона от Солнца и т. п. [2]. Разновидностью бинарного прогнозирования является интервальное прогнозирование (ИП), описанное авторами в работах [6–8]. Суть данного подхода заключается в определении интервала из двух заранее заданных интервалов, в котором будет находиться будущее значение ДП на основе оценок вероятностей этих событий. При этом разделительная граница интервалов задается расчетным способом, исходя из статистических характеристик ДП.

Для любого вероятностного бинарного метода прогнозирования (ВБП) и выбранного ДП существует теоретическая и заранее неизвестная на практике вероятность (p) того, что сделанный прогноз оправдается и вероятность (q) того, что сделанный прогноз не оправдается. При этом выполняется равенство:

$$p+q=1. \tag{1}$$

Фактически вероятность p характеризует точность ВБП и, чем больше эта величина, тем лучше.

Формализуем в общем виде понятие приемлемой точности ВБП.

Определение 1. *Приемлемой точностью ВБП будем считать такое значение* p, *при котором выполняется условие:*

$$p>q. \tag{2}$$

С учетом равенства (1) условие (2) можно переписать так:

$$p>0.5. \tag{3}$$

На практике истинное значение вероятности p неизвестно и ее заменяют точечной оценкой, полученной с использованием предыстории значений ДП. Как правило, для этого большую часть имеющейся выборки значений ДП используют для обучения, а меньшую часть (10–20 % от общего объёма) для проверки результата ВБП и получения точечной оценки вероятности p. Точечная оценка неизвестной вероятности p равна:

$$\tilde{p}(l,n)=l/n. \tag{4}$$

Здесь l – число оправдавшихся прогнозов; n – общее число сделанных прогнозов; $0\le\tilde{p}(l,n)\le1$.

В работах [6–8] предложен эмпирический критерий к определению приемлемой точности ИП для любых n:

$$\tilde{p}(l,n)\ge0.6. \tag{5}$$

То есть минимальной приемлемой точностью ИП здесь является значение 0.6 (эта величина задана эмпирически). Этот критерий является подходящим для определения точности любого метода ВБП.

Однако, учитывая, что величина l в выражении (4) случайна и может рассматриваться как величина с биномиальным законом распределения (то есть как число успехов в серии из n испытаний), то возможно помимо точечной оценки (4) найти доверительный интервал, который с заданной надежностью покрывает неизвестное значение p (интервальная оценка). В связи с этим возникает задача формализации критерия для определения приемлемой точности ВБП для различных значений n с учетом интервальных оценок неизвестной вероятности p.

ИССЛЕДОВАНИЕ СТАТИСТИЧЕСКИХ ЗАКОНОМЕРНОСТЕЙ ТОЧЕЧНОЙ ОЦЕНКИ ВЕРОЯТНОСТИ ТОГО, ЧТО СДЕЛАННЫЙ ПРОГНОЗ ОПРАВДАЕТСЯ

В общем случае, когда $0<l<n$ расчет верхних и нижних границ доверительного интервала для неизвестного значения вероятности p осуществляется по формулам [9]:

$$p^{up}(l,n)=\frac{(l+1)\cdot F(\beta,v_1,v_2)}{n-l+(l+1)\cdot F(\beta,v_1,v_2)},$$
$$v_1=2\cdot(l+1),\quad v_2=2\cdot(n-l);$$
$$p^{low}(l,n)=\frac{l}{l+(n-l+1)\cdot F(\beta,v_1,v_2)},$$
$$v_1=2\cdot(n-l+1),\quad v_2=2\cdot l. \tag{6}$$

где $F(\beta,v_1,v_2)$ – критическое значение распределения Фишера при доверительной вероятности β и степенях свободы v_1, v_2; $p^{up}(l,n)$ – верхняя граница доверительного интервала для значения p; $p^{low}(l,n)$ – нижняя граница доверительного интервала для значения p.

Если $l=0$, то расчет верхних и нижних границ доверительного интервала для значения p осуществляется так [9]:

$$p^{up}(l,n)=1-(1-\beta)^{1/n},\quad p^{low}(l,n)=0. \tag{7}$$

Если $l=n$, то расчет верхних и нижних границ доверительного интервала для значения p осуществляется так [9]:

$$p^{up}(l,n)=1,\quad p^{low}(l,n)=(1-\beta)^{1/n}. \tag{8}$$

Зададим доверительную вероятность $\beta=0.95$.

На рис. 1 представлен график точечных $\tilde{p}(l,n)$ (4) и интервальных оценок $p^{up}(l,n)$, $p^{low}(l,n)$ (6) вероятности p в зависимости от значения $l=0,\ldots,n$ при $n=50$.

На рис. 2 представлен график точечных $\tilde{p}(l,n)$ (4) и интервальных оценок $p^{up}(l,n)$, $p^{low}(l,n)$ (6) вероятности p в зависимости от значения $l=0,\ldots,n$ при $n=100$.

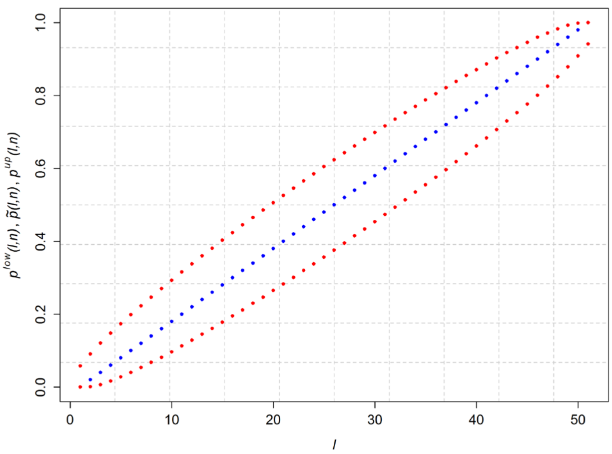

Рис. 1. График точечных и интервальных оценок вероятности p для различных l и $n=50$

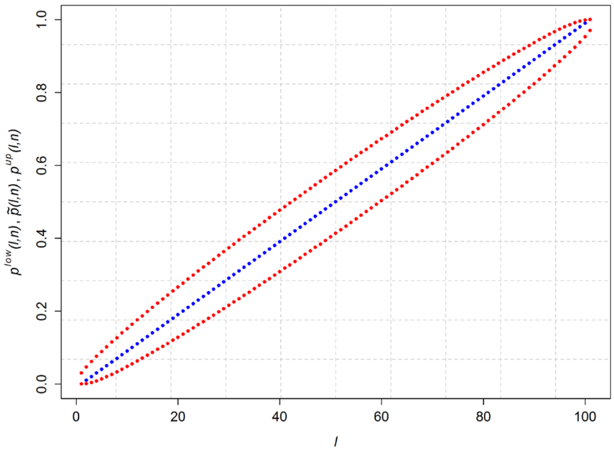

Рис. 2. График точечных и интервальных оценок вероятности p для различных l и $n=100$

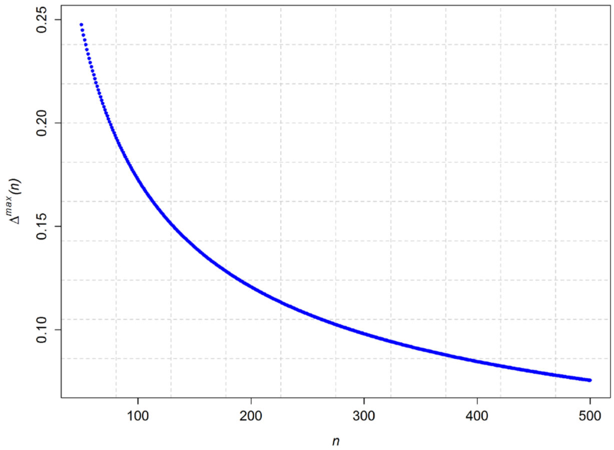

Рис. 3. График значений максимальной ширины доверительного интервала для неизвестной вероятности p при различных n

По представленным графикам видно, что с увеличением значения l доверительный интервал сначала увеличивается, а затем уменьшается.

На рис. 3 представлен график значений максимальной ширины доверительного интервала $\Delta^{\max}(n)=\max\{p^{up}(l,n)-p^{low}(l,n);\ l=0,\ldots,n\}$ для неизвестной вероятности p при $n=50,\ldots,500$.

Например, для при $n=50$ значение $\Delta^{\max}(n)=0.25$, а для $n=500$ значение $\Delta^{\max}(n)=0.075$. То есть с увеличением значения n максимальная ширина доверительного интервала $\Delta^{\max}(n)$ уменьшается.

Таким образом, точечная оценка $\tilde{p}(l,n)$ (4) неизвестной вероятности p корректно характеризует основные статистические тенденции при изменении значения l или n и является корректной характеристикой точности ВБП. При этом желательно выбирать значение n как можно больше, чтобы по максимуму сузить доверительные границы для неизвестной вероятности p, что обеспечит более точную точечную оценку этой величины.

В тоже время возникает вопрос корректности применения эмпирического критерия приемлемой точности ВБП (5) для небольших значений n. В этом случае нижняя доверительная граница неизвестной вероятности p может быть ниже значения 0.5, вследствие значительной ширины доверительного интервала. В статистическом смысле это означает, что точность такого ВБП неприемлема, так как число верных прогнозов может быть меньше половины из всех. При этом критерий (5) оказывается выполнен. Значения n, при которых применение критерия (5) оправдано, будут найдены далее.

СТОХАСТИЧЕСКОЕ ОПРЕДЕЛЕНИЕ И ФОРМАЛИЗАЦИЯ ПРИЕМЛЕМОЙ ТОЧНОСТИ ВЕРОЯТНОСТНОГО БИНАРНОГО ПРОГНОЗИРОВАНИЯ

Используя равенство (1) оценку вероятности q того, что сделанный прогноз не оправдается можно осуществить по формуле:

$$\tilde{q}(l,n) = 1 - \tilde{p}(l,n) = (n-l)/n. \tag{9}$$

Здесь $\tilde{q}(l,n)$ – точечная оценка неизвестной вероятности q того, что сделанный прогноз не оправдается; $n-l$ – число ошибочных прогнозов; n – общее число сделанных прогнозов; $0 \le \tilde{q}(l,n) \le 1$.

Величина $n-l$ (как и величина l в выражении (4)) случайна и может рассматриваться как величина с биномиальным законом распределения (то есть как число неудач в серии из n испытаний).

В общем случае, когда $0 < n-l < n$ расчет верхних и нижних границ доверительного интервала для неизвестного значения вероятности q осуществляется по формулам [9]:

$$\begin{gathered}
q^{up}(l,n) = \frac{(n-l+1)\cdot F(\beta,v_1,v_2)}{l+(n-l+1)\cdot F(\beta,v_1,v_2)}, \\
v_1 = 2\cdot(n-l+1), \quad v_2 = 2\cdot l; \\
q^{low}(l,n) = \frac{n-l}{n-l+(l+1)\cdot F(\beta,v_1,v_2)}, \\
v_1 = 2\cdot(l+1), \quad v_2 = 2\cdot(n-l).
\end{gathered} \tag{10}$$

где $F(\beta,v_1,v_2)$ – критическое значение распределения Фишера при доверительной вероятности β и степенях свободы v_1, v_2; $q^{up}(l,n)$ – верхняя граница доверительного интервала для значения q; $q^{low}(l,n)$ – верхняя граница доверительного интервала для значения q.

Если $n-l=0$, то расчет верхних и нижних границ доверительного интервала для неизвестной вероятности q осуществляется так [9]:

$$q^{up}(l,n) = 1-(1-\beta)^{1/n}, \quad q^{low}(l,n) = 0. \tag{11}$$

Если $n-l=n$, то расчет верхних и нижних границ доверительного интервала для неизвестного значения вероятности q осуществляется так [9]:

$$q^{up}(l,n) = 1, \quad q^{low}(l,n) = (1-\beta)^{1/n}. \tag{12}$$

Определение 2. *Приемлемой точностью ВБП будем считать такое значение $\tilde{p}(l,n)$ (4), при котором это значение статистически «различимо» со значением $\tilde{q}(l,n)$ (9).*

Для этого введем следующий критерий приемлемой точности ВБП:

$$p^{low}(l,n) > q^{up}(l,n), \tag{13}$$

где $p^{low}(l,n)$ – нижняя доверительная граница для неизвестной вероятности p; $q^{up}(l,n)$ – верхняя доверительная граница для неизвестной вероятности q.

Данный критерий был назван стохастическим критерием, так учитывает стохастическую (случайную) природу точечных оценок неизвестной вероятности p.

То есть, согласно критерию (13) доверительные границы для неизвестных вероятностей p и q не должны «перекрываться». Если критерий (13) не выполняется, то точечные оценки $\tilde{p}(l,n)$ и $\tilde{q}(l,n)$ статистически «неразличимы», следовательно точность ВБП неприемлема (в статистическом смысле только половина прогнозов истины).

Следует отметить, что критические значения распределения Фишера, используемые для вычисления величины $p^{low}(l,n)$ в (6) и величины $\tilde{q}(l,n)$ в (10), эквивалентны, так как рассчитываются при одинаковых степенях свободы: $v_1 = 2\cdot(n-l+1)$ и $v_2 = 2\cdot l$.

Докажем, что для любых $l = 0,\dots,n$ и n выполняется равенство:

$$p^{low}(l,n)+q^{up}(l,n)=1. \tag{14}$$

Доказательство. В общем случае, когда $0<l<n$, используя выражения (6) и (10), получим:

$$p^{low}(l,n)+q^{up}(l,n)= \\ =\frac{l}{l+(n-l+1)\cdot F(\beta,v_1,v_2)}+ \\ +\frac{(n-l+1)\cdot F(\beta,v_1,v_2)}{l+(n-l+1)\cdot F(\beta,v_1,v_2)}=1. \tag{15}$$

В частном случае, когда $l=0$, то $p^{low}(l,n)=0$ (7), а $q^{up}(l,n)=1$ (12), следовательно $p^{low}(l,n)+q^{up}(l,n)=1$ и наоборот, когда $l=n$, то $p^{low}(l,n)=(1-\beta)^{\frac{1}{l}}$ (8) и $q^{up}(l,n)=1-(1-\beta)^{\frac{1}{l}}$ (11) следовательно $p^{low}(l,n)+q^{up}(l,n)=1$. Таким образом, равенство (14) доказано.

Используя равенство (14) упростим стохастический критерий (13):

$$p^{low}(l,n)>q^{up}(l,n)\to \\ \to p^{low}(l,n)-q^{up}(l,n)>0\to \\ \to p^{low}(l,n)-(1-p^{low}(l,n))\to \\ \to pl^{low}(l,n)>0.5. \tag{16}$$

Таким образом, точность ВБП приемлема, если выполняется критерий (16).

На рис. 4 для различных $n=50,\dots,500$ показан график минимальных приемлемых значений точечной оценки неизвестной вероятности p ($\tilde{p}^{\min}(n)$), для которых нижняя доверительная граница удовлетворяет (16).

По графику видно, что при возрастающем значении n значения $\tilde{p}^{\min}(n)$ скачкообразно уменьшаются. При $n=50$ значение $\tilde{p}^{\min}(n)=0.64$, а при $n=500$ это значение уже равно $\tilde{p}^{\min}(n)=0.54$.

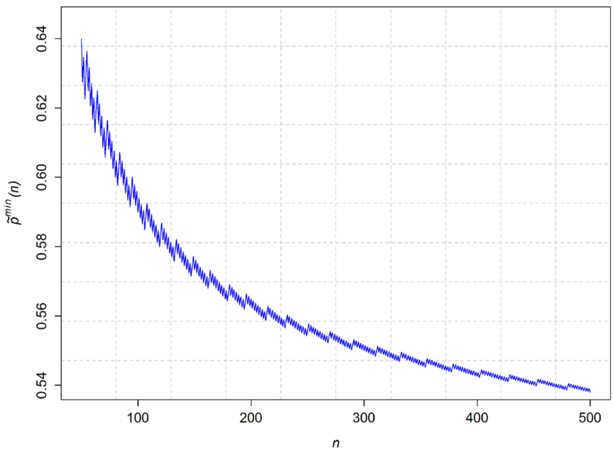

Рис. 4. График значений $\tilde{p}^{\min}(n)$, для которых нижняя доверительная граница удовлетворяет критерию (16)

Полученные данные позволяют определить значение n, при котором применение эмпирического критерия приемлемой точности ВБП (5) допустимо. Данный критерий применим только при $n\geq 96$. Именно в этом случае минимальная приемлемая точность ВБП $\tilde{p}^{\min}(n)$ всегда больше, либо равна 0.6, что удовлетворяет требованию критерия (5) для которого минимальная приемлемая точность должна быть равна 0.6.

При меньших значениях n эмпирический критерий (5) может быть выполнен, однако точность ВБП в статистическом смысле может оказаться неприемлемой. Например, при оценке точности ВБП при $n=50$ и полученном значении $\tilde{p}(l,n)=0.62$ критерий (5) выполнен. Однако здесь нижняя доверительная граница неизвестной вероятности p ниже, либо равна значению 0.5, следовательно, как было показано выше, такую точность ВБП считать приемлемой некорректно.

Для сокращения числа вычислительных операций предлагается рассчитывать минимальные приемлемые значения $\tilde{p}^{\min}(n)$ точечных оценок неизвестной вероятности p для различных n, используя следующую аппроксимацию:

$$\tilde{p}^{\min}(n)\sim\tilde{z}^{\min}(n)=1.47\cdot n^{-3/5}+0.50, \\ 50\leq n\leq 500. \tag{17}$$

Данная аппроксимация была получена путем минимизации по параметрам a, b и c следующей функции с ограничением:

$$\arg\min f(a,b,c) = \sum_{n=50}^{500}\left(a\cdot n^c + b - \tilde{p}^{\min}(n)\right), \quad a\cdot n^c + b - \tilde{p}^{\min}(n) \ge 0. \tag{18}$$

Ограничение в (18) необходимо для того, чтобы обеспечить выполнение условия:

$$\tilde{p}^{\min}(n) \le a\cdot n^c + b. \tag{19}$$

Для минимизации функции (18) возможно использовать различные методы минимизации функций с ограничениями [10].

Аппроксимация (17) соответствует условию $\tilde{p}^{\min}(n) \le 1.47\cdot n^{-3/5} + 0.50$ и, следовательно, удовлетворяет критерию (13). При этом максимальная абсолютная ошибка аппроксимации не превышает 1.8 % от истинного значения. На рис. 5 показан график истинных значений $\tilde{p}^{\min}(n)$ и аппроксимированных $\tilde{z}^{\min}(n)$.

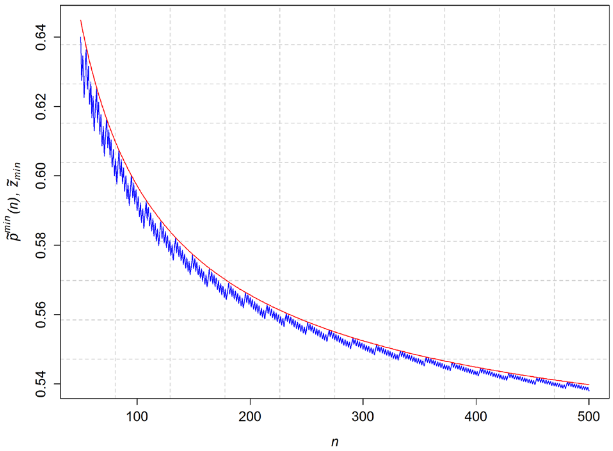

Рис. 5. График истинных значений $\tilde{p}^{\min}(n)$ и аппроксимированных $\tilde{z}^{\min}(n)$

Используя выражение (17) дадим новое альтернативное определение приемлемой точности ВБП.

Определение 3. *Приемлемой точностью ВБП будем считать такое значение точечной оценки $\tilde{p}(l,n)$ (4) неизвестной вероятности p, для которой выполняется стохастический критерий:*

$$\tilde{p}(l,n) \ge 1.47\cdot n^{-3/5} + 0.50, \quad 50 \le n \le 500. \tag{20}$$

Критерий (20) в отличие от критерия (16) не требует расчетов нижней доверительной границы неизвестной вероятности p, что сокращает число вычислительных операций и упрощает проверку того, приемлема или нет точность ВБП, которую характеризует точечная оценка $\tilde{p}(l,n)$ (4) для $n = 50,\ldots,500$.

Следует отметить, что отличие критерия (20) от критерия (5), состоит прежде всего в том, что в критерии (5) минимальная допустимая приемлемая точность ВБП задана эмпирически и неизменно равной 0.6 (что является корректным не для всех значений n, как было показано выше), а в критерии (20) минимальная допустимая приемлемая точность ВБП рассчитывается «динамически» в зависимости от заданного значения n.

Важной преимуществом критерия (20) является то, что в данном критерии уже учтена стохастическая природа точечной оценки неизвестной вероятности p, без необходимости выполнения дополнительных статистических расчетов и обоснований.

Следует также отметить, что в работах [6–8] критерий (5) был выбран корректно, поскольку применялся при $n \ge 100$.

ЗАКЛЮЧЕНИЕ

Проведено исследование оценки вероятности того, что сделанный прогноз на основе вероятностных бинарных моделей с использованием методов машинного обучения оправдается. Формализован стохастический критерий приемлемой точности вероятностного бинарного прогнозирования для различных значений n (16). С использованием аппроксимирующей функции предложен стохастический критерий (20) приемлемой точности вероятностного бинарного прогнозирования для значений $n = 25,\ldots,500$ и доверительной вероятности $\beta = 0.95$, учитывающий стохастическую природу точечных оценок вероятности того, что сделанный прогноз на основе вероятностных бинарных моделей с использованием методов машинного обучения оправдается, без необходимости дополнительных статистических расчетов и обоснований. Показано, что при $n < 96$ использовать эмпирический критерий (5) приемлемой точности вероятностного бинарного прогнозирования некорректно. Учитывая полученные результаты, на практике рекомендуется осуществлять проверку точности вероятностного бинарного прогнозирования, используя оценку (4), а также проверку того, приемлема или нет точность такого прогнозирования, на основе предложенных в данной работе критериев (16) или (20). Все сделанные в работе выводы, предложенные оценки и критерии, полностью справедливы и для интервального прогнозирования.

СПИСОК ЛИТЕРАТУРЫ

1. *James G., Witten D., Hastie T., Tibshirani T.* An Introduction to Statistical Learning with Applications in R / G. James, D. Witten, T. Hastie, T. Tibshirani. – New York: Springer-Verlag, 2014. – 426 p.

2. *Elliott G., Granger C., Timmermann A.* Handbook of Economic Forecasting / G. Elliott, C. Granger, A. Timmermann. – Amsterdam: Elsevier, 2013. – Vol 2. – 1324 p.

3. *Crawford M. M., Ham J., Chen Y., Ghosh J.* Random Forests of Binary Hierarchical Classifiers for Analysis of Hyperspectral Data / M. M. Crawford, J. Ham, Y. Chen, J. Ghosh // In 2003 IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003. – pp. 337–345.

4. *Ng A. Y., Jordan M. I.* On Discriminative vs. Generative classifiers: A Comparison of Logistic Regression and Naive Bayes / A. Y. Ng, M. I. Jordan // Advances In Neural Information Processing Systems, 2002. – Accessed at: https://papers.nips.cc/paper/2020-on-discriminative-vs-generative-classifiers-a-comparison-of-logistic-regression-and-naive-bayes.pdf

5. *Platt C. J.* Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods / C. J. Platt, 1999. – Accessed at: citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.41.1639&rep=rep1&type=pd

6. *Краковский Ю. М., Лузгин А. Н.* Алгоритм интервального прогнозирования динамических показателей на основе робастной вероятностной кластерной модели / Ю. М. Краковский, А. Н. Лузгин // Наука и образование. ФГБОУ ВПО «МГТУ им. Н.Э. Баумана». Электрон. журн. – 2016. – №11. – С. 113 – 126. Режим доступа: http://technomag.neicon.ru/doc/849839.html

7. *Краковский Ю. М., Лузгин А. Н.* Алгоритм интервального прогнозирования динамических показателей на основе вероятностной нейросетевой модели / Ю. М. Краковский, А. Н. Лузгин // Современные технологии. Системный анализ. Моделирование. – 2016. – № 4 (50). – С. 126–132.

8. *Краковский Ю. М., Лузгин А. Н.* Проверка точности интервального прогнозирования на основе доверительных оценок вероятностей / Ю. М. Краковский, А. Н. Лузгин // Прикладная информатика. – 2017. – Т. 12, № 1 (67). – С. 114–120.

9. *Закс Л.* Статистическое оценивание / Л. Закс. – М. : Статистика, 1976. – 598 с.

10. *Kelley C. T.* Iterative Methods for Optimization / C. T. Kelley, 1999. – 180 p.

Краковский Юрий Мечеславович – д-р техн. наук, профессор, профессор кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения.
E-mail: kum@stranzit.ru

Krakovsky Y. M. – Dr. Tech. Sci., Professor, Professor of Department of Information Systems and Information Security, Irkutsk State University of Railway Transport.
E-mail: kum@stranzit.ru

Лузгин Александр Николаевич – канд. техн. наук, преподаватель кафедры «Информационные технологии», Иркутский государственный университет.
E-mail: alexln@mail.ru

Luzgin A. N. – Cand. Tech. Sci. (Eng.), Teacher of the Department of Information Technologies, Irkutsk State University.
E-mail: alexln@mail.ru