Десятая Международная научно-практическая конференция «BIG DATA and Advanced Analytics. BIG DATA и анализ высокого уровня», Минск, Республика Беларусь, 13 марта 2024 год

УДК 519.624.2

О МАКСИМАЛЬНОЙ ПЛОЩАДИ ПОД ИНТЕГРАЛЬНОЙ КРИВОЙ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ ОШИБКИ ПРИ СИНТЕЗЕ ВП

.

И.П. Кобяк
Доцент кафедры ЭВМ БГУИР,
кандидат технических наук, доцент
IPKobyak2012@mail.ru

И.П.Кобяк

Работает в Белорусском государственном университете с 1982 г. Занимаемые должности инженер, ассистент, доцент кафедры ЭВМ. Защитил кандидатскую диссертацию в 1993 г. Область научных интересов: методы идентификации сообщений, проектирование спецпроцессоров.

Аннотация. В работе на основе алгоритма конечных разностей представлена методика формирования равенства для суммы вероятностных моментов производящей функции, соответствующей методу идентификации шумоподобных сообщений оценками наблюдения заданных пар векторов состояний. Полученные зависимости предполагается использовать в технических системах, ориентированных на прием и передачу зашумленных сообщений. Получено аналитическое соотношение для перечисляющей производящей функции, позволяющее определить площадь под кривой вероятности пропуска ошибки при наблюдении векторов переходов, регистрируемых с минимальной вероятностью. В статье также рассмотрен вопрос сравнения в асимптотике площади под интегральной кривой уровней вероятности ошибки, порождаемых предлагаемым методом с известными алгоритмами синтеза контрольных кодов.

Ключевые слова: вектор перехода, субдинамические объекты, сложные события, монообъекты и постобъекты, идентификация последовательностей, вероятность пропуска ошибки, сигнатурный анализ, вектор состояния.

Введение. Применение новых методов идентификации для задач передачи и приема информации по каналам шумоподобной связи является одним из важнейших аспектов повышения эффективности использования проблемно-ориентированных компьютерных систем. Как правило, для идентификации зашумленных сообщений используются коды-идентификаторы на основе выборочных функций [1] или линейной свертки [2], характеризующих свойства последовательностей конечной длины. Реализация данных методов предполагает исследование и анализ алгоритмов сверточного кодирования с учетом ряда научно-технических направлений. В частности, при конкретной реализации сообщений, циркулирующих в сетях передачи данных, необходимо знать уровень коллизий или другие свойства используемой оценки или линейной сигнатуры. Это позволит сократить время на выявление и исправление возможных ошибок в случае появления непредвиденных искажений в данных. С общетеоретической точки зрения необходим системный анализ параметров и их уровней для вероятностей ошибки на основе производящих функций (ПФ). Результатом таких исследований с учетом принципов математической формализации зависимостей является возможность выбора наиболее эффективного из алгоритмов контрольного кодирования путем сопоставления свойств оценок по ряду выбранных показателей [3]. С учетом сказанного в представляемой работе рассматривается метод формирования суммы вероятностных моментов в производящей функции для минимальной вероятности наблюдения векторов переходов заданного вида (ВПЗВ) или лебеговской меры векторов переходов (ВП). Основой для выполненных исследований послужили результаты исследования двоичных последовательностей, которые указали на неравенство аргументов моды и математического ожидания в функции распределения вероятностей ошибки. Таким образом, из указанных результатов вытекает необходимость всестороннего рассмотрения предлагаемого метода оценивания многомерных сообщений в сетях передачи данных для формирования количественных показателей и анализа его преимуществ в соответствии с работами [5,6].

Математические объекты в задаче исследования. С учетом работы [5] при наблюдении ВПЗВ можем записать соотношение для энумератора в форме кусочно-гладкой функции общего вида:

$$\begin{aligned} P_{ifc} = Mo + \sum_{g}^{n-g} \pi(g)\ & \xi_1 p e^t \ k_{1,1} x_1^1 + \sum_{i=2}^{n-4} k_{1,i} x_1^i \frac{1}{2^{i+1}} \beta_{1,i} + k_{1,n-2} x_1^{n-2} \frac{1}{2^{n-1}} \beta_{1,n-2} \ + \\ & + \xi_2 p^2 e^{2t} \ k_{2,1} x_2^1 + \sum_{i=2}^{n-6} k_{2,i} x_2^i \frac{1}{2^{i+1}} \beta_{2,i} + k_{2,n-4} x_2^{n-4} \frac{1}{2^{n-3}} \beta_{2,n-4} \ + \\ & + \ldots + \xi_{0,5(n-5)} p^{\frac{n-5}{2}} e^{\frac{n-5}{2}t} \ k_{\frac{n-5}{2},1} x_{\frac{n-5}{2}}^1 + \sum_{i=2}^{3} k_{\frac{n-5}{2},i} x_{\frac{n-5}{2}}^i \frac{1}{2^{i+1}} \beta_{\frac{n-5}{2},i} + k_{\frac{n-5}{2},5} x_{\frac{n-5}{2}}^5 \frac{1}{2^6} \beta_{\frac{n-5}{2},5} \ + \\ & + \xi_{0,5(n-3)} p^{\frac{n-3}{2}} e^{\frac{n-3}{2}t} \ k_{\frac{n-3}{2},3} x_{\frac{n-3}{2}}^3 \frac{1}{2^4} \beta_{\frac{n-3}{2},3} \ + \xi_{0,5(n-1)} p^{\frac{n-1}{2}} e^{\frac{n-1}{2}t} k_{\frac{n-1}{2},1} x_{\frac{n-1}{2}}^1 . \end{aligned} \tag{1}$$

Математическое ожидание числа последовательных пар ВПЗВ определяется соотношением:

$$j_{mo} = \frac{1}{1-p} = 1 + \xi , \tag{2}$$

где p - асимптотическая вероятность (или удельный вес ВПЗВ в паре ВС при бесконечном n), вида:

$$p = \frac{3^{r-\mu}}{m^2}, \qquad m = 2^r , \tag{3}$$

параметр μ - это число единичных символов в ВП заданного вида.

Однако, в целом, при рассмотрении вероятности (3) с учетом соотношения для математического ожидания (2) можно заключить, что данный параметр для многомерных задач, использующих полиномиальное представление вероятностей, должен быть трансформирован в соотношение:

$$f(p) = \frac{1}{\xi+1}\varphi(p) + \frac{\xi}{\xi+1}\varphi\left(p^2\right), \tag{4}$$

то есть, если $1+\xi$ – это значение математического ожидания параметра, то представление вероятностей в производящей функции должно соответствовать хотя бы минимальному полиному вида (4). Следовательно, в сокращенной форме энумератор (1) может быть представлен суммой:

$$P_{ifc} < Mo + \sum_{g} \pi(g)\ \xi_1 p e^t \left(k_{1,1}x_1^1 + \sum_{i=2}^{n-4} k_{1,i}x_1^i \frac{1}{2^{i+1}}\beta_{1,i} + k_{1,n-2}x_1^{n-2}\frac{1}{2^{n-1}}\beta_{1,n-2} \right. + \tag{5}$$

$$\left. + \xi_2 p^2 e^{2t} \left(k_{2,1}x_2^1 + \sum_{i=2}^{n-6} k_{2,i}x_2^i \frac{1}{2^{i+1}}\beta_{2,i} + k_{2,n-4}x_2^{n-4}\frac{1}{2^{n-3}}\beta_{2,n-4} \right)^{n-g} \right).$$

где множители ξ_j следуют из представления функции в соответствии с общим видом (4).

Далее, следует учитывать, что пара векторов, образующих ВПЗВ, имеет число вариантов $3^{r-\mu}$, а пара ВС, не образующих ВПЗВ - $m^2 - 3^{r-\mu}$. Тогда с учетом производной:

$$\frac{\partial}{\partial r}3^{r-\mu} < \frac{\partial}{\partial r}\left(m^2 - 3^{r-\mu}\right) \tag{6}$$

можем заключить, что значение моды распределения (5) равно:

$$Mo = \frac{1}{2^{n+1}}\beta_n,$$

то есть максимум имеет место при отсутствии событий $k_{j,i}$.

Рассмотрим теперь принцип образования верхней границы для ПФ, представленной соотношением (5).

Верхняя граница для интеграла вероятностей пропуска ошибки. Для определения верхней границы интеграла вероятностей ошибки для всех значений $r > 1$ определим минимальное значение вероятности в данных исследованиях числом $p = \frac{1}{m^2}$, что следует из тривиальных подстановок параметров r и μ в формулу (3).

Теперь рассмотрим разность двух объектов, определяющих функцию правдоподобия для $j = 1$ и произвольных значений i:

$$\frac{1}{2^{i+1}}\beta_{1,i} - \frac{1}{2^{i+2}}\beta_{1,i+1} = \frac{1}{2^{i+1}}\frac{1}{\sqrt{1-4p}}\left[\left(1+\sqrt{1-4p}\right)^{i+1} - \left(1-\sqrt{1-4p}\right)^{i+1}\right] -$$

$$-\frac{1}{2^{i+2}}\frac{1}{\sqrt{1-4p}}\left[\left(1+\sqrt{1-4p}\right)^{i+2} - \left(1-\sqrt{1-4p}\right)^{i+2}\right].$$

Данная разность достаточно просто может быть преобразована к виду:

$$\frac{1}{\sqrt{1-4p}}\left[\left(\frac{1}{2}+\frac{1}{2}\sqrt{1-4p}\right)^{i+1} - \left(\frac{1}{2}-\frac{1}{2}\sqrt{1-4p}\right)^{i+1} - \left(\frac{1}{2}+\frac{1}{2}\sqrt{1-4p}\right)^{i+2} + \left(\frac{1}{2}-\frac{1}{2}\sqrt{1-4p}\right)^{i+2}\right] =$$

$$= \frac{1}{\sqrt{1-4p}}\left[\left(A^{i+1} - B^{i+1}\right) - \left(A^{i+2} - B^{i+2}\right)\right] = \frac{m}{\sqrt{m^2-4}}\left[A^{i+1}(1-A) - B^{i+1}(1-B)\right],$$

здесь

$$1-A=\frac{m-\sqrt{m^2-4}}{2m},\qquad 1-B=\frac{m+\sqrt{m^2-4}}{2m}.$$

Таким образом, из разности моментов следует:

$$\frac{1}{2^{i+1}}\beta_{1,i}-\frac{1}{2^{i+2}}\beta_{1,i+1}=\frac{m}{\sqrt{m^2-4}}\left[\frac{m-\sqrt{m^2-4}}{2m}\left(\frac{m+\sqrt{m^2-4}}{2m}\right)^{i+1}-\frac{m+\sqrt{m^2-4}}{2m}\left(\frac{m-\sqrt{m^2-4}}{2m}\right)^{i+1}\right].$$

Используя в данном соотношении формулу для разности квадратов и тот факт, что при $r>>1$ значение $m-\sqrt{m^2-4}\approx 0$, имеем:

$$\frac{1}{2^{i+1}}\beta_{1,i}-\frac{1}{2^{i+2}}\beta_{1,i+1}\approx\frac{1}{m\sqrt{m^2-4}}\left(\frac{m+\sqrt{m^2-4}}{2m}\right)^{i}\xrightarrow{r>>1}\frac{1}{m\sqrt{m^2-4}}\left(1-\frac{\Delta}{2m}\right)^{i}=\frac{1}{m\sqrt{m^2-4}}\Theta^{i}.$$

Из данного соотношения следует, что:

$$\frac{1}{m\sqrt{m^2-4}}\left(1-\frac{\Delta}{2m}\right)^{i}=\frac{1}{m\sqrt{m^2-4}}\Theta^{i}. \tag{7}$$

Рассмотрим теперь фрагмент функции (5) для $j=1$ при $\xi_1=1$ и всех $k_{1,i}=1$. При этом сумма элементов в скобках, выраженная с использованием результата (7), может быть приведена к виду:

$$\sum_{i=2}^{n-4}\frac{1}{2^{i+1}}\beta_{1,i}=(n-5)\frac{1}{2^{2+1}}\beta_{1,2}-\frac{1}{m\sqrt{m^2-4}}\left(\Theta^2\sum_{i=2}^{n-4}1+\Theta^3\sum_{i=3}^{n-4}1+\Theta^4\sum_{i=4}^{n-4}1+\ldots+\Theta^{n-5}\sum_{i=n-5}^{n-4}1+\Theta^{n-4}\sum_{i=n-4}^{n-4}1\right). \tag{8}$$

где

$$\frac{1}{2^3}\beta_{1,2}=\frac{1}{\sqrt{1-4\frac{1}{m^2}}}\left[\left(\frac{1}{2}+\frac{1}{2}\sqrt{1-4\frac{1}{m^2}}\right)^{2+1}-\left(\frac{1}{2}-\frac{1}{2}\sqrt{1-4\frac{1}{m^2}}\right)^{2+1}\right]\approx 1-\frac{1}{m^2},$$

Полагая далее, что $\Theta^{i}\xrightarrow{r>>1}1$ получаем минимальное (возможно отрицательное, но далее корректируемое) значение функции (8):

$$\sum_{i=2}^{n-4} \frac{1}{2^{i+1}} \beta_{1,i} = (n-5) \frac{1}{2^{2+1}} \beta_{1,2} - \frac{1}{m\sqrt{m^2-4}} \frac{(n-5)(n-4)}{2}.$$

Учитывая, что каждая разность $n-3-i$ в (8) умножается на дробь, данное соотношение с большей степенью достоверности может быть представлено в виде:

$$\sum_{i=2}^{n-4} \frac{1}{2^{i+1}} \beta_{1,i} = (n-5) \frac{1}{2^{2+1}} \beta_{1,2} - \frac{1}{m\sqrt{m^2-4}} \frac{(n-5)(n-4)}{2} B_1,$$

где B_1 - абстрактный множитель, уменьшающий сумму членов арифметической прогрессии до реального значения.

Из полученного равенства при $\frac{1}{2^{2+1}} \beta_{1,2} = \frac{15}{16} \approx 1$ и $m\sqrt{m^2-4} \approx m^2$ достаточно просто получить соотношение:

$$\sum_{i=2}^{n-4} \frac{1}{2^{i+1}} \beta_{1,i} = (n-5) - \frac{1}{2m^2} (n-5)(n-4) B_1. \tag{9}$$

Учитывая, что сумма моментов в (9) должна быть больше нуля, из правой части приведенного равенства имеем квадратное неравенство:

$$-n^2 B_1 + \left(2m^2 + 9B_1\right) n - \left(10m^2 + 20B_1\right) > 0.$$

откуда следует: $n \approx \frac{2m^2}{B_1}$.

Ряд преобразований для второй части функции (5) приводят к равенству:

$$\sum_{i=2}^{n-6} \frac{1}{2^{i+1}} \beta_{2,i} = (n-7) \frac{1}{2^{2+1}} \beta_{2,2} - \frac{1}{m\sqrt{m^2-4}} \frac{(n-7)(n-6)}{2} B_2$$

и, соответственно к результату: $n \approx \frac{2m^2}{B_2}$.

Тогда из (5) следует:

$$P_{ifc} \approx \left(1-\frac{\Delta}{2m}\right)^{n+1} + \left(\frac{1}{1+\xi}\left(p+2\left(\frac{1}{B_1}-\frac{1}{B_1^2}\right)\right)+\frac{\xi}{1+\xi}\left(p^2+2p\left(\frac{1}{B_2}-\frac{1}{B_2^2}\right)\right)\right)^{n}.$$

Соответственно, для разностей в скобках можем составить два неравенства. В первом случае:

$$2\left(\frac{1}{B_1}-\frac{1}{B_1^2}\right) < p,$$

при этом некоторая бесконечно малая величина

$$\psi_1 = \frac{1}{B_1}-\frac{1}{B_1^2} < \frac{1}{2}p.$$

Во втором случае имеем

$$2p\left(\frac{1}{B_2}-\frac{1}{B_2^2}\right) < p^2,$$

$$\psi_2 = \frac{1}{B_2}-\frac{1}{B_2^2} < \frac{1}{2}p.$$

Тогда соотношение для ПФ может быть представлено в виде:

$$P_{ifc} \approx \left(1-\frac{\Delta}{2m}\right)^{n+1} + \left(\frac{1}{1+\xi}\left(p+2\left(\psi_1 < \frac{1}{2}p\right)\right)+\frac{\xi}{1+\xi}\left(p^2+2p\left(\psi_2 < \frac{1}{2}p\right)\right)\right)^{n}.$$

Учитывая теперь степень n у дроби во второй части формулы, практически без потери общности, можем записать:

$$P_{ifc} \approx \left(1-\frac{\Delta}{2m}\right)^{n+1} + \frac{1}{2^n}p^n, \qquad (10)$$

$$p^2 \ll p.$$

Полученное соотношение позволяет заключить, что при $n \rightarrow$ численное значение ПФ для вероятностей наблюдения ВПЗВ стремиться к моде распределения вероятностей.

Сравнение методов вероятности наблюдения ВПЗВ с классическими алгоритмами свертки. Классические алгоритмы свертки цифровых последовательностей, как правило, основываются на суммировании выбранных элементарных событий в идентифицируемых сообщениях. При этом площадь под огибающей кривой плотности распределения вероятностей пропуска ошибки принципиально равна единице, что следует из факта полной группы событий для вероятностей наблюдаемых объектов. Так при сигнатурном анализе (*sa*) функция распределения вероятностей ошибки, представляющая собой прямую линию, при суммировании всех аргументов приводит к единичному результату. Рассматриваемый же в данных исследованиях алгоритм представляет собой методологию особого вида, основанную на явлении переключения технических систем из состояния $A(t)$ в состояние $A(t+1)$ под действием синхронизирующих импульсов. Соответственно задача наблюдения ВПЗВ или лебеговской меры ВП не позволяет получить сумму вероятностей ошибки равную единице. Следовательно, сравнение параметров предлагаемого в данной работе метода и алгоритмов сигнатурного анализа, а также наблюдения векторов состояний, сводится к получению обратных функций для интегральных сумм моментов при минимальном или максимальном значениях рассмотренного в данной работе энумератора.

Сравнение интеграла вероятности пропуска ошибки при сигнатурном анализе (или регистрации векторов состояний) с методом наблюдения ВПЗВ при максимальном значении ПФ показывает, что имеет место примерное равенство при малых значениях n. Однако, с ростом длины выборки преимущество метода наблюдения ВПЗВ возрастает весьма существенно.

Список литературы

[1] Кобяк И.П. Идентификация случайных сообщений статистической оценкой векторов переходов // АиТ. 1997. №9. С.81-86.

[2] Nadig H.G. Signature analysis - concepts, examples and guidelines // Hewlett Packard J. - May. 1977. P. 5-21.

[3] Кобяк И.П. Теория внутрисхемного наблюдения СБИС с использованием автокорреляционных функций // Автоматика и вычислительная техника. 2009. № 2. С. 37-46.

[4] Кобяк И.П. Производящая функция для распределения вероятностей наблюдения векторов переходов // Автоматика и вычислительная техника. 2006. № 6. С. 60-67.

[5] Кобяк И.П. Производящая функция для вероятности пропуска ошибки при наблюдении векторов переходов. В кн.: BIG DATA и анализ высокого уровня 2023, BIG DATA and advanced analytics 2023: сборник материалов 9-й международной научно-практической конференции, часть 2, Минск, 17-18 мая 2023 г. / Белорусский государственный университет информатики и радиоэлектроники, Минск, 2023. С. 16-23.

[6] Кобяк И.П. Сравнительный оценка достоверности методов сигнатурного анализа и счета состояний // Электрон. моделирование. 1996. Т.18. №3. С. 58-62.

Авторский вклад

Кобяк Игорь Петрович – руководство исследованием о максимальной площади под интегральной кривой распределения вероятностей ошибки при синтезе ВП, формирование структуры статьи

ON THE MAXIMUM AREA UNDER THE INTEGRAL CURVE OF THE ERROR PROBABILITY DISTRIBUTION IN THE SYNTHESIS OF VP

I.P. Kabiak,
PhD, Associate Professor,
Chair of ECM, BSUIR.

Abstract. Based on the finite difference algorithm, the paper presents a method for generating equality for the sum of the probabilistic moments of the generating function corresponding to the method of identifying noise-like messages by observing estimates of given pairs of state vectors. The obtained dependencies are supposed to be used in technical systems focused on receiving and transmitting noisy messages. An analytical relation for the enumerating generating function is obtained, which makes it possible to determine the area under the probability curve of missing an error when observing transition vectors recorded with minimal probability. The article also considers the issue of comparing, in the asymptotic of the area under the integral curve, the error probability levels generated by the proposed method with known algorithms for the synthesis of control codes.

Keywords: Keywords: transition vector, sub dynamic objects, complex events, mono objects and post objects, sequence identification, probability of missing an error, signature analysis, state vector.