ОПТИМАЛЬНАЯ ИДЕНТИФИКАЦИЯ СЛУЧАЙНЫХ СОБЫТИЙ

УДК 681.518

Борис Николаевич Чугаев
к. т. н., доцент кафедры вычислительные машины, системы и сети Московского авиационного института (МАИ)
Тел. 8 (499) 158-43-82
Эл. почта: b.915@yandex.ru

Александр Юрьевич Аржененко
д. ф-м. н., профессор кафедры Информационные технологии Московского авиационного института (МАИ)

В статье излагается метод оптимизации идентификации случайных событий с помощью теории вопросников. Теория вопросников – раздел математики, возникший на стыке математической статистики, комбинаторики, теории графов и теории информации. Её основоположник К.Ф.Пикар разработал эту теорию как аппарат описания логических взаимосвязей задач, методов их решения и условий, при которых принимаются эти решения. П.П. Пархоменко предложил использовать теорию вопросников как математический аппарат для наглядного построения, представления и описания оптимальных условных алгоритмов диагностирования.
Основной задачей теории вопросников стало построение оптимального в некотором смысле вопросника.

Ключевые слова: вопросник, стоимость вопросника, событие, весовая функция события, трудоемкость алгоритма.

Boris N. Chugaev
PhD in Engineering, Associate Professor, Department of computers, systems and networks, Moscow Aviation Institute (MAI)
Tel. 8 (499) 158-43-82
E-mail: b.915@yandex.ru

Alexander Y. Arzhenenko
Doctorate of Physics and Mathematics, Professor, Department of Information Technologies, Moscow Aviation Institute (MAI)

OPTIMAL IDENTIFICATION OF RANDOM EVENTS

The paper presents a method for optimizing identification of random events by using the theory of questionnaires. Theory questionnaires - a branch of mathematics that arose at the turn of mathematical statistics, combinatorics, graph theory and information theory. Its founder K.F.Pikar developed this theory as a description of the device logical relationships of tasks and methods of their solutions and the conditions under which these decisions are made. PP Parkhomenko suggested using questionnaires theory as a mathematical device to illustrate the construction, presentation and description of the best conventional diagnostic algorithms. The main task of the theory of the questionnaires was to build optimal in some sense of the questionnaire.

Keywords: questionnaire, the cost of the questionnaire, event, the weight function of the events, complexity of the algorithm.

1. Введение

В журналах прикладного и теоретического характера часто встречаются задачи, относящиеся к теории принятия решений, теории дискретного поиска, теории искусственного интеллекта, технической диагностике цифровых систем и т.п., которые при различной постановке имеют схожее математическое описание. Часто для их решения используют единый математический аппарат. Таким аппаратом является теория вопросников [2,3,4].

Ниже излагаются основные понятия и определения теории вопросников.

Задано конечное множество Y, состоящее из N элементов y_j, именуемых событиями. Каждому событию $y_j \in Y$ приписана не отрицательная весовая функция $\omega(y_j)$, называемая весом события y_j. Задано также конечное множество T разбиений множества Y на классы, число элементов в множестве T есть $|T| = R$. Элементы t множества T называются вопросами. Число подмножеств $a(t)$, на которые вопрос t разбивает множество Y, называют основанием вопроса t. Очевидно, что $2 \le a(t) \le N$. Классы событий, которые множество Y разбивается вопросом $t \in T$, называются ответами или исходами вопроса t. Каждому вопросу $t \in T$ приписана положительная весовая функция $c(t)$, называемая стоимостью вопроса.

Описание разбиения множества событий Y вопросами из T на подмножества принято проводить при помощи анкеты. Анкета представляет собой матрицу A размерности $R \times N$, элементы которой Z_{ij} определяется значением исходного вопроса t_i для события y_j [2,4].

Совокупность вопросов $Q \in T$ и последовательности, в которых задаются эти вопросы, для полной идентификации каждого из N событий множества Y, образуют вопросник для Y. При одних и тех же множествах T и Y могут быть построены вопросники, отличающиеся последовательностью постановки, так и множествами поставленных вопросов. Поэтому возникает основная задача теории вопросников, а именно выбор оптимального, в некотором смысле, вопросника.

Вопросник представляется нагруженным орграфом $G(X, \Gamma)$, у которого

$$X: \{x \in (Q \cup Y)\};$$
$$x = y \in Y \Rightarrow |\Gamma y| = \varnothing;$$
$$x = t \in Q \Rightarrow |\Gamma t| = a(t);$$
$$Q \cap Y = \varnothing.$$

В графе $G(X,\Gamma)$ существует единственная вершина $x_o = t_1 \in Q$, для которой $|\Gamma^{-1}{}_{Xo}| = 0$. Очевидно, что такой граф является корневым деревом с корнем в вершине X_o, оканчивается в $X_к \in Y$.

Стоимостью пути из X_o в $X_к \in Y$ называют сумму стоимостей вопросов, принадлежащих этому пути:

$$C(t_1, x_k) = \Sigma C(d_i),$$

где $i = 0,\dots k$.

Пронормируем весовые функции $\omega(y)$ событий $y \in Y$ следующим образом:

$$p(y) = \omega(y) / W,$$

где

$$W = \Sigma\omega(y).$$
$$\Sigma p(y) = 1,\ 0 \le p(y) \le 1.$$

Таким образом, в качестве весовой функции $p(y)$ можно задавать вероятности событий у полной системы событий Y. Такой подход наиболее распространен в теории вопросников [6].

Затраты на идентификацию событий в целом (стоимость вопросника), определяется как математическое ожидание идентификации всех событий из Y. Другими словами, под стоимостью под стоимостью обхода графа $G(X, \Gamma)$:

$$C(t_1, Y) = \Sigma c(t_1, y_j)p(y_j).$$

Вопросник с равными основаниями $a(t)$ всех вопросов называется однородным, а однородный вопросник, у которого $a(t) = 2$ – дихотомичным или бинарным. Бинарному вопроснику соответствует бинарное дерево.

Таблица 1.

Вопросы T	События Y.							Стоимости $C(t)$
T_i	y_1	y_2	y_3	y_4	y_5	y_6	y_7	$C(t_i)$
t_1	0	0	0	0	1	1	1	2
t_2	0	1	0	1	0	0	1	1
t_3	0	0	0	0	0	0	0	3
t_4	0	1	1	0	1	0	0	2
t_5	0	0	1	0	1	0	1	4
Вероятности $p(y_j)$	0,1	0,15	0,07	0,1	0,2	0,08	0,3	

Отметим, что вопросник с любым основанием $a(t)$ достаточно просто сводится к бинарному вопроснику [4,6].

В дальнейшем будем рассматривать бинарные вопросники.

Минимальное число вопросов k, необходимое для полной идентификации N событий, определяется из соотношения

$$K = [\log 2^N],$$

где $[\log 2^N]$ – наибольшее целое данного выражения.

Очевидно, что неизбыточное множество вопросов T, позволяющее идентифицировать каждое событие $y \in Y$, задается таким образом, чтобы $|T| \geq K$. Если $|T| = K$, то вопросник называется компактным. В противном случае ($|T| > K$) – вопросник некомпактный.

Рассмотрим в качестве примера построение вопросника по заданной анкете (таблица. 1).

Как уже отмечалось, одной и той же анкете соответствуют различные вопросники. На рис. 1 приведены три варианта возможных вопросников, построенных по заданной анкете (таблица 1). Приведенные варианты вопросников отличаются последовательностью задаваемых вопросов.

Приведенные на рис. 1 варианты вопросников отличаются последовательностью задаваемых вопросов.

Подсчитаем стоимость вопросника, приведенного на рис. 1 *а*.

$$\begin{gathered} Ca(Y) = [c(t_1) + c(t_2) + c(t_3)] \times [p(y_1) + \\ + p(y_3)] + [c(t_1) + c(t_2) + c(t_4)] \times \\ \times [p(y_2) + p(y_4)] + [c(t_1) + c(t_5)]p(y_6) + \\ + [c(t_1) + c(t_5) + c(t_4)] \times [p(y_7) + p(y_5)] = \\ = (2 + 1 + 3)(0{,}1 + 0{,}07) + (2 + 1 + 2) \times \\ \times (0{,}15 + 0{,}1) + (2 + 4) \times 0{,}08 + \\ + (2 + 4 + 2)(0{,}3 + 0{,}2) = 6 \times 0{,}17 + \\ + 5 \times 0{,}25 + 6 \times 0{,}08 + 8 \times 0{,}5 = 6{,}75. \end{gathered}$$

Подсчитав аналогично стоимость вопросников, приведенных на рис. 1*б*, и рис.1*в*, получим соответственно $Cб(Y) = 7{,}87$ и $Cв(Y) = 8{,}51$.

Из приведенного примера видно, что стоимость вопросника существенно зависит от последовательности задания вопросов.

Выше отметилось, что вопросник, имеющий минимальную стоимость, называется оптимальным.

2. Алгоритм оптимизации вопросника

Для построения оптимального вопросника при наличии полной системы бинарных вопросов и вероятностей событий, предлагается метод сокращений.

Очевидно, что вопросник с максимальной стоимостью можно построить, задавая все вопросы из T для каждого события $y_j \in Y$. Стоимость такого вопросника:

$$C(Y) = \Sigma c(t_i)p(y_j).$$

Длина пути от корня дерева x_1 до любой конечной вершины в вопроснике G равна R. Следовательно, число висячих вершин S в бинарном вопроснике определяется как:

$$S = 2^R$$

Но множество событий Y состоит из N элементов. Остальные $F = S - N$ событий образуют подмножество событий с нулевыми вероятностями. Все события, принадлежащие этому подмножеству, обозначим y_0.

Будем поочередно удалять из воп-

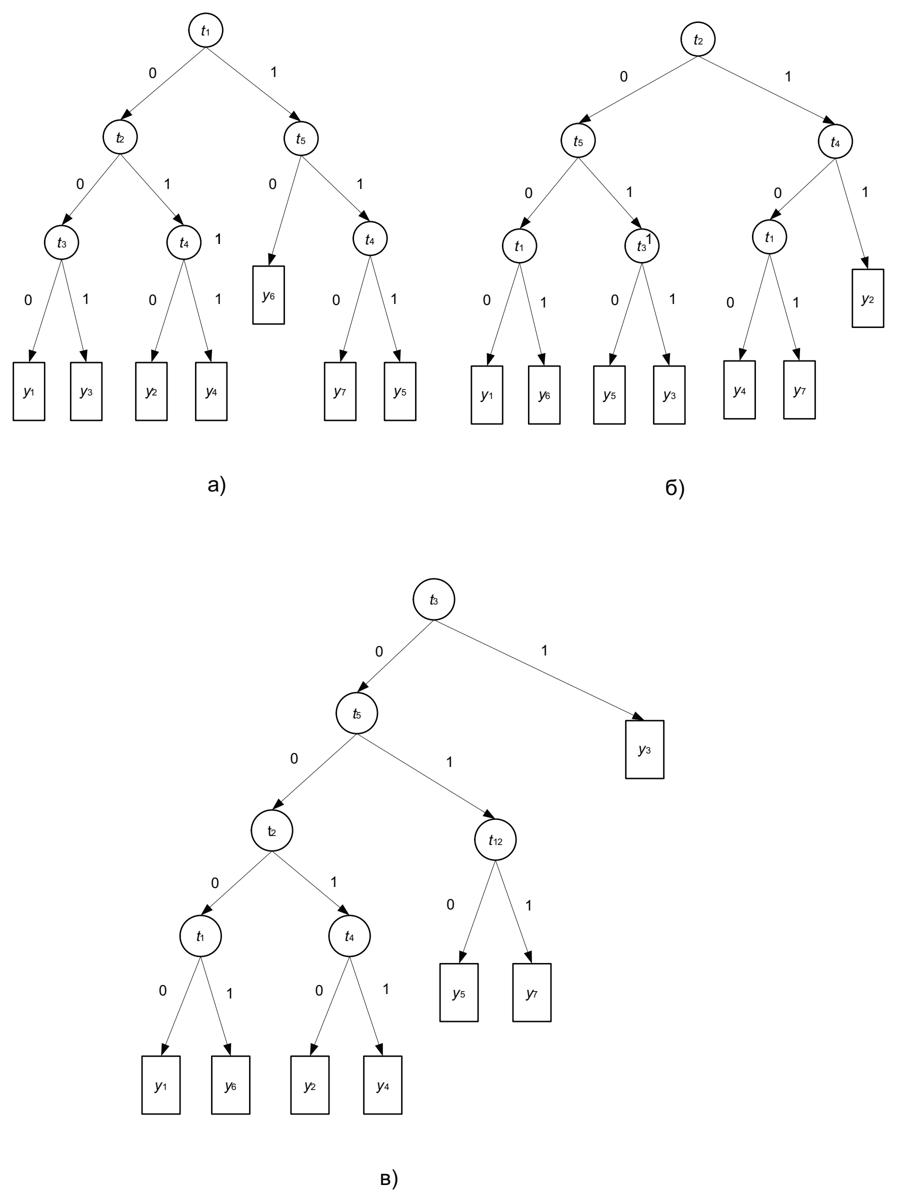

Рис. 1. Варианты возможных вопросников

росника такие вопросы, у которых один из исходов на рассматриваемом множестве событий принадлежит подмножеству F.

Удалив вопрос, стоящий на (r-1) ранге вопросника, число висячих вершин, имеющих нулевые вершины, сократится на величину

$$2^{R-r-2},$$

а стоимость вопросника при этом уменьшится на величину

$$c(Y', Z) = C(t_{r-1})p(Y').$$

Из сказанного следует, что оптимальный вопросник $G1$ можно построить в том случае, если удаляя все вершины $Y_j \in F$ получим минимум $C(G)$. Этого можно достичь последовательно удаляя из вопросника наибольшее слагаемые $C(y_j, t_r) = C(t_{r-1})p(y_j)$.

Исходя из сказанного, вопросник $G1$ строится так, чтобы вопросы, имеющие меньшую стоимость, имели бы меньший ранг в вопроснике. Упорядочим все вопросы в порядке возрастания их цен.

Очевидно, сокращение стоимости вопросника следует начинать с удаления максимального слагаемого $C\max(t_{r-2})P\max(y_j)$. При этом множество F уменьшится на единицу, т.е. $F1 = F$-1, а вершина y_j переместится с ранга r на ранг (r-1). Отметим, что удалив теперь из стоимости $C(G1)$ слагаемое $C(tr\text{-}1)P\max(yj)$, число нулевых вершин Fj сократится на 2 единицы. Вообще говоря, сокращение каждой из вершин $y_0 \in F$ уменьшает $C(G1)$ на величину

$$C(tr\text{-}2)P\max(yj)/2$$

Поставим в соответствие каждому событию y_j некоторое число, определяемое как

$$\delta r(y_j) = c(t_{r-1})p(y_j)/2^{R-r-1},$$

которое назовем функцией сокращения для события y_j имеющего ранг r в формируемом вопроснике $r = R\text{-}m\text{-}1$.

Коэффициентом сокращения $\delta_r(y^j)$ для рассматриваемого события y^j в вопроснике $G1$ $\delta_r(y^j) = 2m$.

Удаляя из вопросника $G1$ вопросы, определяющие максимальное слагаемое в сумме $C(G1)$, удовлетворяем при этом соответствующее число вершин y_0 до тех пор, пока коэффициент сокращения $\delta_{r\min}(y_j)$ для события y_j, имеющего минимальный ранг, не окажется больше числа $F_к$ оставшихся вершин с нулевыми вероятностями. На этом этапе возникает задача выбора $Mк$ чисел $\delta_r(y_j)$ так, чтобы их сумма была максимальной.

Достаточно просто доказать, что эта задача является обобщением известной NP – полной задачи «целочисленный» рюкзак и, следовательно, тоже NP полной [3]. Это означает, что оптимальный вопросник не может быть построен за полиноминальное, относительно числа событий, число шагов.

Таким образом, метод сокращения позволяет построить оптимальный вопросник, но при этом метод имеет экспоненциальную трудоемкость.

Приведем алгоритм оптимизации по методу сокращений.

Алгоритм:

1. Вычислим значение функции сокращения $\delta_r(y_j)$ для каждого события Y; $F_0 = 2^r - N$; $i = 0$.
2. Выбрать событие y_j для которого $\delta_r(y_j)$ имеет максимальное значение и $\delta_r(y_j) \le F_i$.
3. Удалить последний вопрос t_k стоящий на пути от корня вопросника к висячей вершине $y_j \in F$.
4. Вычислить новые значения $\delta_{r-1}(y_j)$, $\zeta_r(y_j)$, F_i.
5. Если $F_i = 0$, то алгоритм заканчивает работу.
6. Если $\delta_{r\min}(y_n) < M_i$, то перейти к пункту 2. В противном случае методом перебора выделить F_i значений функции сокращения, образующих максимальную сумму и провести удаление соответствующих вопросов.

Для получения алгоритма более, приемлемого с точки зрения трудоемкости, заменим в приведенном алгоритме пункт 6 безусловным переходом к пункту 2. В результате получим алгоритм построения квазиоптимального вопросника, который будет иметь полиномиальную трудоемкость [2].

3. Заключение

Большое распространение теория вопросников получила в технической диагностике. В этом случае множеству вопросов ставится в соответствие множество тестов, а множеству событий – множество неисправностей технической системы. Такое применение теории вопросников объясняется тем, что в настоящее время большое внимание уделяется быстрому и точному определению исправности, работоспособности и правильности функционирования технических систем. Особенно это касается автоматических систем управления, работающих в реальном масштабе времени [4,5].

Метод сокращения может быть использован для точной оценки нижней границы стоимости условных алгоритмов диагностирования технических устройств при полном множестве текстов различной стоимости, а так же для оптимальной стратегии идентификации множества неисправностей и системы цен.

Литература

1. Алискеров Ф.Т., Хабина Э.Л., Шварц Д.А. Бинарные отношения, графы и коллективные решения. – М.: ГУ ВШЭ, 2006. – 300с.

2. Аржененко А.Ю., Чугаев Б.Н. Оптимальные бинарные вопросники. – М.: Энергоатомиздат, 1989.-128 с.

3. Пападимитриу Х., Стайглиц К. Комбинаторная оптимизация. Алгоритмы и сложность: перев. с англ. – М.: Мир, 1985. 512с.

4. Пархоменко П.П., Согомонян Е.С. Основы технической диагностики. – М.: Энергоатомиздат, 1981. – 410 с.

5. Филин В.М., Пчелинцев Л.А., Денчик В.Н., Задеба В. Оптимизация диагностики космического разгонного блока. – М.: УРСС, 2004. – 184 с.

6. Picard C. F. Theorie des questionnaires. Paris: Ganthier-Villars. 1972. P. 182.

References

1. Aliskerov F.T., Habina Je.L., Shvarc D.A. Binarnye otnoshenija, grafy i kollektivnye reshenija. – M.: GU VShJe, 2006. – 300s.

2. Arzhenenko A.Ju., Chugaev B.N. Optimal'nye binarnye voprosniki. – M.: Jenergoatomizdat, 1989.-128 s.

3. Papadimitriu H., Stajglic K. Kombinatornaja optimizacija. Algoritmy i slozhnost': perev. s angl. – M.: Mir, 1985. 512s.

4. Parhomenko P.P., Sogomonjan E.S. Osnovy tehnicheskoj diagnostiki. – M.: Jenergoatomizdat, 1981. – 410 s.

5. Filin V.M., Pchelincev L.A., Denchik V.N., Zadeba V. Optimizacija diagnostiki kosmicheskogo razgonnogo bloka. – M.: URSS, 2004. – 184 s.

6. Picard C. F. Theorie des questionnaires. Paris: Ganthier-Villars. 1972. P. 182.