ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 1986
Seria: AUTOMATYKA z. 84 Nr kol.894

Lech Jamroż
Politechnika Krakowska

BŁĄD OBCIĘCIA W PRZYBLIŻONYCH ROZWIĄZANIACH ZAGADNIEŃ KOMBINATORYCZNYCH

Streszczenie.Otrzymywanie rozwiązań dokładnych "dużych" zagadnień kombinatorycznych jest nie tylko nieefektywne ale często niemożliwe.W zasadzie w takich przypadkach korzystamy z rozwiązań przybliżonych.Powstaje wówczas problem szacowania dokładności tych rozwiązań w stosunku do rozwiązań dokładnych.W artykule przedstawiono podejście do zagadnienia szacowania rozwiązań przybliżonych oparte na metodzie statystycznej.

1. Wstęp

Wiele praktycznych problemów NP-zupełnych, np. problem szeregowania, komiwojażera , załadunku,trójpodziału itp. charakteryzują się kombinatoryczną /dyskretną/ strukturą.Fakt ten umożliwił,między innymi,zastosowanie optymalizacji kombinatorycznej do rozwiązywania tych problemów. Zagadnienia kombinatoryczne,często proste w sformułowaniu,nie są łatwe do rozwiązania ze względu na liczbę rozwiązań dopuszczalnych.

Otrzymywanie optymalnych rozwiązań "dużych" zagadnień kombinatorycznych,dotyczy to przede wszystkim zagadnień praktycznych,jest nie tylko nieefektywne ale często niemożliwe.W zasadzie w takich przypadkach korzystamy z rozwiązań przybliżonych,traktując metody przybliżone lub heurystyczne jako podejście do rozwiązywania problemów kombinatorycznych. Wówczas powstaje zagadnienie szacowania uzyskiwanych rozwiązań w stosunku do rozwiązań optymalnych.

Znanych jest kilka podejść do tego zagadnienia.Wymienić należy przede wszystkim testowanie empiryczne [1],ocenę najgorszego przypadku [3] czy też analizę statystyczną [6,7] .W niniejszym artykule zostanie przedstawione podejście do szacowania dokładności rozwiązań przybliżonych problemów kombinatorycznych w przypadku minimalizacji,oparte na metodzie statystycznej.Przypadek maksymalizacji można znaleźć w pracy [7].

Przedstawiona metoda polega na wyznaczeniu rozkładu granicznego dla estymatora nieznanej wartości określonego parametru,będącego z założenia rozwiązaniem dokładnym.Wartość ta wyznacza tzw. punkt cięcia /truncation point/ w zbiorze wartości rozwiązań przybliżonych.
W celu wyznaczenia granicznego rozkładu wspomnianego estymatora określona zostanie funkcja rozkładu występowania poszczególnych rozwiązań.

2. Funkcja szacowania dokładności i jej podstawowe własności.

W zagadnieniach aproksymacyjnych,czy też identyfikacji,zwykle bada się określone cechy elementów pewnego zbioru zwanego populacją /populacją generalną/ na podstawie zaobserwowanych cech zestawu elementów wylosowanych z tego zbioru.Stwierdzenie,że elementy badane są losowane z populacji opiera się na założeniu,że badana cecha jest zmienną losową tzn. istnieje rozkład prawdopodobieństwa,który charakteryzuje daną populację.W dalszej części artykułu,realizacją wspomnianej zmiennej losowej będzie wartość funkcji celu wyznaczonej dla konkretnego rozwiązania problemu kombinatorycznego.Oznaczmy przez

$$x_1 \ , \ x_2 \ , \ \ldots\ldots\ldots\ldots \ , \ x_n \qquad /1/$$

ciąg otrzymanych n wartości rozwiązań przybliżonych.Ciąg ten traktujemy jako n zaobserwowanych wartości wspomnianej zmiennej losowej,którą oznaczymy przez X.Ciąg wartości /1/ nazwiemy n elementową próbą losową. Oznacza to,że wartości te są jakby "losowane" z pewnych zbiorów liczbowych o określonych własnościach,a dokładniej,że dla każdego n istnieje rozkład prawdopodobieństwa występowania poszczególnych wartości zmiennej losowej X.Gdybyśmy wielokrotnie powtarzali doświadczenie polegające na wyborze $x_1,x_2, \ldots ,x_n$,to otrzymalibyśmy za każdym razem inne wartości. W ogólności zestaw liczb postaci /1/ można zatem uważać jako zestaw wartości zaobserwowanych zmiennych losowych

$$X_1 \ , \ X_2 \ , \ \ldots\ldots\ldots\ldots \ , \ X_n \qquad /2/$$

Wprowadźmy za pracą [7] uporządkowany ciąg statystyczny

$$X_{r_1} < X_{r_2} < \ldots\ldots\ldots < X_{r_n} \qquad /3/$$

zmiennych losowych postaci /2/.
Oznaczmy przez

$$x_{r_1} < x_{r_2} < \ldots\ldots\ldots\ldots < x_{r_n} \qquad /4/$$

uporządkowany ciąg wartości będący realizacją ciągu zmiennych losowych postaci /3/. Załóżmy tymczasową znajomość empirycznego rozkładu prawdopodobieństwa występowania odpowiednich wartości ciągu /4/.
Znajomość tę przyjmujemy jako informację aprioryczną do wyznaczenia empirycznej dystrybuanty wspomnianego rozkładu.
Przyjęcie określonej postaci tego rozkładu może być wynikiem arbitralnego założenia/np. rozkład jest normalny/ albo też może być rezultatem obserwacji dużej liczby wartości.
W drugim przypadku na podstawie wyników obserwacji ustalamy empiryczny rozkład prawdopodobieństwa będący przybliżeniem rozkładu dokładnego.

Oznaczmy przez θ nieznany parametr, zwany punktem cięcia w zbiorze wartości rozwiązań, będący z założenia rozwiązaniem dokładnym. Dla wyznaczenia wartości θ określmy graniczny rozkład dla estymatora tej wartości. W zadaniach minimalizacji, parametr θ identyfikujemy z lewostronnym cięciem w zbiorze wartości rozwiązań takim, że $\theta < x_i$, $i = \overline{1,n}$. Określmy teraz postać empirycznej dystrybuanty zmiennej losowej X i oznaczmy ją przez $F(x)$. Natomiast przez $F_\theta(x)$ oznaczmy dystrybuantę charakteryzującą ciąg postaci /4/, w którym poprowadzono cięcie na poziomie wartości θ. Jako empiryczne prawdopodobieństwo występowania poszczególnych wartości ciągu /4/ przyjmuje się częstość ich występowania, tj.

$$F(x) = P(X_i < x) \tag{/5/}$$

A zatem dystrybuantę empiryczną można otrzymać przez zastąpienie prawdopodobieństwa $P(X_i < x)$ odpowiednią częstością. Jeżeli założymy, że $K(x)$ jest liczbą elementów ciągu /4/ o wartościach mniejszych od x to wówczas

$$F(x) = \frac{K(x)}{n} \tag{/6/}$$

Określmy teraz postać funkcji $F_\theta(x)$.
W związku z tym poprowadźmy cięcie w zbiorze wartości /4/ na poziomie x_{r_k}. Oznaczając przez $X^k = \{x_{r_{k+1}}, \ldots\ldots, x_{r_n}\}$ mamy

$$F(\theta) = \frac{K(\theta)}{n} = \frac{k}{n},$$

natomiast

$$F_\theta(x) = P(X_{r_i} < x \mid x_{r_i} \in X^k) = \frac{K(x) - k}{n-k} = \frac{n\,K(x)}{n(n-k)} - \frac{k}{n-k} = \tag{/7/}$$

$$= \frac{F(x)}{\frac{n-k}{n}} - \frac{k}{n-k} = \frac{F(x) - F(\theta)}{1-F(\theta)}$$

Z powyższej zależności widać, że $0 \leqslant F_\theta(x) \leqslant 1$ dla $x \in \langle 0, x_{r_n} \rangle$ a więc $F_\theta(x)$ jest dystrybuantą.

Dla zadania maksymalizacji postępujemy podobnie. Identyfikujemy wartość parametru θ z prawostronnym cięciem. Przyjmując $X^k = \{x_{r_1}, \ldots x_{r_k}\}$ otrzymujemy

$$F_\theta(x) = \frac{K(x)}{k} = \frac{K(x)}{k}\,\frac{n}{n} = \frac{F(x)}{F(\theta)}, \tag{/8/}$$

co jest zgodne z postacią funkcji otrzymanej w pracy [7].

Funkcja $F_\theta(x)$ postaci /7/ jako funkcja ciągła i różniczkowalna jest rozwijalna w szereg Taylora .

Oznaczmy przez $E_n(X_1, X_2, \ldots\ldots, X_n)$ /9/

funkcję szacowania wartości parametru θ i nazwijmy ją estymatorem.

Estymator E_n posiada następujące własności:

i/ $\lim_{n \to \infty} E_n(X_1, X_2, \ldots, X_n) = \theta$

ii/ $E\, E_n(X_1, X_2, \ldots, X_n) = \theta$

iii/ E_n - jest funkcją rozwijalną w szereg Taylora w otoczeniu θ .

Własność i/ oznacza, że dla dużej liczebności próby wartość szacowana zmierza w probabilistycznym sensie do rzeczywistej /dokładnej/ wartości parametru θ tzn. $\lim_{n \to \infty} (P|E_n - \theta| > \varepsilon) = 0$.

Własność ii/ oznacza, że dla wielokrotnie powtarzanych doświadczeń przy tej samej liczebności próby średnia arytmetyczna z oszacowań zdąża do wartości θ, natomiast własność iii/ wymaga od estymatora E_n przynależności do klasy funkcji ciągłych i różniczkowalnych.

3. Graniczny rozkład estymatora

Nie zawsze mamy do czynienia z populacją generalną, która ma rozkład normalny. Często jednak populacja ta ma rozkład odmienny od normalnego. W zasadzie w takich przypadkach trudno jest wyznaczyć rozkład interesującego nas estymatora. Nawet jeśli rozkład taki wyznaczymy to z reguły jest on na tyle skomplikowany, że trudno jest w praktyce z niego korzystać. Z drugiej strony często rozkład populacji generalnej jest nieznany. Zarówno w jednym jak i w drugim przypadku można wyznaczyć, z czego skorzystamy, graniczny rozkład estymatora przy $n \to \infty$, gdzie n jest wielkością próby, na podstawie której określamy wymagany estymator.

Korzystanie z rozkładu granicznego estymatora ma ten praktyczny sens, że przy dostatecznie dużym n możemy zastąpić nieznany rozkład rozkładem granicznym.

Wyznaczenie rozkładu granicznego estymatora wiąże się z rozwinięciem dystrybuanty postaci /7/ w otoczeniu punktu θ. Z własności dystrybuanty mamy

$$y = F_\theta(x) \qquad x \in \langle 0 , X_{r_n} \rangle \qquad y \in \langle 0 , 1 \rangle \tag{10}$$

Wprowadźmy funkcję odwrotną do funkcji F_θ, tzn.

$$R_{F\theta} = F_\theta^{-1} \tag{11}$$

wówczas:

$$x = F_\theta^{-1}(y) = R_{F\theta}(y) \tag{12}$$

Podamy teraz niektóre własności funkcji $R_{F\theta}(y)$.

W ł a s n o ś ć 1/

Funkcja $R_{F\theta}$ spełnia następujące warunki

i/ $R_{F\theta}$ – należy do klasy funkcji rozwijalnych w szereg Taylora ,

ii/ $R_{F\theta}(y) = \theta$ dla $y=0$,

iii/ $x = R_{F\theta}(y)$. □

Dowód. Warunek i/ wynika z faktu, że $R_{F\theta}$ jest funkcją odwrotną do funkcji ciągłej i różniczkowalnej, a więc jest również rozwijalną w szereg Taylora, natomiast ii/ oraz iii/ wynika bezpośrednio z /12/ cnd.

W celu wyprowadzenia zależności na wartość parametru θ skorzystamy z własności 1/ rozwijając funkcję $R_{F\theta}(y)$ w otoczeniu punktu $y=0$, wówczas

$$x = R_{F\theta}(0) + y\, R'_{F\theta}(0) + \frac{y^2}{2!} R''_{F\theta}(0) + \ldots + \frac{y^k}{k!} R^{k}_{F\theta}(0) + \ldots \qquad /13/$$

Wyznaczmy wartości pochodnych funkcji $R_{F\theta}$ /ograniczymy się do 3 pochodnej/.

$$F'_{\theta}(x) = \frac{d\, F_{\theta}(x)}{d\, x} = \frac{F'(x)}{1-F(\theta)} \qquad /14/$$

$$F''_{\theta}(x) = \frac{d\, F'_{\theta}(x)}{d\, x} = \frac{F''(x)}{1-F(\theta)} \qquad /15/$$

$$F'''_{\theta}(x) = \frac{dF''_{\theta}(x)}{d\, x} = \frac{F'''(x)}{1-F(\theta)} \qquad /16/$$

Traktując $R_{F\theta}(y)$ jako funkcję uwikłaną względem zmiennej x otrzymujemy

$$R'_{F\theta}(y) = \frac{1}{F'_{\theta}(x)} = \frac{1-F(\theta)}{F'(x)} \qquad /17/$$

Dla pozostałych pochodnych ze względu na objętość wyprowadzeń, ograniczymy się do podania końcowych zależności.

$$R''_{F\theta}(y) = \frac{d}{d\,y} R'_{F\theta}(y) = \frac{d}{d\,y}\left[\frac{1}{F'_{\theta}(R_{F\theta}(y))}\right] = -\frac{F''(x)}{F'(x)}\left[\frac{1-F(\theta)}{F'(x)}\right]^2 \qquad /18/$$

$$R'''_{F\theta}(y) = \frac{d}{d\,y}\left[R''_{F\theta}(y)\right] = \frac{1-F(\theta)}{F'^{5}(\theta)}\left\{F''^{2}(\theta) + 2F''(\theta)\left[1-F(\theta)\right] - F'(\theta)\,F'''(\theta)\right\} \qquad /19/$$

Wstawiając wyznaczone wartości pochodnych do zależności /13/ określimy wartość przeciętną zmiennej X_{r_1}.

$$E X_{r_1} = \theta + R'_{F\theta}(0)\, E\, Y_{r_1} + \frac{R''_{F\theta}(0)}{2!} E\, Y^2_{r_1} + \dots + \frac{R^k_{F\theta}(0)}{k!} E\, Y^k_{r_1} + \dots \qquad /20/$$

W celu wyznaczenia $E\, Y^k_{r_1}$ zauważmy następującą własność

W ł a s n o ś ć 2/

Jeżeli ciąg $y_{r_1} = F_\theta(x_{r_1})$, $i = \overline{1,n}$ jest ciągiem wartości o rozkładzie równomiernym na przedziale $\langle 0,1 \rangle$ oraz $R^{-1}_{F\theta} = R^{-1}_\theta$ to ciąg $x_{r_1} = R_{F\theta}(y_{r_1})$ jest ciągiem wartości o rozkładzie z dystrybuantą $F_\theta(x_{r_1})$. □

Dowód. Zauważmy, że dla $x_{r_1} = R_{F\theta}(y_{r_1})$, $i = \overline{1,n}$ słuszna jest następująca zależność $P(X < x_{r_1}) = P(R_{F\theta}(Y) < x_{r_1}) = P(F_\theta(R_{F\theta}(Y)) < F_\theta(x)) =$

$$= P(Y < F_\theta(x_{r_1})) = F_\theta(x_{r_1}),$$

z której wynika, że zmienna X ma rozkład o dystrybuancie F_θ, cnd.

Zmienne losowe X_{r_1} z dystrybuantą $F_\theta(x_{r_1})$, $i = \overline{1,n}$ są przekształcane w zmienne losowe Y_{r_1}, $i = \overline{1,n}$ o rozkładzie równomiernym na przedziale $\langle 0,1 \rangle$.

Korzystając z powyższego, wartości $E(Y^k_{r_1})$ w zależności /20/ są k-tymi momentami zmiennej losowej Y_{r_1} o rozkładzie równomiernym.

Dla wygody wyraźmy k-ty moment $Y^k_{r_1}$ za pomocą funkcji gamma Γ wówczas

$$E\, Y^k_{r_1} = \int_0^1 Y^k_{r_1}(1 - Y_{r_1})^{n-1} = \frac{\Gamma(k+1)\,\Gamma(n)}{\Gamma(n+k+1)} = \frac{k!}{n \dots (n+k)} = k!\, n^{-k} \qquad /21/$$

Wstawiając /21/ do /20/ otrzymujemy

$$E X_{r_1} = \theta + n^{-1} R'_{F\theta}(0) + n^{-2} R''_{F\theta}(0) + n^{-3} R'''_{F\theta}(0) + \dots \qquad /22/$$

Analogicznie można wyprowadzić zależności na wartość $E X_{r_2}$

$$E X_{r_2} = \theta + 2\, n^{-1} R'_{F\theta}(0) + 3\, n^{-2} R_{F\theta}(0) + \dots + (k+1) n^{-k} R^k_{F\theta}(0) \qquad /23/$$

Dysponując ciągiem /3/, naturalnym będzie jeżeli estymator $E_n(X_1, X_2 \dots X_n)$ określimy w postaci

$$E_n(X_1, X_2, \dots, X_n) = X_{r_1} \qquad /24/$$

wówczas

$$E\, E_n(X_1, X_2 \ldots X_n) = E(X_{r_1}) = \theta + \frac{1}{n} + \frac{2}{n^2} + \ldots + \frac{k}{n^k} \qquad /25/$$

Oznaczmy przez E_n^k estymator szacujący wartość θ z dokładnością $\frac{\varrho^k}{n^k}$.
Jeżeli funkcja estymatora daje się przedstawić w postaci szeregu /25/, to E_n^k można wyrazić następującą zależnością [7]

$$E_n^k = \sum_{v=0}^{k} (-1)^v \binom{k+1}{v+1} X_{r_{v+1}} \qquad /26/$$

Powyższy wzór pozwala oszacować wartość parametru θ.
Zilustrujmy to przykładem wyznaczając wartość estymatora E_n^1.
Z zależności /26/ otrzymujemy

$$E_n^1 = \binom{2}{1} X_{r_1} - \binom{2}{2} X_{r_2} = 2\, X_{r_1} - X_{r_2}$$

zatem

$$E\ E_n^1 = 2\, E(X_{r_1}) - E\, X_{r_2} = 2(\theta + n^{-1} R' + n^{-2} R'' + \ldots +) - \theta + 2n^{-1} R' +$$
$$+ 3\, n^{-2} R'' + 4\, n^{-3} R''' + \ldots + \ldots) = \theta - 2\, n^{-2} R'' - 3\, n^{-3} R''' \ldots$$

$$\lim_{n \to \infty} E\ E_n^1 = \theta \qquad /27/$$

Z drugiej strony jeżeli estymator E_n jest postaci /24/ wówczas [2]

$$E_n^1 = n\, E_n - \frac{n-1}{n} \Big(\sum_{i}^{n} E_{n-1,i} \Big), \qquad /28/$$

gdzie $E_{n-1,i} = E_n(x_{r_1}, \ldots, x_{r_{i-1}}, x_{r_{i+1}}, \ldots, x_{r_n})$ jest wartością estymatora dla przypadku wyeliminowania z ciągu /4/ wartości x_{r_i}.
Z uporządkowania tego ciągu mamy również następującą zależność:

$$E_{n-1,i} = \min(x_{r_1}, \ldots, x_{r_{i-1}}, x_{r_{i+1}}, \ldots, x_{r_n}) = \begin{cases} x_{r_1} & \text{dla } x_{r_1} = x_{r_i} \\ x_{r_2} & \text{dla } x_{r_i} = x_{r_1} \end{cases} \qquad /29/$$

Wobec powyższego

$$E_n^1 = n\, x_{r_1} - \frac{n-1}{n} \big((n-1) x_{r_1} + x_{r_2} \big) \qquad /30/$$

$$\lim_{n\to\infty} E_n^1 = \lim_{n\to\infty} \frac{2n-1}{n} x_{r_1} - \frac{n-2}{n} x_{r_2} = 2x_{r_1} - x_{r_2} \qquad /30/$$

Zgodnie z /27/ oraz /30/ otrzymujemy zależność na wyznaczenie wartości parametru θ

$$\theta = 2x_{r_1} - x_{r_2} \qquad /31/$$

4. Uwagi końcowe

W artykule przedstawiono metodę statystycznego oszacowania rozwiązań przybliżonych problemów kombinatorycznych.Zasadniczym przedmiotem uwagi były problemy,dla których istniejące algorytmy pozwalają szybko uzyskać zbiór wartości dopuszczalnych rozwiązań.
Podane wyrażenie /26/ na wyznaczenie wartości parametru θ /estymatora E_n^k/ będącego z założenia rozwiązaniem dokładnym,nie wymaga znajomości rozkładu prawdopodobieństwa występowania poszczególnych wartości rozwiązań i jest proste w zastosowaniu praktycznym.

LITERATURA

[1] Dannenbrig D.G.: A evaluation of flow shop sequencing problems, Management Science Vo 23,No 11,1977.

[2] Dannenbrig D.G.: Procedures for estimating optimal solution values for large combinatorial problems,Management Science Vo 23,No 12,1977.

[3] Fisher M.L.: Worst-case analysis heuristic algorithms,Management Science,Vo 26,No 1,1980.

[4] Jamroż L.: Heurystyczne podejście do zagadnienia szeregowania w dyskretnym procesie produkcyjnym,Zeszyty Naukowe AGH,No 866,z. 146,1981.

[5] Jamroż L.: Zagadnienie szacowania dokładności rozwiązań problemów kombinatorycznych,Zeszyty Naukowe AGH,No 973 , z. 34,1983.

[6] Markov I.M.: Statical error estimation in ekstremum problems,Avtom. Telemechanika,Vo 42,No 3,1981.

[7] Robson D.S., Whitlock J.H.: Estimation of a truncation point, Biometrika,Vo 51,1974.

Recenzent: Prof.dr inż.Henryk Kowalowski

Wpłynęło do Redakcji do 1986.04.30

ОШИБКА УРЕЗАНИЯ В ПРИБЛИЖЕННЫХ РЕШЕНИЯХ КОМБИНАТОРНЫХ ЗАДАЧ

Р е з ю м е

В работе рассматривается оценка стоимости приближенных решений комбинаторных задач.Представленная методика основывается на определении предельного распределения для оценки неизвестной стоимости параметра, выходя из предпосылки , которым является точное решение. Эта стоимость в множестве приближенных решений определяет точку разреза . В задачах минимизации во внимание принимается левостороннюю точку разреза.

Введённая оценка основана на статистической последовательности приближенных решений.

TRUNCATION ERROR IN SUBOPTIMAL SOLUTIONS OF COMBINATORIAL PROBLEMS

S u m m a r y

This paper deals with the statistical approach to estimation the value of suboptimal solutions of combinatorial problems. The wide class of NP - complete practical problems such as,sequencing problems,selection,knapsack, 3-partition etc. is characterized by combinatorial / discrete/ structure. This fact gives oportunity to utylize combinatorial optimization in solution of mentioned problems. The combinatorial problems often in form are difficult to solve because the number of admissible solution is finite but numerous.Obtaining of optimal solution of practical problems may be uneconomic or impossible.In this case estimation of suboptimal solutions is needed in relation to optimal solutions. The method is based on the boundary distribution of estimation of unknown value of parameter.This value determines the truncation point in the set of suboptimal solutions.In minimization problem we are concerned with the identification of the left truncation point.