Testowanie hipotezy o kompletności zbioru argumentów

Paweł Szołtysek

Relacja podobieństwa i testowanie kompletności zbioru argumentów

Relacja podobieństwa - założenia

- Proces jest opisany za pomocą funkcji wymiarowej, wymiarowo niezmienniczej i jednorodnej

$$Z = \Phi(A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_n)$$

- Proces, opisany tę samą funkcją, o argumentach równych wymiarom, jednak innych co do wartości:

$$\hat{Z} = \Phi(\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_n, \hat{B}_1, \hat{B}_2, \ldots, \hat{B}_n)$$

Co można zapisać jako

$$A_i = \alpha_i \hat{A}_i, B_j = \beta_j \hat{B}_j, i = 1,2,\ldots,m, \ j = 1,2,\ldots,r$$

Relacja podobieństwa – stosunek

Ile wynosi stosunek tych wartości?

$$\lambda = \frac{Z}{\hat{Z}}, \lambda \in \Pi_0$$

Z twierdzenia π:

$$Z = f(\varphi_1, \varphi_2, ..., \varphi_r)\prod_{i=1}^{m} A_i^{a_i}, \varphi_j = \frac{B_j}{\prod_{i=1}^{m} A_i^{a_{ji}}}, j = 1,2,...,r$$

$$\hat{Z} = f(\hat{\varphi}_1, \hat{\varphi}_2, ..., \hat{\varphi}_r)\prod_{i=1}^{m} \hat{A}_i^{a_i}, \hat{\varphi}_\mathrm{j} = \frac{\hat{B}_j}{\prod_{i=1}^{m} \hat{A}_i^{a_{ji}}}, j = 1,2,...,r$$

Czyli

$$\lambda = \frac{f(\varphi_1, \varphi_2, ..., \varphi_r)}{f(\hat{\varphi}_1, \hat{\varphi}_2, ..., \hat{\varphi}_r)}\prod_{i=1}^{m} \alpha_i^{a_i}$$

Relacja podobieństwa – definicja

Dwie realizacje procesu są podobne, jeśli spełnione jest kryterium podobieństwa:

$$\hat{\varphi}_j = \varphi_j, \; j = 1,2,\ldots,r$$

Zauważmy, że wtedy nasz stosunek

$$\lambda = \prod_{i=1}^{m} \alpha_i^{a_i}$$

nie zależy od postaci funkcji f.

Relacja podobieństwa a teza o kompletności zbioru argumentów

Z podobieństwa dwóch realizacji $\hat{R}_p, \hat{R}_t$

$$f(\varphi_{1p}, \varphi_{2p}, \ldots, \varphi_{rp}) = f(\varphi_{1t}, \varphi_{2t}, \ldots, \varphi_{rt})$$

czyli

$$\frac{Z_{0p}}{\prod_{i=1}^{m} Z_{ip}^{a_i}} = \frac{Z_{0t}}{\prod_{i=1}^{m} Z_{it}^{a_i}}$$

Jeśli zbiór argumentów funkcji jest kompletny, możemy obliczyć Z_{0p}, a także Z_{ot}. Oznacza to zależność między wejściem i wyjściem.

Kompletność zbioru argumentów

Zbiór argumentów funkcji wymiarowej, wymiarowo niezależnej i jednorodnej jest kompletnym opisem procesu wtedy i tylko wtedy gdy funkcja f jest różnowartościowa, i dla każdej pary podobnych realizacji procesu $\hat{R}_p, \hat{R}_t$ zachowana jest równość

$$\frac{Z_{0p}}{\prod_{i=1}^{m} Z_{ip}^{a_i}} = \frac{Z_{0t}}{\prod_{i=1}^{m} Z_{it}^{a_i}}$$

Zachodzi także twierdzenie odwrotne do powyższego.

Relacja podobieństwa i testowanie kompletności zbioru argumentów w przypadku wystąpienia błędów pomiaru

Problemy z błędami pomiaru

Mamy realizację $R_t = (Z_{1t},\dots,Z_{mt},\dots,Z_{m+r},Z_{0t})$

Możemy ją opisać podobnie jak wcześniej:

$$R_t = (Z_{1t},\dots,Z_{mt},\varphi_{1t}\dots,\varphi_{rt},\varphi_{0t})$$

Realizacje są podobne, gdy

$$\varphi_{j1} = \varphi_{j2},\ j = 1,2,\dots,r$$

Tak jednak nie będzie, ponieważ na pomiar wpływają błędy.

$$Z_{lt} = Z_{lt}^{*} + \Delta Z_l,\ l = 0,1,2,\dots,m+r,\ t = 0,1,2,\dots,q$$

Ustala się zwykle, że błąd ma rozkład normalny $N(0,\sigma^2)$. Jednocześnie, przez wzgląd na to, należy uznać, że φ_{jt} także jest zmienną losową.

Określanie przedziałów błędów

Możemy dla każdej realizacji znaleźć jednak taki przedział

$$[Z_{lt} - \Delta Z_l, Z_{lt} + \Delta Z_l], l = 0,1,\ldots, m+r, t = 1,2,\ldots, q$$

że wartość będzie należała do tego przedziału ze znanym, bliskim 1 prawdopodobieństwem. Znając te przedziały możemy zdefiniować podobnie takie przedziały dla $[\varphi_{jt}^{L}, \varphi_{jt}^{R}]$, też z prawdopodobieństwem bliskim 1.

Określanie metryki między relacjami

Bierzemy gęstości wystąpień tych parametrów, dla wszystkich realizacji, i budujemy metrykę

$$d(f,g) = 1 - \int_R \min(f,g)xdx$$

Metryka ta opisuje odległość między dwiema realizacjami jako

$$d(f_1, f_2) \leq \varepsilon_j,\ j = 1,\ldots,r$$

Tą metryką możemy podmienić warunek podobieństwa. Wartości epsilon są ustalane w sposób taki, by statystyczna ocena podobieństwa była później satysfakcjonująca.

Relacja podobieństwa

Aby określić które realizacje spełniają warunek podobieństwa, najpierw dzielimy je na klasy podobieństwa, a później oceniamy podobieństwo statystycznie w tych, czysto hipotetycznych klasach. W końcu, każdy taki zbiór jest ponownie dzielony, tym razem w relacji z parametrem fi.

W tym momencie nasz warunek musi być sprawdzony statystycznie. Procedura ta jest podzielona na dwa kroki – badanie jakości przeprowadzonego dzielenia oraz obliczenie miary podobieństwa.

Sformułowanie hipotezy o kompletności zbioru

Postawmy hipotezę. W każdej hipotetycznej klasie podobieństwa wartości parametrów fi są stałe, jeśli tylko j jest ustalone (nie jest zmienną).

Tworzymy zredukowaną realizację złożoną z m+1 parametrów potrzebnych do obliczenia wartości fi

$$R_{jt}^{e} = (Z_{1t}, \dots, Z_{mt}, Z_{m+j,t}); t = 1, \dots, q$$

Każda wartość parametru fi odpowiada w przestrzeni R^{m+1} m-wymiarowej hiperpłaszczyźnie

$$H_{\varphi_j} = \left\{ R_j^e = (Z_1, \dots, Z_m, Z_{m+j}) : \frac{Z_{m+j}}{\prod_{i=1}^{m} Z_i^{a_{ji}}} = \varphi_j \right\}$$

Transformacja zmiennych

Z hipotezy: wartości fi z każdej klasy są sobie równe

Ta wartość (oznaczmy φ_j^l) z kolei wyznacza nam nową hiperpłaszczyznę, którą mamy wyznaczyć.

Transformujemy zmienne Z umożliwiając linearyzację w sposób następujący:

$$L^j : R^{m+1} \to R^{m+1}$$

$$\dot{\varphi}_j = \dot{Z}_{m+j} - a_{j1}\dot{Z}_1 - \ldots - a_{jm}\dot{Z}_m$$

Gdzie $\dot{\varphi}_j = \ln \varphi_j$ opisuje konkretną hiperpłaszczyznę w przestrzeni R^{m+1}.

Modyfikacja metryki Euklidesa

Jeśli hipoteza jest prawdziwa, wtedy wszystkie transformowane realizacje $\dot{R}^e_{j1},\ldots,\dot{R}^e_{jq_b}$ leżały będą na tej samej hiperpłaszczyźnie określoną wartością φ^l_j.

Opiszmy odchylenie kwadratowe parametru $\dot{\varphi}_{jb}$:

$$Q\left(\dot{\varphi}_j\right)=\frac{1}{q_b}\sum_{t=1}^{q_b} d^2\left(\dot{R}_{jt}, H_{\dot{\varphi}_{jt}}\right)$$

Ustalmy też zmodyfikowaną metrykę Euklidesa:

$$d\left(\dot{R}_{j1},\dot{R}_{j2}\right)=\sqrt{\sum_{i=1}^{m} d_i\left(\dot{Z}_{i1}-\dot{Z}_{i2}\right)^2+d_{m+j}\left(\dot{Z}_{m+j,1}-\dot{Z}_{m+j,2}\right)^2}$$

ESTYMACJA REALIZACJI

Optymalna wartość φ_{jb}^{l} zostanie uzyskana przez minimalizację odchylenia kwadratowego

$$\dot{\varphi}_{jb}^{l} = \frac{1}{q_b} \sum_{t=1}^{q_b} \dot{\varphi}_{jt}$$

Wartość ta jest niezależna od współczynników d.

Z hipotezy, realizacje z klasy należą do tej samej hiperprzestrzeni, więc korzystając z metryki estymujemy realizację

$$\dot{R}_{jt}^{*e} = \left(\dot{Z}_{1t} - \frac{D_t a_{j1}}{d_1}, \ldots, \dot{Z}_{mt} - \frac{D_t a_{jm}}{d_m}, \dot{Z}_{m+j,t} - \frac{D_t}{d_{m+j}} \right)$$

$$D_t = \frac{\left(\dot{Z}_{m+j,t} - \sum_{i=1}^{m} a_{ji} \dot{Z}_{it} - \dot{\varphi}_j^{l} \right)}{\left(\sum_{i=1}^{m} \frac{a_{ji}}{di} + \frac{1}{d_{m+j}} \right)}$$

Estymacja wartości Z

Używamy tu metryki d jako kryterium podobieństwa.

W tym miejscu dokonujemy odwrotnej transformacji L, uzyskując estymowane wartości Z w naszej klasie. Wyglądają one następująco:

$$Z^{*}_{pb} = Z_{pt} \exp\left[\frac{\left(\frac{a_p}{d_p}\right)\left(\dot{\varphi}^{l}_{jt} - \dot{Z}_{m+j,t} + \sum_{i=1}^{m} a_{ji}\dot{Z}_{it}\right)}{\sum_{i=1}^{m} \frac{a_i^2}{d_i} + \frac{1}{d_{m+j}}} \right]$$

Konstrukcja poziomu ufności

Jeśli hipoteza jest prawdziwa, możemy określić funkcję gęstości dla klasy, korzystając z rozkładu trójkątnego. Używamy do tego statystyki

$$V = -2\sum_{b=1}^{w}\sum_{t=1}^{q_b}\ln\left\{\int_{-\infty}^{\varphi_{jk}} f_{jk}(x)dx\right\}$$

Jeśli hipoteza jest prawdziwa, powyższa statystyka ma rozkład χ^2 z $2(q-w_0)$ poziomami wolności. Z miary jakości podziału na klasy otrzymujemy rodzinę zbiorów

$$V_\alpha = \{x \in R : P(V > x) \le \alpha\}, 0 \le \alpha \le 1$$

I wykorzystując statystykę, mamy poziom ufności

$$p(v_{cal}) = \inf\{\alpha : v_{cal} \in V_\alpha\}$$

Określenie miary podobieństwa

Równanie to nazywane jest miarą podobieństwa $\kappa(j)$ rozpatrywanego podziału realizacji R. Jeśli jest on poprawny, wartości fi są stałe dla każdej klasy. Jeśli choćby jedna z nich miałaby nie spełniać tego założenia, hipoteza powinna być odrzucona.

Statystyka $Q = \min\{\kappa(1),\dots,\kappa(r)\}$ może być więc użyta do testowania hipotezy o kompletności. Ogólnie, miara przyjmie postać

$$\kappa(R) = 1 - (1 - q_{cal})^r$$

Kompletność zbioru

Tak określona miara podobieństwa charakteryzuje się następującymi własnościami dotyczącymi weryfikowania hipotezy o kompletności zbioru:

- $0 \leq \kappa(R) \leq 1$
- Dla $\kappa(R) = 0$ hipoteza jest odrzucona, natomiast dla $\kappa(R) = 1$ – zaakceptowana.
- Im większa miara podobieństwa, tym lepiej w sensie spełnienia warunku kompletności zbioru.

Wartości miary możemy zmieniać zmieniając wartości ε_j.