PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 176

Taksonomia 18. Klasyfikacja i analiza danych – teoria i zastosowania 2011

Bartłomiej Jefmański

Uniwersytet Ekonomiczny we Wrocławiu

ZASTOSOWANIE ROZMYTEJ FUNKCJI REGRESJI W OCENIE POZIOMU SATYSFAKCJI KLIENTÓW Z USŁUG

Streszczenie: Model regresji rozmytej, zaproponowany przez Tanakę, a określany mianem posybilistycznej funkcji regresji, jest nieparametryczną metodą użyteczną w estymacji rozmytych zależności między zmiennymi. Ma on na celu minimalizację stopnia rozmycia związku między zmiennymi poprzez rozwiązanie zadania programowania matematycznego. W artykule zaprezentowano możliwości wykorzystania regresji rozmytej w badaniach satysfakcji konsumentów, a w szczególności szacowania ich poziomu satysfakcji. Podstawy teoretyczne proponowanego podejścia zaprezentowano na przykładzie analizy wyników otrzymanych z badania satysfakcji studentów jednej z niepublicznych szkół wyższych.

Słowa kluczowe: rozmyta funkcja regresji, estymacja satysfakcji konsumenta.

1. Wstęp

Badacze zajmujący się problematyką satysfakcji klienta często podejmują w swoich badaniach próbę oszacowania ogólnego poziomu satysfakcji klientów z nabywanych usług. Jedną z częściej stosowanych metod statystycznych w tym zakresie jest analiza regresji. W wypadku badań satysfakcji z usług zmienną zależną jest ogólny poziom satysfakcji z nabywanej usługi, a zmiennymi niezależnymi są oceny zadowolenia konsumentów z poszczególnych atrybutów usługi.

Należy podkreślić, że satysfakcja konsumenta jest zmienną, którą trudno zdefiniować i mierzyć. Postrzega się ją jako stan emocjonalny, na który oprócz jakości atrybutów usługi wpływają również między innymi czynniki subiektywne czy też sytuacyjne. Związek między ogólnym poziomem satysfakcji a oceną poszczególnych atrybutów usługi jest zatem niewyraźny, ogólnikowy, rozmyty. Uwzględnienie tego faktu staje się możliwe dzięki wykorzystaniu w analizie satysfakcji rozmytej modyfikacji modelu regresji. Zasadniczym celem artykułu jest zaprezentowanie nowego podejścia w szacowaniu poziomu satysfakcji z nabytej usługi poprzez zastosowanie rozmytej funkcji regresji.

2. Charakterystyka rozmytej funkcji regresji

Regresja rozmyta to metoda nieparametryczna, użyteczna w estymacji zależności między zmiennymi o ograniczonej liczbie mało precyzyjnych obserwacji oraz gdy interakcje między zmiennymi są niepewne i rozmyte [Wang, Tsaur 2000, s. 355]. Jest rozszerzeniem klasycznej analizy regresji w przypadku, kiedy niektóre elementy modelu wyrażone są w postaci liczb rozmytych [Modarres i in. 2005, s. 977].

Ze względu na sposób estymacji współczynników regresji wyróżnia się w literaturze przedmiotu podział rozmytych funkcji regresji na dwa rodzaje [Romano 2006, s. 34; Taheri 2003, s. 242]:

- Rozmyta posybilistyczna funkcja regresji (*Fuzzy Possibilistic Regression*) – estymacja funkcji regresji poprzez rozwiązanie zagadnienia optymalizacji z zastosowaniem programowania matematycznego. Celem jest minimalizacja rozmycia oszacowanego modelu poprzez minimalizację całkowitej rozpiętości rozmytych współczynników regresji, które mają postać trójkątnych liczb rozmytych.
- Rozmyta funkcja regresji najmniejszych kwadratów (*Fuzzy Least Squares Regression*) – znacznie bardziej zbliżona do klasycznej funkcji regresji. Celem jest zminimalizowanie odległości między rzeczywiście zaobserwowanymi a oszacowanymi na podstawie modelu zmiennymi rozmytymi.

Pierwsza z nich zaproponowana została przez Tanakę i w pierwotnej wersji zakładała:

– rozmyty związek między zmienną zależną a zmiennymi niezależnymi,
– dokładne dane empiryczne (w pierwotnej wersji modelu wymóg ten dotyczył zarówno zmiennej zależnej, jak i zmiennych niezależnych) oraz rozmyte wartości teoretyczne otrzymane na podstawie oszacowanego modelu,
– współczynniki regresji w postaci trójkątnych, symetrycznych liczb rozmytych.

Rozmyta funkcja regresji zaproponowana przez Tanakę dla *m* zmiennych niezależnych i *n* obserwacji ma postać [Tanaka i in. 1988, s. 275-289]:

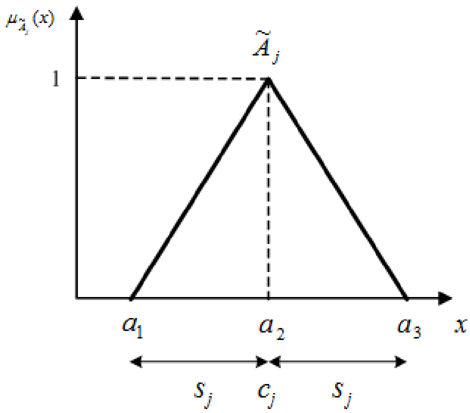
$$\tilde{Y}_i = \tilde{A}_0 + \tilde{A}_1 X_{i1} + \ldots + \tilde{A}_m X_{im} = \tilde{\mathbf{A}}\mathbf{X}_i,\ i = 1,\ 2, \ldots,\ n,\ j = 1,\ 2, \ldots,\ m, \qquad (1)$$

gdzie: $\tilde{Y}_i$ – oszacowana wartość zmiennej zależnej dla *i*-tego obiektu,

$\mathbf{X}_i = [X_{i1}, X_{i2} \ldots, X_{im}]^T$ – wektor wartości zmiennych niezależnych dla *i*-tego obiektu,

$\tilde{\mathbf{A}} = [\tilde{A}_0, \tilde{A}_1, \ldots, \tilde{A}_m]^T$ – wektor oszacowanych rozmytych współczynników regresji.

Ze względu na to, że współczynniki funkcji (1) szacowane są w postaci trójkątnych liczb rozmytych, wartości teoretyczne zmiennej zależnej również mają postać trójkątnych liczb rozmytych. Graficzną interpretację rozmytych współczynników funkcji regresji przedstawiono na rys. 1.



* Na rysunku wyszczególniono dwa rodzaje oznaczeń środka zakresu dziedziny trójkątnej liczby rozmytej stosowane w literaturze przedmiotu.

Rys. 1. Zakresy dziedzin przyporządkowane trójkątnym liczbom rozmytym charakteryzującym poszczególne wartości lingwistyczne

Źródło: opracowanie własne.

Funkcja przynależności dla trójkątnej liczby rozmytej zaprezentowanej na rys. 1 ma następującą postać [Lasek 2002, s. 98]:

$$\mu(x) = \begin{cases} \dfrac{x-a_1}{a_2-a_1}, \ a_1 < x < a_2, \\ \dfrac{a_3-x}{a_2-a_3}, \ a_2 < x < a_3, \\ 0 \text{ dla } x \le a_1 \text{ lub } x \ge a_3, \\ 1 \text{ dla } x = a_2. \end{cases} \tag{2}$$

Wyznaczenie parametrów funkcji regresji wymaga rozwiązania zadania programowania liniowego, polegającego na minimalizacji następującej funkcji celu [Tanaka i in. 1988, s. 275-289]:

$$J = \min\left\{ ns_0 + \sum_{i=1}^{n}\sum_{j=1}^{m} s_j x_{ij} \right\}, \tag{3}$$

przy następujących warunkach ograniczających:

$$c_0 + \sum_{j=1}^{m} c_j x_{ij} + (1-h)\left(s_0 + \sum_{j=1}^{m} s_j |x_{ij}| \right) \geq y_i + (1-h)e_i \quad \forall_i = 1, 2, ..., n,$$
$$c_0 + \sum_{j=1}^{m} c_j x_{ij} - (1-h)\left(s_0 + \sum_{j=1}^{m} s_j |x_{ij}| \right) \leq y_i + (1-h)e_i \quad \forall_i = 1, 2, ..., n, \quad (4)$$
$$s_j \geq 0,\ c_j \in R,\ x_{i0} = 1,\ 0 \leq h \leq 1,\ \forall_j = 1, 2, ..., m,\ i = 1, 2, ..., n.$$

Symbol e_i oznacza połowę zakresu dziedziny i-tej obserwacji zmiennej zależnej w postaci trójkątnej liczby rozmytej. Oszacowanie rozmytych współczynników regresji wymaga również ustalenia parametru *h*, którego wartości mogą mieścić się w przedziale $h \in <0;1>$. Wartość parametru ustalana jest arbitralnie przez badacza w zależności od jego pewności co do rzetelności danych podlegających analizie. Im większa pewność, tym większa wartość parametru *h* [Chen 2001, s. 261-262].

Do oceny stopnia dopasowania oszacowanego modelu do danych empirycznych proponuje się w literaturze przedmiotu wykorzystanie dwóch statystyk [Taheri i in. 2006, s. 4]:

– statystyka MSE (*Mean Squared Error*):

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left[Y_i - \mathrm{def}\left(\tilde{Y}_i\right) \right]^2, \quad (5)$$

gdzie $\mathrm{def}\left(\tilde{Y}_i\right)$ – wyostrzona wartość oszacowanej zmiennej zależnej dla *i*-tej obserwacji,

– statystyka MPC (Mean of Predictive Capabilities):

$$MPC = \frac{1}{n} \sum_{i=1}^{n} \tilde{Y}_i(Y_i), \quad (6)$$

gdzie $\tilde{Y}_i(Y_i)$ – wartość stopnia przynależności empirycznej wartości zmiennej zależnej dla *i*-tej obserwacji w oszacowanym przedziale wartości teoretycznej zmiennej zależnej.

3. Przykład empiryczny – ocena poziomu satysfakcji studentów

W czerwcu 2010 r. wykonano badanie satysfakcji studentów drugiego roku studiów niestacjonarnych jednej ze niepublicznych szkół wyższych działających na terenie powiatu świdnickiego. Przeprowadzono wywiady z 10 spośród 25 wszystkich słuchaczy tego roku studiów. Studenci ocenili ogólny poziom satysfakcji na porządkowej, pięciostopniowej skali pomiaru o następujących wartościach lingwistycznych: 1. „bardzo niezadowolony(a)”, 2. „niezadowolony(a)”, 3. „trudno powiedzieć”, 4. „zadowolony(a)”, 5. „bardzo zadowolony(a)”. Pomiaru oceny atrybutów usługi dydaktycznej, w tym fachowości i uprzejmość pracowników dziekanatu i sekretariatu, dokonano na skali pozycyjnej o wartościach z przedziału <0; 100>, gdzie wartości skrajne oznaczają: 0 – „bardzo źle” a 100 – „bardzo dobrze” .

Celem niniejszej analizy była konstrukcja modelu umożliwiającego szacowanie ogólnego poziomu satysfakcji studentów (zmienna zależna Y) na podstawie oceny fachowości i uprzejmość pracowników dziekanatu i sekretariatu (zmienna niezależna X). Wartości zmiennych dla poszczególnych respondentów zestawiono w tab. 1.

Tabela 1. Zbiór danych pierwotnych

Lp.	*X*	*Y*	Lp.	*X*	*Y*
1	30	2	6	45	3
2	85	4	7	60	4
3	30	2	8	50	4
4	70	5	9	10	1
5	80	5	10	30	3

Źródło: opracowanie własne.

Ze względu na małą liczbę zgromadzonych w badanu obserwacji oraz istnienie w literaturze przedmiotu poglądu, że związki między percepcją jakości atrybutów usługi a ogólnym poziomem satysfakcji konsumentów są niewyraźne i niepewne, zastosowano rozmytą funkcję regresji w wersji zaproponowanej przez Tanakę.

Za pomocą programowania liniowego (zob. wzór 4) oszacowano rozmytą funkcję regresji o następującej postaci:

$$\tilde{Y}_i = [1{,}05;\, 1{,}9] + [0{,}045;\, 0{,}01] X_i\,. \tag{7}$$

Model oszacowano dla wartości parametru $h = 0{,}5$. Wartości statystyk $MSE = 0{,}03$ oraz $MPC = 0{,}79$ świadczą o dobrym dopasowaniu oszacowanego modelu do danych empirycznych.

Wyniki oszacowania ogólnego poziomu satysfakcji studentów z usług administracji uczelni z wykorzystaniem rozmytej funkcji regresji (6) wyszczególniono w tab. 2.

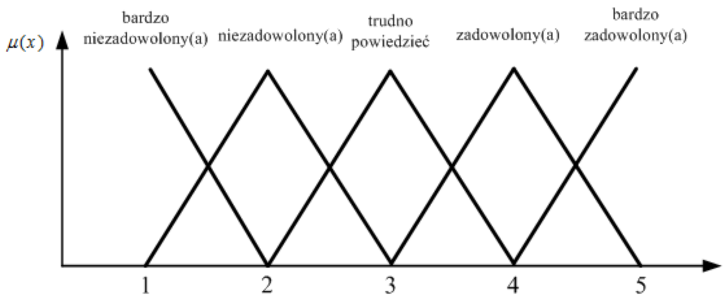
Jedną z własności modelu Tanaki jest to, że oszacowane wartości zmiennej zależnej (w tym przypadku poziom satysfakcji) mają postać liczb rozmytych i trudno o ich bezpośrednie porównanie lub też stwierdzenie, czy poziom satysfakcji jest zadowalający czy też nie. Pewnych wskazówek w tym zakresie mogą dostarczyć jedynie wartości środkowe oszacowanych liczb rozmytych. Dlatego w niniejszym artykule zaproponowano taki sposób transformacji uzyskanych z modelu wyników, aby można było jednoznacznie określić poziom satysfakcji każdego z respondentów. Podejście to wymaga rozmycia wartości empirycznych zmiennej zależnej i wyostrzenia wartości zmiennej zależnej, oszacowanych na podstawie modelu w postaci trójkątnych liczb rozmytych.

Tabela 2. Wartości teoretyczne zmiennej zależnej

Numer respondenta	$\tilde{Y}$			MSE	MPC	Poziom satysfakcji
	środek	zakres	$\mathrm{def}(\tilde{Y})$			
1	2,40	2,2	2,40	0,02	0,82	niezadowolony(a)
2	4,88	2,8	4,88	0,08	0,68	bardzo zadowolony(a)
3	2,40	2,2	2,40	0,02	0,82	niezadowolony(a)
4	4,20	2,6	4,20	0,06	0,69	zadowolony(a)
5	4,65	2,7	4,65	0,01	0,87	bardzo zadowolony(a)
6	3,08	2,4	3,08	0,00	0,97	trudno powiedzieć
7	3,75	2,5	3,75	0,01	0,90	zadowolony(a)
8	3,30	2,4	3,30	0,05	0,71	trudno powiedzieć
9	1,50	2,0	1,50	0,03	0,75	niezadowolony(a)
10	2,40	2,2	2,40	0,04	0,73	niezadowolony(a)

Źródło: obliczenia własne.

Rozmycie porządkowej, pięciostopniowej skali pomiaru ogólnej satysfakcji respondentów wymaga przyporządkowania poszczególnym kategoriom liczb rozmytych. W niniejszym artykule zastosowano trójkątne liczby rozmyte, a sposób ich przyporządkowania wartościom lingwistycznym zaprezentowano na rys. 2.

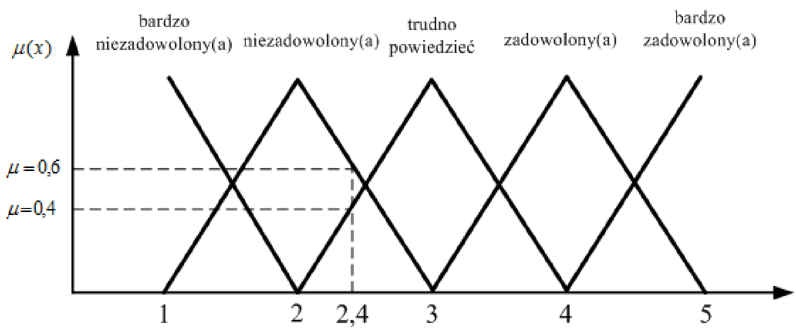


Rys. 2. Zakresy dziedzin przyporządkowane trójkątnym liczbom rozmytym charakteryzującym poszczególne wartości lingwistyczne

Źródło: opracowanie własne.

Drugi krok proponowanego podejścia wymaga wyostrzenia wartości zmiennej zależnej, oszacowanych z modelu w postaci trójkątnych liczb rozmytych. W tym celu zastosowano metodę środków ciężkości, scharakteryzowaną m.in. w opracowaniu Opricovica i Tzenga [2003], a wyniki zestawiono w tab. 2. Wartości wyostrzone oznaczono symbolem $\mathrm{def}(\tilde{Y})$.

Dysponując wyostrzonymi wartościami zmiennych zależnych możemy ocenić ich położenie względem zakresu dziedzin liczb rozmytych zaprezentowanych na rys. 2 i oszacować za pomocą wzoru (2) stopień przynależności do poszczególnych wartości lingwistycznych. Przykład oszacowania poziomu satysfakcji w postaci wartości lingwistycznej dla pierwszego respondenta zaprezentowano na rys. 3.



Rys. 3. Stopień przynależności wyostrzonej wartości oszacowanej na podstawie modelu zmiennej zależnej – przypadek pierwszego respondenta

Źródło: opracowanie własne.

Wyostrzona wartość oszacowanej z modelu (7) zmiennej zależnej dla pierwszego respondenta wyniosła $\operatorname{def}(\tilde{Y}) = 2{,}4$. Jak wynika z rys. 3, ma ona większy stopień przynależności do liczby trójkątnej charakteryzującej wartość lingwistyczną „niezadowolony(a)” niż do liczby trójkątnej opisującej wartość lingwistyczną „trudno powiedzieć”. Należy zatem uznać, że pierwszy z respondentów jest niezadowolony z nabywanej usługi edukacyjnej. W podobny sposób oszacowano poziom satysfakcji w postaci wartości lingwistycznych dla pozostałych respondentów.

4. Podsumowanie

W artykule zaprezentowano badanie satysfakcji, w którym związek między ogólnym poziomem satysfakcji a oceną poszczególnych atrybutów usługi jest niewyraźny, ogólnikowy i niepewny. W takim przypadku zastosowanie znajduje model rozmytej liniowej funkcji regresji, której współczynniki szacuje się za pomocą programowania liniowego. Stosunkowo proste i szybkie obliczenia są główną przyczyną popularności tej metody.

W pierwotnej wersji modelu zaproponowanego przez Tanakę zakresy dziedzin oszacowanych współczynników regresji miały tendencję do przyjmowania wartości 0. Ponadto model jest wrażliwy na występowanie obserwacji odstających, a uwzględnienie w nim większej liczby zmiennych niezależnych może skutkować większym zakresem dziedziny oszacowanej w postaci liczby trójkątnej zmiennej zależnej.

Należy podkreślić, że w literaturze przedmiotu dostępne są również inne podejścia w rozmytej analizie regresji, umożliwiające m.in. oszacowanie modelu w przypadku rozmytych zmiennych zależnych i niezależnych czy też estymację współczynników funkcji regresji w postaci niesymetrycznych, trójkątnych liczb rozmytych.

Literatura

Chen Y.S., *Outliers detection and confidence interval modification in fuzzy regression*, „Fuzzy Sets and Systems" 2001, no. 119.

Lasek M., *Data Mining. Zastosowania w analizach i ocenach klientów bankowych*, Biblioteka Menedżera i Bankowca, Warszawa 2002.

Modarres M., Nasrabadi E., Nasrabadi M.M., *Fuzzy linear regression models with least square errors*, „Applied Mathematics and Computation" 2005, no. 163.

Opricovic S., Tzeng G.H., *Defuzzification within a multicriteria decision model*, „International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems" 2003, no. 5.

Romano R., *Fuzzy Regression and PLS Path Modeling*: *A Combined Two-stage Approach for Multiblock Analysis*, praca doktorska, University of Naples [2006].

Taheri S.M., *Trends in fuzzy statistics*, „Austrian Journal of Statistics" 2003, vol. 32, no. 3.

Taheri S.M., Tavanai H., Nasiri M., *Possibilistic Regression in False-Twist Texturing*, http://www.wseas.us/e-library/conferences/2006elounda1/papers/537-303.pdf.

Tanaka H., Hayashi J., Watada J., *Possibilistic linear regression analysis for fuzzy data*, „European Journal of Operational Research" 1989, no. 40.

Wang H.F., Tsaur R.Ch., *Bicriteria variable selection in a fuzzy regression equation*, „Computers and Mathematics with Applications" 2000, no. 40.

AN APPLICATION OF FUZZY REGRESSION ANALYSIS IN CUSTOMER SATISFACTION ESTIMATION

Summary: Tanaka's model (fuzzy possibilistic regression) approach is a well known fuzzy regression technique used for the prediction problems including fuzzy type of uncertainty. The aim of that model is to minimize fuzziness between the dependent and the independent variables and the solution to this optimization problem is obtained through an extensive use of the mathematical programming. The aim of the paper is to present a potential of a fuzzy linear regression in customer satisfaction surveys particularly in customer satisfaction estimation. An empirical example of student satisfaction estimation was used to present methodological aspects of this approach.