

JEDNOSTAVNA METODA ZA ANALIZU PROMJENA NA JEDNOM ENTITETU OPISANOM NAD SKUPOM KVALITATIVNIH VARIJABLI

Branko Nikolić

Fakultet za defektologiju
Sveučilišta u Zagrebu

Originalni znanstveni članak
UDK:519.8

SAŽETAK

Predložen je model i algoritam za analizu kvalitativnih promjena na jednom entitetu opisanom nad skupom binarnih varijabli u inicijalnom i finalnom stanju. U tu svrhu izračunata je apsolutna i relativna Hammingova udaljenost. Kao mjera kvalitativnih promjena predlaže se koeficijent kvalitativnih promjena (FI) temeljen na Hammingovoj udaljenosti. Značajnost kvalitativnih promjena testirana je HI—kvadrat testom uz broj stupnjeva slobode 1. U tu svrhu izračunata je (P) vjerojatnost pogreške kod odbacivanja nul—hipoteze (H_0). Osim toga izvršena je analiza na kojim varijablama je došlo do kakvih promjena.

0. UVOD

Razvijeno je mnogo metoda i postupaka za analizu kvantitativnih i kvalitativnih promjena na uzorcima entiteta nad kojim je opisan skup kvantitativnih varijabli u više vremenskih točaka, mogu se promatrati tri specijalna slučaja, i to:

1. Analiza promjena na skupu entiteta opisanih nad jednom kvantitativnom varijablom u više distinktnih vremenskih točaka (Momirović i Karaman, Coldiff);

2. Analiza promjena na jednom entitetu nad kojim je opisan skup kvantitativnih varijabli u više vremenskih točaka (Momirović i Karaman, Indiff);

3. Analiza promjena na uzorku entiteta nad kojim je opisan skup varijabli u dvije vremenske točke (Morrison, 1967; Momirović, Gredelj i Sirovitza, 1977, Ssdif i Dif—G).

Kao specijalni slučaj točke 3 javlja se problem analize promjena na jednom entitetu opisanom nad skupom binarnih varijabli u dvije vremenske točke. Taj će problem se pokušati riješiti primjenom teo-

rije informacija, odnosno algoritam za analizu kvalitativnih promjena temeljit će se na Hammingovoj udaljenosti i prijenosu poruka kroz binarni simetrični kanal (Berlekamp, 1968; Peterson, 1961; Wolfowitz, 1964).

1. CILJ

Predložiti će se algoritam i napisati program za kvalitativnu analizu promjena na jednom entitetu, opisanom nad skupom binarnih varijabli u dvije vremenske točke, odnosno u inicijalnom i finalnom stanju. Kao mjera stupnja kvalitativnih promjena bit će definiran koeficijent promjena, izveden pomoću Hammingove udaljenosti (Hamming, 1950), značajnost razlika između dva stanja testirat će se HI—kvadrat testom.

Osim toga bit će provedena analiza kod kojih varijabli su se dogodile, i kakve, promjene.

2. HAMMINGOVA UDALJENOST

Kod definiranja Hammingove udaljenos-

ti (Hamming, 1950) poslužiti ćemo se svojstvima binarnog simetričnog kanala (Peterson, 1967; Aurer, 1972; Pauše, 1974).

Na ulazu binarnog simetričnog kanala (BSC) pojavljuje se poruka kodirana s m -članim binarnim nizom odnosno $x^e \{0,1\}^m$, a na izlazu se pojavljuje ta poruka kodirana kao drugi M -člani binarni niz $\{Y^e 0,1\}^m$

Može se definirati neka funkcija udaljenosti $d(x,y)$ na skupu $0,1^m$ svih m -članih nizova koja predstavlja broj koordinata u kojima se x i y razlikuju.

Vidi se da ta funkcija udaljenosti ima ova svojstva:

1. $d(x,y) \geq 0$
2. $d(x,y) = 0 \iff x = y$
3. $d(x,y) = d(y,x)$
4. $d(x,z) \leq d(x,y) + d(y,z)$

gdje su $x,y,z, e \in \{0,1\}^m$

Funkcija udaljenosti tako definirana naziva se Hammingova udaljenost i izražava broj promjena nastalih prijenosom m -članog binarnog niza kroz binarni simetrični kanal bez memorije.

3. MODEL

Neka je neki entitet opisan vektorom X u nultoj vremenskoj točki (u inicijalnom ispitivanju) nad skupom binarnih varijabli na ovaj način:

$$X = (x_1 \ x_2 \ \dots \ x_m)^T$$

Isti entitet neka je opisan vektorom Y u prvoj vremenskoj točki (u finalnom ispitivanju) nad skupom binarnih varijabli

$$Y = (y_1, y_2, \dots, y_m)^T$$

gdje m predstavlja broj binarnih varijabli.

Definirajmo vektor z kao razliku vektora X i Y

$$z = Y - X$$

Hammingova udaljenost kao apsolutna mjera kvalitativnih promjena između inici-

jalnog i finalnog ispitivanja dobiva se na ovaj način:

$$D = Z^T \cdot Z$$

može se definirati neka relativna mjera kvalitativnih promjena kao:

$$H = \frac{1}{m} \cdot D = \frac{1}{m} \cdot Z^T \cdot Z$$

Za razliku od BSC-a ovdje moramo uočiti dva stanja, i to:

1. prelazak 0 u 1 ($0 \rightarrow 1$)
2. prelazak 1 u 0 ($1 \rightarrow 0$)

Prema tome, Hammingova udaljenost kao mjera kvalitativnih promjena ne može se koristiti u svom izvornom obliku, već je potrebno kreirati jedan novi koeficijent kvalitativnih promjena vodeći računa o smjeru promjena.

Može se predložiti koeficijent kvalitativnih promjena kao:

$$\emptyset = \frac{1}{m} \cdot Z^T \cdot \underline{1}$$

gdje su:

Z – vektor razlika između Y i X

$\underline{1}$ – vektor jedinica dužine m

koeficijent promjena može poprimiti ove vrijednosti

$$-1 \leq \emptyset \leq 1$$

Svojstva \emptyset koeficijenta kvalitativnih promjena

- a) Ako su sve varijable u inicijalnom stanju jednake 0, a sve varijable u finalnom stanju jednake 1, onda je $\emptyset = 1$, pa se može zaključiti da je došlo do "pozitivnih" promjena na svim varijablama nad kojima je opisan neki entitet.
- b) Ako su sve varijable u inicijalnom stanju jednake 1, a sve varijable u finalnom stanju jednake 0, onda je $\emptyset = -1$, pa se može zaključiti da je došlo do "negativnih" promjena na svim varijablama nad kojima je opisan neki entitet.

c) Ako su vektori X i Y jednaki ili ako je broj "pozitivnih" jednak broju "negativnih" promjena, onda je $\emptyset = 0$.

d) Ako vektor X sadrži više varijabli u stanju $\underline{1}$ od vektora Y, tada će \emptyset imati vrijednost u rasponu

$$-1 \leq \emptyset \leq 0$$

e) Ako vektor X sadrži više varijabli u stanju 0 od vektora Y, tada će \emptyset imati vrijednost u rasponu

$$0 < \emptyset < \underline{1}$$

Pomoću koeficijenta kvalitativnih promjena određuje se intenzitet i smjer promjena, ali to se znači da su te promjene statistički značajne.

Moguće je testirati hipotezu da se finalno stanje nekog entiteta ne razlikuje od inicijalnog stanja na skupu kvalitativnih varijabli. Za testiranje te hipoteze predlaže se HI–kvadrat test uz broj stupnjeva slobode jednak $\underline{1}$, a sastoji se u testiranju razlika između dviju koreliranih proporcija (Guilford, J.P., 1968).

HI–kvadrat u ovom slučaju možemo računati na ovaj način:

$$Z = Y - X \quad \chi^2 = (Z^T \cdot \underline{1})^2 \cdot (Z^T \cdot Z)^{-1}$$

odnosno

$$\chi^2 = m^2 \cdot \emptyset^2 \cdot D^{-1}$$

Da bi se mogao primijeniti χ^2 test, Hammingova udaljenost D mora biti veća ili jednaka 10.

Testiranje značajnosti χ^2 svodi se na računanje vjerojatnosti pogreške kod odbacivanja H_0 hipoteze, odnosno izračunvanje P (χ^2/DF) (Veldman, 1967).

Imamo da je

$$F = \chi^2/DF$$

zamijenimo li ∞ s A 1000, normalna aproksimacija Snedecorove F distribucije glasi:

$$T = ((1 - 2/9000)F^{1/3} - (1 - 2/(9 \cdot DF))) A / ((2/9000)F^{2/3} + 2/(9DF) - 1/2)$$

odnosno

$$P(\chi^2/DF) = 0.5 / (1 + C_1 \cdot T + C_2 \cdot T^2 + C_3 \cdot T^3 + C_4 \cdot T^4)$$

gdje su koeficijenti C

$$C_1 = 0.196854$$

$$C_2 = 0.115194$$

$$C_3 = 0.000344$$

$$C_4 = 0.019527$$

Značajnost kvalitativnih promjena testirali smo HI–kvadrat testom uz DF=1, a smjer i intenzitet promjena odredili smo pomoću koeficijenta kvalitativnih promjena \emptyset . Zato, da bi opisali kvalitativne promjene na jednom entitetu u dvije vremenske točke (inicijalno i finalno stanje), potrebno je izračunati oba parametra \emptyset i χ^2 . Pogotovo je to evidentno kad je $\emptyset = 0$. Ako je u tom slučaju $\chi^2 = 0$ onda nije došlo ni do kakvih promjena, međutim ako je $\chi^2 > 0$ broj "pozitivnih" jednak je broju "negativnih" promjena. Taj model podrazumijeva da sve varijable nad kojima je opisan neki entitet u dvije vremenske točke imaju istu težinu, tj. jednako sudjeluju u definiranju neke pojave i pored toga imaju isti "pozitivan" smjer.

4. ALGORITAM I PROGRAM

Program Hamming¹ napisan je tako da korisniku omogućava interaktivan rad, a sve izvršene radnje, izlazni podaci štampaju se u posebnu datoteku. Po završetku rada korisnik može, ako hoće, štampati sve ono

¹Hamming-program, implementiran je u SRCE-u, nalazi se u programskoj datoteci Fakulteta za defektologiju (DEF LIB). Izvođenje ovog programa aktivira se naredbom: DADD DEF LIB. HAMMING

što je radio. Program se sastoji od 5 dijelova, i to:

1. učitavanje i ispis ulaznih podataka
2. izračunavanje Hammingove udaljenosti kao apsolutne i relativne mjere promjena
3. izračunavanje koeficijenta kvalitativnih promjena
4. izračunavanje HI-kvadrata i $P(\chi^2/DF)$ kao vjerojatnosti pogreške kod odbacivanja nul-hipoteze
5. štampanje rezultata analize.

Kao test-podaci poslužilo je 18 binarnih varijabli opisanih nad jednim entitetom o inicijalnom i finalnom stanju. U tablici 1, (prilog) prikazana je kompletna kvalitativna analiza promjena izvedena programom Hamming.

Pregledom tablice 1 može se uočiti da χ^2 iznosi 8,33, a P, kao vjerojatnost pogreške pri odbacivanju H_0 hipoteze, iznosi $P = .023$. Kako je $P < .05$ može se zaključiti da se inicijalno i finalno stanje nekog entiteta međusobno značajno razlikuju na svim binarnim varijablama nad kojima je opisan taj entitet. Relativna Hammingova udaljenost iznosi $H = 0,6667$, a koeficijent kvalitativnih promjena $FI = 0,5557$, pa se može zaključiti da je došlo do pozitivnih promjena na varijablama pod utjecajem određenog tretmana. Pošto je $H \neq FI$ može se zaključiti da se pojavio određeni broj negativnih promjena (1, prešle u 0). Što je broj negativnih promjena veći, to je veća razlika između koeficijenta FI i H. U tablici 1 vrijednost FI

koeficijenta kvalitativnih promjena je visoka, a značajnost promjena testirana je HI-kvadrat testom. FI koeficijenti su međusobno komparibilni. Apsolutna Hammingova udaljenost kao mjera promjena nema osobito značenje jer nije uporediva s drugim mjerama, ali služi za izračunavanje HI-kvadrata i relativne Hammingove udaljenosti.

Analiza promjena ne bi bila potpuna ako se ne uoči na kojim varijablama je došlo, i do kakvih, promjena. Vidi se da je samo na varijabli 2 došlo do "negativne" promjene, dok su 6 i 13 zadržale stanje "0". Očito da upotrijebljeni tretman nije očekivano djelovao na varijable 6 i 13, dok je na varijabli 2 proizveo obrnut efekat od očekivanog.

5. ZAKLJUČAK

Ovaj jednostavni algoritam rješava problem kvalitativne analize kvalitativnih promjena na jednom entitetu opisanom nad skupom binarnih varijabli, gdje sve varijable jednako sudjeluju u definiranju nekog stanja. Koeficijent kvalitativnih promjena predstavlja jednu od boljih mjera pomoću koje se može ocijeniti intenzitet promjena. Značajnost promjena testirana je HI-kvadrat testom. Model i algoritam temelji se na nekim svojstvima binarnog simetričnog kanala, odnosno na Hammingovoj udaljenosti kao mjeri kvalitativnih promjena.

¹OVO JE PROGRAM *HAMMING* KOJI VRŠI ANALIZU KVALITATIVNIH PROMJENA

UPIŠITE BROJ VARIJABLI U FORMATU 15

18

UPIŠITE PODATKE ZA INICIJALNO STANJE U FORMATU 5011

010000010011000001

UPIŠITE PODATKE ZA FINALNO STANJE U FORMATU 5011

101110111111011111

PROGRAM HAMMING SADA IZRAČUNAVA I ISPISUJE SVE PODATKE

APSOLUTNA HAMMINGOVA UDALJENOST	RELATIVNA HAMMINGOVA UDALJENOST
D= 12	H= .66666666

KOEFICIJENT RELATIVNIH PROMJENA	HI-KVADRAT ST. SLOBODE	P(HI/DF)
.555556	8.333	1
		.022563

VARIJABLE KOD KOJIH JE DOŠLO DO POZITIVNIH PROMJENA (0 PREŠLE U 1)

1 3 4 5 7 9 10 14 15 16
17

VARIJABLE KOD KOJIH JE DOŠLO DO NEGATIVNIH PROMJENA (1 PREŠLE U 0)

2

VARIJABLE KOD KOJIH NIJE DOŠLO DO PROMJENA (OSTALE SU JEDNAKE 0)

6 13

VARIJABLE KOD KOJIH NIJE DOŠLO DO PROMJENA (OSTALE SU JEDNAKE 1)

8 11 12 18

LITERATURA

1. AURER, B.: Određivanje optimalnih uvjeta sigurnosnih kodova za otkrivanje i korekciju grešaka kod spremnika podataka, Doktorska disertacija, Elektrotehnički fakultet, Zagreb, 1972, 1–26.
2. BERLEKAMP, E.R.: Algebraic coding theory, McGraw Hill book company, New York, 1968.
3. GUILFORD, J.P.: Osnove psihološke i pedagoške statistike, Savremena administracija, Beograd, 1986, 216.
4. HAMMING, R.W.: Error detecting and error correcting codes, Bell system Techn.J. 29, 1950.
5. MOMIROVIĆ, K., M. GREDELJ i L. SIROVICA: Metode multivarijatne analize, ZPR, Zagreb, 1977.
6. MOMIROVIĆ, K., Ž. KARAMAN: Coldif: Algoritam i program za analizu kvantitativnih promjena spektralnom dekompozicijom univarijatnih trajektorija, Kineziologija, Zagreb, 1981, 11, 1–2.
7. MOMIROVIĆ, K, i Ž. KARAMAN: Indif: Model, algoritam i program za analizu promjena stanja nekog objekta opisanog na skupom kvantitativnih varijabli, Kineziologija, Zagreb, 1981, 11, 1–2.
8. MORRISON, D.F.: Multivariate statistical methods, McGraw–Hill New York, 1961, (133–141).
9. PETERSON, W.W.: Prufbare und korrigierbare Codes. R. Oldenbourg Verlag Munchen–Wien, 1967.
10. PAUŠE, Ž.: Vjerojatnost–informacija–stohastički procesi, Školska knjiga, Zagreb, 1974, 199–215.

A SIMPLE METHOD FOR ANALYSIS OF CHANGES OF THE ONLY ONE ENTITY OVER A GROUP OF QUALITATIVE VARIABLES

summary

The model and algorithm are proposed for the analysis of qualitative changes of the only one entity as described over a group of binary variables in the initial and final position. For this purpose Hamming's distance (absolute and relative) was calculated. As a measure of qualitative changes the coefficient of qualitative changes (FI), which is based on the Hamming's distance, is proposed. Significance of qualitative changes is tested by the Chi-square test with 1 degree of freedom. For this purpose the probability of error (p) in rejecting the null hypothesis is calculated (H_0). Besides, an analysis is done for the obtaining the answer in which variables changes are manifested.