Funkcije za kvantifikovanje uzajamne sličnosti telekomunikacionih servisa

Milan Bjelica, *Student Member, IEEE*

***Sažetak*—Da bi se omogućila personalizacija telekomunikacionih servisa, potrebno je kvantitativno iskazati njihove odlike i uzajamnu sličnost. U ovom radu je dat pregled funkcija koje se tom prilikom koriste.**

***Ključne reči*—Funkcije sličnosti, metričke funkcije, vektorski prostori, personalizacija, telekomunikacioni servisi.**

I. Uvod

JEDAN od izazova koji se postavlja pred operatore multiservisnih telekomunikacionih mreža je modeliranje njihovih korisnika (u ovom radu se za osobu koja koristi telekomunikacioni servis koristi naziv *korisnik*, bez želje da se implicira njena rodna pripadnost). *Model korisnika* pri tome predstavlja skup kvalitativnih i kvantitativnih atributa koji opisuju interesovanja, želje ili potrebe posmatranog korisnika. Kada bi raspolagali ovim podacima, operatori bi mogli svakom od korisnika da ponude *personalizovane* servise, koji bi se u najvećoj meri podudarali s procenjenim modelom.

Dva ključna elementa personalizacije telekomunikacionih servisa su učenje modela korisnika i određivanje sličnosti procenjenog modela s raspoloživim servisima. U svojoj doktorskoj disertaciji [1], pokazao sam da se potonji problem može elegantno rešiti korišćenjem elemenata matematičke teorije metričkih prostora. U ovom radu, sažeto ću predstaviti osnovne rezultate do kojih sam tom prilikom došao.

II. Vektorski prostori

Na samom početku, podsetićemo se modela vektorskog prostora i videćemo kako se on koristi za predstavljanje telekomunikacionih servisa.

A. Osnovne definicije

Neka su dati skup Ω i polje $(\mathbb{F}, +, \cdot)$. Za skup Ω se kaže da predstavlja *vektorski prostor* nad poljem $\mathbb{F}$, ako je zatvoren nad operacijom sabiranja,

$$(\forall \mathbf{x}, \mathbf{y} \in \Omega)\ \mathbf{x} + \mathbf{y} \in \Omega \tag{1}$$

i množenja elementima polja $\mathbb{F}$,

$$(\forall \alpha \in \mathbb{F})\,(\forall \mathbf{x} \in \Omega)\ \alpha\mathbf{x} \in \Omega \tag{2}$$

i ako za sve elemente $\mathbf{x}$, $\mathbf{y}$ i $\mathbf{z} \in \Omega$ i α i $\beta \in \mathbb{F}$ važe sledeće aksiome:

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}, \tag{3}$$

$$\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}, \tag{4}$$

$$(\exists \mathbf{0} \in \Omega)\ \mathbf{0} + \mathbf{x} = \mathbf{x}, \tag{5}$$

$$(\exists (-\mathbf{x}) \in \Omega)\ \mathbf{x} + (-\mathbf{x}) = \mathbf{0}, \tag{6}$$

$$1\mathbf{x} = \mathbf{x}, \tag{7}$$

$$\alpha\,(\beta\mathbf{x}) = (\alpha\beta)\,\mathbf{x}, \tag{8}$$

$$\alpha\,(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}, \tag{9}$$

$$(\alpha + \beta)\,\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}. \tag{10}$$

Elementi vektorskog prostora nazivaju se vektorima, a elementi polja $\mathbb{F}$ skalarima. Aksiome (3)-(6) pokazuju da je Ω komutativna ili Abelova grupa u odnosu na operaciju sabiranja vektora. Ostale aksiome odnose se na množenje vektora skalarom.

Primetimo da postojanje vektorskog prostora implicira linearnost sistema. Bilo koja dva elementa sistema mogu se sabrati i, takođe, bilo koji element sistema može se pomnožiti konstantom, da bi se dobio novi element sistema.

B. Skalarni proizvod vektora

Neka su dati vektorski prostor Ω sa vektorima $\mathbf{x}$, $\mathbf{y}$ i $\mathbf{z}$ i polje $\mathbb{F}$ sa skalarima α i β. Preslikavanje $\pi :\ \Omega \times \Omega \to \mathbb{F}$ za koje važi

$$\pi\,(\mathbf{x} + \mathbf{y}, \mathbf{z}) = \pi\,(\mathbf{x}, \mathbf{z}) + \pi\,(\mathbf{y}, \mathbf{z})\,, \tag{11}$$

$$\pi\,(\alpha\mathbf{x}, \mathbf{y}) = \alpha\pi\,(\mathbf{x}, \mathbf{y})\,, \tag{12}$$

$$\pi\,(\mathbf{x}, \mathbf{y}) = \pi\,(\mathbf{y}, \mathbf{x})\,, \tag{13}$$

$$\mathbf{x} \neq \mathbf{0}\ \Rightarrow\ \pi\,(\mathbf{x}, \mathbf{x}) > 0 \tag{14}$$

naziva se skalarnim proizvodom vektora. U vektorskom prostoru nad poljem realnih brojeva, uobičajeno je da se skalarni proizvod definiše kao

$$\pi\,(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y} = |\mathbf{x}|\,|\mathbf{y}| \cos\theta, \tag{15}$$

gde je sa $||$ označen moduo vektora, dok je θ ugao između vektora $\mathbf{x}$ i $\mathbf{y}$.

Vektorski prostor nad kojim je definisan skalarni proizvod naziva se euklidskim prostorom. Za dva vektora u euklidskom prostoru se kaže da su ortogonalni ako je njihov skalarni proizvod jednak nuli. Ako bazu prostora čine uzajamno normalni vektori jediničnog modula, tada se takav prostor naziva ortonormalnim.

Milan Bjelica, Elektrotehnički fakultet Univerziteta u Beogradu, Bulevar kralja Aleksandra 73, Beograd, Srbija (e-mail: milan@etf.rs).

C. Predstavljanje telekomunikacionih servisa u modelu vektorskog prostora

Po analogiji s teorijom pronalaženja informacija, [2], svakom telekomunikacionom servisu možemo pridružiti njegovu surogatsku apstrakciju u vidu n-dimenzionalnog *vektora odlika* $\mathbf{x} = [x_1\, x_2\, \cdots\, x_n]$. Bazisnim vektorima ovakvog prostora možemo pridružiti značenja „ključnih reči”, pa bi u tom slučaju vrednosti pojedinih koordinata $x_1, x_2, \ldots, x_n$ odgovarale „udelu” ključnih reči u posmatranom servisu; stoga je uobičajeno da vrednosti koordinata budu nenegativne. Primetimo da se, ako su ove vrednosti 0 ili 1, radi o jednostavnom označavanju prisustva ili odsustva ključne reči, dok se u slučaju kada one pripadaju intervalu $[0, 1]$ implicitno uvode elementi fazi skupova.

Treba naglasiti da pitanja definisanja bazisnog skupa za opisivanje telekomunikacionih servisa i određivanja koordinata u ovakvom vektorskom prostoru za sada nisu standardizovana. U oblasti multimedije, moguće rešenje mogli bi predstavljati standardi MPEG-7 i MPEG-21.

Kada su poznati vektorski surogati raspoloživih servisa, moguće je jednostavno kvantifikovati njihovu uzajamnu sličnost.

III. KVANTIFIKOVANJE SLIČNOSTI

Neka su dati vektorski prostor Ω i vektori $\mathbf{x} = [x_1\, x_2\, \cdots\, x_n]$, $\mathbf{y} = [y_1\, y_2\, \cdots\, y_n]$ i $\mathbf{z} = [z_1\, z_2\, \cdots\, z_n]$. U najgrubljoj podeli, funkcije za kvantifikovanje uzajamne sličnosti telekomunikacionih servisa možemo svrstati u dve grupe, mere razdaljine i mere sličnosti.

A. Mere razdaljine

Najpoznatije mere razdaljine su metričke funkcije. Neka je data funkcija f, takva da važi

$$f(\mathbf{x}, \mathbf{y}) \geqslant 0, \tag{16}$$

$$f(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}, \tag{17}$$

$$f(\mathbf{x}, \mathbf{y}) = f(\mathbf{y}, \mathbf{x}), \tag{18}$$

$$f(\mathbf{x}, \mathbf{y}) + f(\mathbf{y}, \mathbf{z}) \geqslant f(\mathbf{x}, \mathbf{z}). \tag{19}$$

Funkcija f naziva se *metrikom* i ima interpretaciju rastojanja između svojih argumenata – što su dva vektora „sličniji”, njihovo rastojanje u posmatranom prostoru je manje i obrnuto.

U nekim primenama su od interesa i funkcije koje ne zadovoljavaju sve uslove (16)-(19). Ako nije ispunjen uslov (17), funkcija f naziva se *pseudometrikom*. Ako ne važe uslovi (17) i (18), radi se o *hemimetrici*. Ako ne važi samo (18), funkcija f je *kvazimetrika*. Za funkciju f kaže se da je *semimetrika* ako ne zadovoljava samo nejednakost trougla, (19). Konačno, ako je zadovoljen samo uslov (16) uz $f(\mathbf{x}, \mathbf{x}) = 0$, funkcija f se naziva *prametrikom*.

Prva mera razdaljine koju ćemo razmotriti je metrika Minkowskog, koja je data izrazom

$$f(\mathbf{x}, \mathbf{y}) = \sqrt[p]{\sum_{i=1}^{n} |x_i - y_i|^p}, \tag{20}$$

pri čemu je parametar p realan broj koji nije manji od 1. Primetimo da su funkcije koje imaju podešljiv parametar naročito pogodne za preciznije kvantifikovanje sličnosti srodnih servisa, jer se promenom vrednosti tog parametra „razvlači skala” metrike.

Metrika Minkowskog ima tri važna posebna slučaja. Za $p = 1$, dobija se Manhattan ili *taxi cab* metrika,

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i|. \tag{21}$$

Za $p = 2$, radi se o Euklidovoj metrici

$$f(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}, \tag{22}$$

koja se, uprkos kritikama da ne opisuje dobro karakteristike multimedijalnih servisa, u literaturi i dalje najčešće koristi.

Kada $p \to \infty$, dobija se metrika supremuma,

$$f(\mathbf{x}, \mathbf{y}) = \max_i |x_i - y_i|. \tag{23}$$

Metrici Minkowskog slična je l_p metrika,

$$f(\mathbf{x}, \mathbf{y}) = \sqrt[p]{\frac{1}{n} \sum_{i=1}^{n} |x_i - y_i|^p}, \tag{24}$$

pri čemu je sada p prirodan broj. Normalizacioni član $1/n$ u literaturi se ponekad izostavlja.

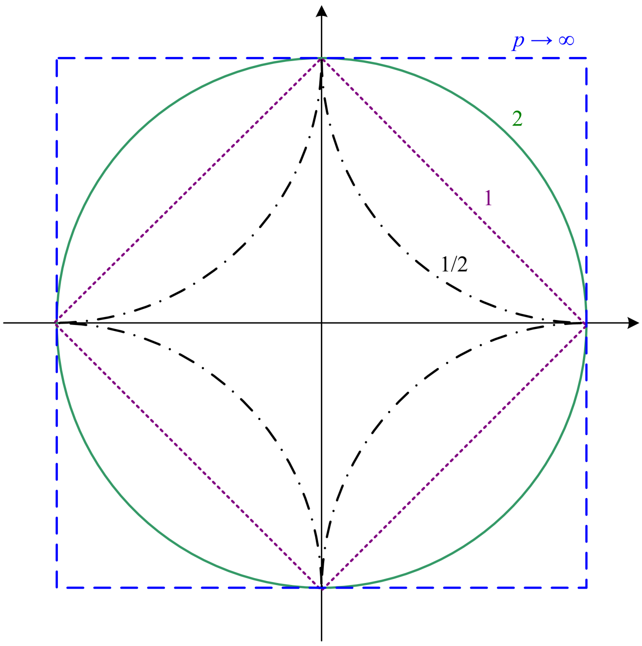
Ponderisana varijanta Euklidove metrike naziva se Mahalanobisovim rastojanjem,

$$f(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} w_i (x_i - y_i)^2}, \tag{25}$$

gde su sa w_i, $i = 1, \ldots, n$ označeni težinski koeficijenti. Izborom njihovih vrednosti se uzima u obzir nejednak uticaj pojedinih koordinata vektorskog prostora na percepciju servisa.

Aggarwal i dr. su primetili da sa povećavanjem dimenzije vektorskog prostora, n, performanse metrike Minkowskog, u smislu sposobnosti distinkcije bliskih vektora, rastu sa smanjivanjm vrednosti parametra p [3]. Zbog toga, predlažu da se u prostorima velikih dimenzija koristi frakciona metrika, koja se dobija kada se u (20) dozvoli da vrednost parametra p pripada intervalu $(0, 1)$. Sledeći ovu ideju, u radu [4], pokazao sam da je uočena pojava još izraženija u hipersfernim vektorskim prostorima, u kojima su moduli svih vektora uzajamno jednaki. Pod takvim okolnostima, optimalna vrednost parametra p, koja pruža najveću distinktivnost metrike, nalazi se upravo unutar intervala $(0, 1)$. François i dr. su pokazali da su frakcione metrike pogodne za primene u uslovima kada šum koji deluje nije beli Gaussov [5].

Na slici 1, prikazane su izodistante linije funkcije (20), za različite vrednosti parametra p. Radi preglednosti, posmatran je dvodimenzionalni prostor, pri čemu je pretpostavljeno da se jedan od vektora nalazi u koordinatnom početku.



Slika 1: Primeri izodistantnih linija funkcije (20), za različite vrednosti parametra p, u dvodimenzionalnom vektorskom prostoru.

Od ostalih mera razdaljine, pomenućemo Gaussovo rastojanje [6],

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \exp\left[-\frac{(x_i - y_i)^2}{2\sigma_i^2}\right], \tag{26}$$

gde su σ_i, $i = 1, \ldots, n$ podešljivi parametri.

Zeta rastojanje [7] dato je izrazom

$$f(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i^2 - y_i^2)^2}. \tag{27}$$

Ova funkcija je primer pseudometrike. Ona postaje metrika onda kada su sve koordinate vektora $\mathbf{x}$ i $\mathbf{y}$ istog znaka.

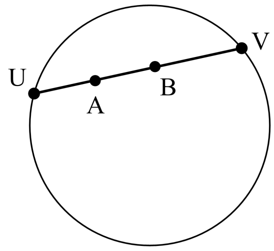
U radu [8], predložena je upotreba hiperboličkog rastojanja,

$$f(\mathbf{x}, \mathbf{y}) = \left|\ln \frac{r - \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}}{r + \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}}\right|, \tag{28}$$

pri čemu je r podešljivi parametar. Geometrijska interpretacija ovog rastojanja u ravni bila bi sledeća. Posmatrani hiperbolički prostor predstavlja kružnu površ. Neka su date tačke A i B koje pripadaju ovom prostoru. Neka je, dalje, kroz ove tačke konstruisana tetiva kruga, čije su krajnje tačke U i V, kao što je to prikazano na slici 2. Hiperboličko rastojanje tačaka A i B tada je

$$f(\mathrm{A}, \mathrm{B}) = \left|\ln \frac{d_E(\mathrm{A}, \mathrm{U})\, d_E(\mathrm{B}, \mathrm{V})}{d_E(\mathrm{A}, \mathrm{V})\, d_E(\mathrm{B}, \mathrm{U})}\right|, \tag{29}$$

gde je sa $||$ označen operator apsolutne vrednosti i sa d_E Euklidovo rastojanje račaka.



Slika 2: Geometrijska interpretacija hiperboličkog rastojanja (28) u ravni.

Još jedan primer neeuklidskog rastojanja nalazimo u [9]. Funkcija f definisana izrazom

$$f(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} g_{i,j}(x_i - y_i)(x_j - y_j)}, \tag{30}$$

gde je $g_{i,j}$ lokalna reprezentacija tenzora Riemannove metrike naziva se tangentnim rastojanjem. Primetimo da se u slučaju $g_{i,j} = \delta_{i,j}$ tangentno rastojanje svodi na Euklidovu metriku.

Ovaj pregled funkcija razdaljine završićemo tzv. merama divergencije, koje se definišu nad prostorima koji su ograničeni na pozitivni hiperkvadrant [10].

Kullback-Leiblerova divergencija data je izrazom

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} x_i \log\frac{x_i}{y_i}. \tag{31}$$

Ova funkcija ima smisla samo ako nijedna koordinata vektora $\mathbf{y}$ nije jednaka nuli. Primetimo da ona nije ni prametrika, jer ne zadovoljava uslov (16); takođe, nije zadovoljen ni uslov simetrije (18). Definiše se i simetrična varijanta Kullback-Leiblerove divergencije, koja ga zadovoljava:

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} x_i \log\frac{x_i}{y_i} + y_i \log\frac{y_i}{x_i}. \tag{32}$$

Jeffreyjeva divergencija data je izrazom

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} x_i \log\frac{x_i}{\xi_i} + y_i \log\frac{y_i}{\xi_i}, \tag{33}$$

gde je $\xi_i = (x_i + y_i)/2$, $i = 1, \ldots, n$.

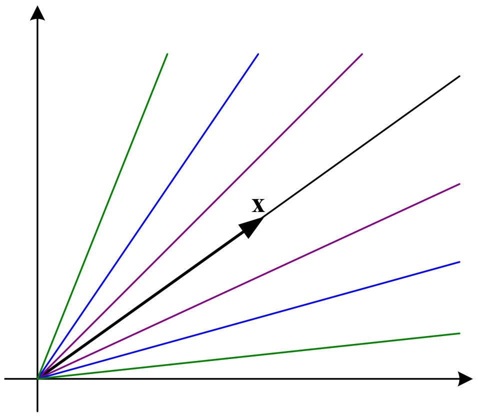
B. *Mere sličnosti*

Funkcije sličnosti se mnogo koriste u teoriji pronalaženja informacija. Nemaju formalnu matematičku definiciju, mada se u mnogim praktičnim primenama insistira da zadovoljavaju sledeće uslove:

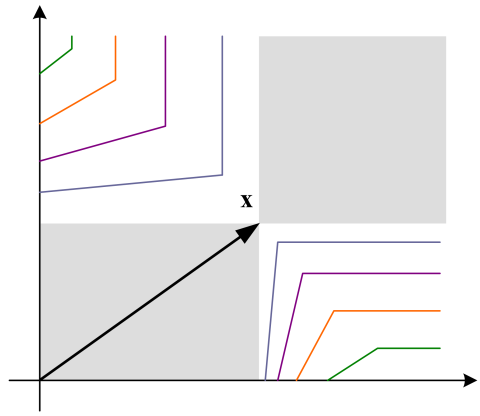
$$0 \leqslant f(\mathbf{x}, \mathbf{y}) \leqslant 1, \tag{34}$$

$$f(\mathbf{x}, \mathbf{y}) = f(\mathbf{y}, \mathbf{x}), \tag{35}$$

$$f(\mathbf{x}, \mathbf{x}) = \kappa = const. \tag{36}$$



Slika 3: Primeri izosimilarnih linija kosinusne sličnosti (37), u dvodimenzionalnom vektorskom prostoru.



Slika 4: Primeri izosimilarnih linija koeficijenta preklapanja (40), u dvodimenzionalnom vektorskom prostoru (prema [11]).

Kosinusna sličnost je neposredno preuzeta iz teorije pronalaženja informacija, gde se i dalje intenzivno koristi. Definisana je izrazom

$$f(\mathbf{x},\mathbf{y}) = \frac{\mathbf{x}\cdot\mathbf{y}}{|\mathbf{x}|\,|\mathbf{y}|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}. \tag{37}$$

Ova funkcija se može shvatiti kao kosinus ugla između posmatranih vektora, kao njihov normalizovani skalarni proizvod, ili kao korelacija. Ako je prostor Ω ograničen na prvi hiperkvadrant, tada kosinusna sličnost zadovoljava uslove (34)-(36), uz $\kappa = 1$.

Na slici 3, prikazane su izosimilarne linije funkcije (37). Ponovo je posmatran dvodimenzionalni prostor, pri čemu je sada pretpostavljen položaj jednog od vektora kao na slici.

Pseudokosinusna sličnost data je izrazom

$$f(\mathbf{x},\mathbf{y}) = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}. \tag{38}$$

Zanimljivo je da u prostoru ove sličnosti postoje vektori koji nisu najsličniji samim sebi.

Dve naredne funkcije spadaju u grupu tzv. mera asocijacije. Diceov koeficijent je dat izrazom

$$f(\mathbf{x},\mathbf{y}) = \frac{2\sum_{i=1}^{n} x_i \cdot y_i}{\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i}. \tag{39}$$

Koeficijent preklapanja, koji je korišten u prvim sistemima za automatsko pronalaženje informacija, definisan je izrazom

$$f(\mathbf{x},\mathbf{y}) = \frac{\sum_{i=1}^{n} \min(x_i, y_i)}{\min\left(\sum_{j=1}^{n} x_j, \sum_{k=1}^{n} y_k\right)}. \tag{40}$$

Na slici 4, prikazane su izosimilarne linije ove funkcije. I ovde je posmatran dvodimenzionalni prostor, pri čemu je pretpostavljeni položaj jednog od vektora vidljiv sa slike.

Ako je data (pseudo)metrička funkcija μ, tada se iz nje može izvesti funkcija sličnosti f koja zadovoljava uslove (34-36) na sledeći način:

(1) Funkcija μ se normalizuje na posmatranom prostoru Ω, tako da važi

$$(\forall \mathbf{x},\mathbf{y} \in \Omega)\;\; \mu_n(\mathbf{x},\mathbf{y}) \in (0,1), \tag{41}$$

(2) Formira se funkcija

$$f(\mathbf{x},\mathbf{y}) = 1 - \mu_n(\mathbf{x},\mathbf{y}), \tag{42}$$

koja predstavlja traženu funkciju sličnosti.

Na ovaj način, polazeći od hiperboličkog rastojanja, (28), može se izvesti hiperbolička sličnost

$$f(\mathbf{x},\mathbf{y}) = \left(\ln\left(e\,\frac{r + \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}}{r - \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}}\right)\right)^{-1}. \tag{43}$$

IV. Diskusija

Korisnicima savremenih multiservisnih telekomunikacionih mreža nudi se veliki broj servisa, koji su različite prirode i karakteristika. Zbog toga, pridruženi vektorski prostori se odlikuju izuzetno velikim dimenzijama. Ako tome dodamo i potrebu da se u realnom vremenu opsluži veliki broj heterogenih korisnika, jasno je zašto je pravilan izbor funkcije za kvantifikovanje sličnosti važan činilac uspeha personalizacije.

U tipičnim scenarijima primene, kvantifikaciona funkcija, bez obzira na to da li je u pitanju mera razdaljine ili mera sličnosti, trebalo bi da bude što manjeg reda računske složenosti. Često nije od interesa znati tačnu vrednost sličnosti, već je bitno uporediti sličnosti nekoliko raspoloživih servisa sa procenjenim modelom korisnika. Stoga je velika prednost ako postoji heuristika koja bi brzo dovela do rešenja, makar ono bilo i suboptimalno. Eventualno postojanje inverzne funkcije svakako je jedna od prednosti, naročito ako je inverzan problem lakše rešiv.

Jednostavnu heuristiku pruža nam geometrijska interpretacija kvantifikacione funkcije, koja može da nas uputi u kojoj oblasti da tražimo servise od interesa za posmatranog korisnika.

Da bi se omogućilo fer poređenje servisa, u radu [12] sam predložio normalizaciju vektorskog prostora, tako da predstavlja hipersfernu površ.

Postojanje *jednog* podešljivog parametra može biti od koristi u primenama sa malim brojem servisa i korisnika, za koje je onda pojedinačno moguće odrediti „optimalne" metrike; to nije racionalno raditi u većim mrežama. Iz istog razloga, nisu podesne ni funkcije sa više od jednog podešljivog parametra, što je diskutovano i u [13].

Imajući u vidu ove napomene, ne iznenađuje što i dalje veliki značaj imaju Euklidova metrika i kosinusna sličnost.

V. Zaključak

U ovom radu su prikazane funkcije koje se u procesu personalizacije telekomunikacionih servisa koriste za kvantifikovanje njihove sličnosti sa procenjenim modelom korisnika. Još uvek se ne može reći da postoji funkcija koja bi imala najbolje performanse u svim praktičnim primenama. Paralelno sa konstruisanjem novih funkcija, koje bi bile pogodne za primene u vektorskim prostorima ogromnih dimenzija, svojstvenim za multiservisna okruženja, stoga će se i dalje koristiti „stare", proverene i dobro poznate funkcije, poput metrike Minkowskog ili kosinusne sličnosti.

Literatura

[1] M. Bjelica, "Algoritam za personalizaciju telekomunikacionih servisa," doktorska disertacija, Elektrotehnički fakultet, Beograd, 2009.

[2] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Commun. ACM*, vol. 18, no. 11, pp. 613–620, 1975.

[3] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional spaces," in *ICDT '01: Proceedings of the 8th International Conference on Database Theory*. London, UK: Springer-Verlag, 2001, pp. 420–434.

[4] M. Bjelica and Z. Petrovic, "Choosing between multimedia services by means of Minkowski metric," in *The International Conference on Computer as a Tool, EUROCON 2005*, vol. 2, Nov. 2005, pp. 1324–1327.

[5] D. François, V. Wertz, and M. Verleysen, "Non-Euclidean metrics for similarity search in noisy datasets," in *ESANN2005, European Symposium on Artificial Neural Networks*, Bruges (Belgium), 2005, pp. 339–344.

[6] P. Muneesawang and L. Guan, *Multimedia database retrieval: A human-centered approach*. Springer Science, 2006.

[7] S. Cheung and A. Zakhor, "Fast similarity search and clustering of video sequences on the world-wide-web," *IEEE Transactions on Multimedia*, vol. 7, no. 3, pp. 524–537, 2005.

[8] J. Góth and A. Skrop, "Varying retrieval categoricity using hyperbolic geometry," *Information Retrieval*, vol. 8, no. 2, pp. 265–283, April 2005.

[9] Z. W. Wang, R. B. Maguire, and Y. Y. Yao, "A non-Euclidean model for web retrieval," in *Web-Age Information Management*, ser. Lecture Notes in Computer Science, H. Lu and A. Zhou, Eds., vol. 1846. Springer, 2000, pp. 233–238.

[10] M. Koskela, A. F. Smeaton, and J. Laaksonen, "Measuring concept similarities in multimedia ontologies: Analysis and evaluations," *IEEE Transactions on Multimedia*, vol. 9, no. 5, pp. 912–922, 2007.

[11] W. P. Jones and G. W. Furnas, "Pictures of relevance: A geometric analysis of similarity measures," *J. Am. Soc. Inf. Sci.*, vol. 38, no. 6, pp. 420–442, 1987.

[12] M. Bjelica and Z. Petrovic, "A novel service retrieval scheme," *Communications Letters, IEEE*, vol. 11, no. 7, pp. 637–639, July 2007.

[13] E. Balmashnova and L. M. Florack, "Novel similarity measures for differential invariant descriptors for generic object retrieval," *J. Math. Imaging Vis.*, vol. 31, no. 2–3, pp. 121–132, 2008.

Abstract

To deliver a set of personalized telecommunication services to the end user, it is necessary to quantify their features as well as to measure the mutual similarities. This paper gives an overview of some mathematical functions that are used for this.

QUANTIFYING THE MUTUAL SIMILARITIES OF TELECOMMUNICATION SERVICES

Milan Bjelica