УДК 515.124.2, 512.652
MSC 2020: 54E35, 06B99
DOI: 10.31471/2304-7399-2024-19(73)-67-73

ГРАТКА ЧАСТКОВИХ УЛЬТРАПСЕВДОМЕТРИК

І.В. Никифорчин, О.Р. Никифорчин, В.М. Пенгрин

Прикарпатський національний університет імені Василя Стефаника; 76018, Шевченка 57, Івано-Франківськ, Україна;

e-mail: oleh.nykyforchyn@pnu.edu.ua, iryna.nykyforchyn@pnu.edu.ua, volodymyrpenhryn@gmail.com

Запроваджено частковий порядок на множині всіх ультрапсевдометрик, визначених на підмножинах фіксованої множини. Запропоновано операції склейки та стягування для ультрапсевдометрик з різними областями визначення. З їх допомогою доведено, що побудована частково впорядкована множина є граткою.

Ключові слова: *гратка, ультраметрика, псевдометрика, зворотний спектр.*

Вступ

У цій праці ми розглядаємо частково впорядковану множину ультраметрик з областями визначення, що є підмножинами фіксованої множини (часткових ультрапсевдометрик). Ультрапсевдометрики є "гібридом" ультраметрик і псевдометрик, які узагальнюють метрики (відстані). Ультра-(псевдо)метрики природно виникають у задачах ієрархічної (багаторівневої) класифікації, тому знайшли численні застосування у біології, комп'ютерних науках, біології тощо (див. огляд [4] Фіонна Муртага, який заслужено вважається лідером цього напрямку).

Практичні потреби вимагають розробки ефективних обчислювальних алгоритмів, як точних, так і наближених, для різних пов'язаних з ультра(псевдо)метриками цілей – ієрархічної кластеризації [5, 6], наближення ультраметриками інших функцій, що описують (не-)подібність об'єктів [9], а також для побудови і використання ультраметричних (еволюційних) дерев, що є зручним засобом подання ультраметрик [1].

Якщо ультраметрика на деякій множині описує багаторівневу класифікацію її елементів, то часткова ультраметрика відповідає класифікації частини елементів. Природно порівнювати часткові класифікації за шириною охоплення і грубістю/детальністю, а також об'єднувати класифікації на множинах у класифікацію на їх об'єднанні. Метою цієї статті є формальне обґрунтування коректності потрібних для цього операцій, зокрема, граткових операцій інфімуму і супремуму двох ультрапсевдометрик.

Нагадаємо, що ультрапсевдометрикою на множині X ми називаємо функцію $d : X \times X \to \mathbb{R}$, таку, що для довільних $x, y, z \in X$ виконано

- $d(x, y) \geqslant 0$ (невід'ємність);
- $d(x, y) = d(y, x)$ (симетричність);
- $d(x, z) \leqslant \max\{d(x, y), d(y, z)\}$ (посилена нерівність трикутника);
- $d(x, x) = 0$.

Якщо ж остання вимога виконана у сильнішій формі

- $d(x, y) = 0$, якщо і тільки якщо $x = y$ (невиродженість),

то функцію d називаємо ультраметрикою [4].

1. Склеювання ультрапсевдометрик

Розглянемо множини X_1 та X_2 з непорожнім перетином X_0. Для спрощення їх можна уявляти скінченними, тоді точні нижні грані нижче будуть мінімумами, хоча істинність тверджень цього розділу не залежить від скінченності.

Припустимо, що на X_1 та X_2 зафіксовано відповідно псевдометрики d_1 та d_2, що збігаються на X_0. Нашою метою ж продовження d_1 та d_2 до деякої ультрапсевдометрики d на об'єднанні $X = X_1 \cup X_2$. Потрібно обрати значення $d(x, z) = d(z, x)$ для всіх $x \in X_1 \setminus X_2$ та $z \in X_2 \setminus X_1$. Для кожного $y \in X_0$ воно повинне задовольняти посилену нерівність трикутника $d(x, z) \leqslant \max\{d(x, y), d(y, z)\} = \max\{d_1(x, y), d_2(y, z)\}$, звідки

$$d(x, z) \leqslant \inf_{y \in X_0} \max\{d_1(x, y), d_2(y, z)\}.$$

Теорема 1. *Якщо ультрапсевдометрики $d_1 : X_1 \times X_1 \to \mathbb{R}$ та $d_2 : X_2 \times X_2 \to \mathbb{R}$ збігаються на $X_0 = X_1 \cap X_2 \neq \varnothing$, то функція $d : X \times X \to \mathbb{R}$, де $X = X_1 \cup X_2$, означена як*

$$d(x, z) = \begin{cases} d_1(x, z), & x, z \in X_1, \\ d_2(x, z), & x, z \in X_2, \\ \inf\limits_{y \in X_0} \max\{d_1(x, y), d_2(y, z)\}, & x \in X_1, z \in X_2, \\ \inf\limits_{y \in X_0} \max\{d_2(x, y), d_1(y, z)\}, & x \in X_2, z \in X_1, \end{cases} \qquad x, z \in X,$$

є найбільшою (поточково) ультрапсевдометрикою, що продовжує d_1 і d_2 на X.

Зауважимо, що умови у формулі вище можуть виконуватись одночасно, але неважко переконатися, що у випадку, коли x, z задовольняють більш, ніж одну умову, відповідні їм значення збігаються.

Доведення. Невід'ємність, симетричність функції d та її рівність нулю для кожної пари однакових аргументів очевидні. Вище також показано, що значення кожної ультрапсевдометрики, що продовжує d_1 і d_2 на X, не перевищують відповідних значень d.

Посилену нерівність трикутника потрібно перевірити тільки для трійок точок, які не лежать одночасно у X_1 чи у X_2. Нехай, наприклад, $x \in X_1 \setminus X_2, z, z' \in X_2 \setminus X_1$. Покажемо, що $d(z, z') \leqslant \max\{d(x, z), d(x, z')\}$, від супротивного. Якщо $\max\{d(x, z), d(x, z')\} < d(z, z') = a$, тобто $d(x, z) < a$, $d(x, z') < a$, то існують $y, y' \in X_0$, для яких $\max\{d_1(x, y), d_2(y, z)\} < a$, $\max\{d_1(x, y'), d_2(y', z')\} < a$. Тоді з $d_1(x, y) < a$, $d_1(x, y') < a$ випливає $d_1(y, y') = d_2(y, y') < a$, а тоді з врахуванням $d_2(y, z) < a$, $d_2(y', z') < a$ отримуємо $d_2(z, z') < a$, що суперечить припущенню.

Доведемо неможливість $\max\{d(x, z), d(z, z')\} < d(x, z') = b$. Тоді були б істинними нерівності $d(x, z) < b$, $d(z, z') = d_2(z, z') < b$, і існувала б точка $y \in X_0$, для якої $\max\{d_1(x, y), d_2(y, z)\} < b$. Тоді з $d_2(y, z) < b$, $d_2(z, z') < b$ випливало б $d_2(y, z') < b$, звідки $d(x, z') \leqslant \max\{d_1(x, y), d_2(y, z')\} < b$, тобто теж отримуємо суперечність. Доведення неможливості $\max\{d(x, z'), d(z', z)\} < d(x, z)$ аналогічне. ☐

Зауваження 1.1. Якщо перетин $X_0 = X_1 \cap X_2$ вище порожній, то формула вище задає невід'ємну симетричну функцію d, яка є нульовою для пар однакових аргументів і задовольняє посилену нерівність трикутника, але набуває значення $+\infty$, якщо аргументи лежать у різних з множин X_1, X_2. Якщо розглядати тільки ультрапсевдометрики зі значеннями, не більшими за фіксоване число M, і покласти $\inf \varnothing = M$, то отримаємо функцію d, що є ультрапсевдометрикою у звичайному сенсі.

Назвемо побудовану вище функцію d *склейкою* ультрапсевдометрик d_1 і d_2, що збігаються на перетині $X_0 \times X_0$ їх областей визначення, і позначимо її $d_1 \veebar d_2$.

Зауваження 1.2. Якщо перетин $X_0 = X_1 \cap X_2$ вище є скінченним, а функції d_1 і d_2 є ультраметриками, то їх склейка теж є ультраметрикою. Неважко побудувати приклад з нескінченним X_0, коли d_1 і d_2 є ультраметриками, а їх склейка не є ультраметрикою, тобто порушується невиродженість.

2. Стягування ультрапсевдометрик

Теорема 2. *Нехай ультрапсевдометрику d задано на множині X, а ультрапсевдометрику d_0 – на її підмножині X_0, і $d_0(x, y) \leqslant d(x, y)$ для всіх $x, y \in X_0$. Тоді функція $d' : X \times X \to \mathbb{R}$, задана як*

$$d'(x, y) = \min\Big\{d(x, y), \inf_{u,v \in X_0} \max\{d(x, u), d_0(u, v), d(v, y)\}\Big\}$$

для $x, y \in X$, є найбільшою серед ультрапсевдометрик на X, які поточково не перевищують d, а їх звуження на X_0 не перевищують d_0.

Доведення. Якщо ультрапсевдометрика ρ на X поточково не перевищує d, а її звуження на X_0 не перевищує d_0, то з посиленої трикутника випливає, що $\rho(x, y) \leqslant d'(x, y)$ для всіх $x, y \in X$. Очевидно, що $d'(x, y) = d_0(x, y)$ для довільних $x, y \in X_0$ і $d' \leqslant d$. Залишається довести, що d' є ультрапсевдометрикою.

Невід'ємність, симетричність і рівність d' нулю при збігу аргументів очевидні. Потрібно довести тільки виконання посиленої нерівності трикутника. Припустимо, що $a = d'(x, z) > \max\{d'(x, y), d'(y, z)\}$ для деяких $x, y, z \in X$. Якби виконувались рівності $d'(x, y) = d(x, y)$, $d'(y, z) = d(y, z)$, то було б

$$d'(x, z) \leqslant d(x, z) \leqslant \max\{d(x, y), d(y, z)\} = \max\{d'(x, y), d'(y, z)\},$$

тому істинне принаймні одне з тверджень

$$d'(x, y) = \inf_{u,v \in X_0} \max\{d(x, u), d_0(u, v), d(v, y)\} < d(x, y)$$

та

$$d'(y, z) = \inf_{s,t \in X_0} \max\{d(y, s), d_0(s, t), d(t, z)\} < d(y, z).$$

Якщо істинне тільки одне з них, наприклад, перше, то маємо

$$\inf_{u,v \in X_0} \max\{d(x, u), d_0(u, v), d(v, y)\} < a, \ d(y, z) < a,$$

тому існують $u, v \in X_0$, для яких $d(x, u) < a$, $d_0(u, v) < a$, $d(v, y)\} < a$. Враховуючи $d(y, z) < a$, маємо $d(v, z) < a$, а тоді

$$d'(x, z) \leqslant \inf_{u,v \in X_0} \max\{d(x, u), d_0(u, v), d(v, z)\} < a,$$

що суперечить припущенню.

Якщо ж істинні обидва твердження, то існують точки $u, v, s, t \in X_0$, для яких $d(x,u) < a$, $d_0(u,v) < a$, $d(v,y)\} < a$ і $d(y,s) < a$, $d_0(s,t) < a$, $d(t,z)\} < a$. Звідси випливає $d_0(v,s) \leqslant d(v,s) \leqslant \max\{d(v,y), d(y,s)\} < a$, а тоді і нерівність $d_0(u,t) \leqslant \max\{d_0(u,v), d_0(v,s), d_0(s,t)\} < a$. Остаточно отримуємо

$$d'(x,z) \leqslant \inf_{u,t \in X_0} \max\{d(x,u), d_0(u,t), d(t,z)\} < a,$$

що теж суперечить припущенню. □

Кажемо, що ультрапсевдометрику d' отримано *стягуванням* d до d_0, і позначаємо її $d \downarrow d_0$.

Зауваження 2.1. Знову ж, якщо множина X_0 є скінченною, і d та d_0 є ультраметриками, то $d \downarrow d_0$ теж є ультраметрикою. Це твердження хибне без припущення про скінченність X_0.

3. Точні нижня і верхня грані двох ультраметрик з різними областями визначення

Загальновідомо (див., наприклад, [7]), що для ультрапсевдометрик d_1, d_2 на множині X їх точна нижня грань (найбільша ультрапсевдометрика, що поточково не перевищує обох d_1 і d_2) визначається формулою

$$d_1 \wedge d_2(x,y) = \inf\bigl\{\max\{d(z_0,z_1), d(z_1,z_2), \ldots, d(z_{n-1},z_n)\} \bigm| n \in \mathbb{N}_+,\ z_0 = x,\ z_n = y,\ z_1, \ldots, z_{n-1} \in X\bigr\}$$

для довільних $x, y \in X$.

Практично інфімум $d_1 \wedge d_2$ знаходиться застосуванням до функції відстані (невід'ємної симетричної функції двох змінних, рівної нулю при збігу аргументів) $\min\{d_1(x,y), d_2(x,y)\}$ алгоритму каркасного дерева (spanning tree algorithm) [8] або алгоритму Флойда-Воршалла, у якому замість звичного додавання вживається максимум.

Розглянемо загальнішу задачу. Позначимо $PUP(X)$ множину всіх ультрапсевдометрик, визначених на підмножинах фіксованої множини X (часткових ультрапсевдометрик). Для $d_1, d_2 \in PUP(X)$, $d_1 : X_1 \times X_1 \to \mathbb{R}$, $d_2 : X_2 \times X_2 \to \mathbb{R}$, вважаємо, що $d_1 \preceq d_2$, якщо $X_1 \supset X_2$ (зауважте напрямок включення!) і $d_1(x,y) \leqslant d_2(x,y)$ для всіх $x, y \in X_2$. Очевидно, що $\preceq$ є частковим порядком на множині $PUP(X)$.

Теорема 3. *Частково впорядкована множина* $(PUP(X), \preceq)$ *є граткою з найбільшим і найменшим елементами.*

Доведення. Найменшим елементом є нульова ультрапсевдометрика на X, а найбільшим – порожня (визначена на порожній множині) ультрапсевдометрика. Точною верхньою гранню ультрапсевдометрик d_1, d_2, де $d_1 : X_1 \times X_1 \to \mathbb{R}$, $d_2 : X_2 \times X_2 \to \mathbb{R}$, у $PUP(X)$ є ультрапсевдометрика на $X_0 = X_1 \cap X_2$, визначена як поточковий максимум d_1 і d_2, тобто $\max\{d_1(x, y), d_2(x, y)\}$ для всіх $x, y \in X_0$.

Нетривіальним є тільки знаходження точної верхньої грані d_1 і d_2. Позначимо d_0 точну нижню грань обмежень d_1 та d_2 на X_0 (її знаходження описано вище). Для довільної ультрапсевдометрики $\rho \in PUP(X)$ з $\rho \preceq d_1$, $\rho \preceq d_2$ випливає $\rho \preceq d_0$, а тоді згідно з Теоремою 2 і $\rho \preceq d_1 \downarrow d_0$ і $\rho \preceq d_2 \downarrow d_0$. Враховуючи, що звуження $d_1 \downarrow d_0$ і $d_2 \downarrow d_0$ на X_0 рівні d_0, на підставі Теореми 1 маємо $\rho \preceq (d_1 \downarrow d_0) \veebar (d_2 \downarrow d_0)$. З іншого боку, склейка $(d_1 \downarrow d_0) \veebar (d_2 \downarrow d_0)$ у $(PUP(X), \preceq)$ передує d_1 і d_2. Це завершує доведення того, що $(d_1 \downarrow d_0) \veebar (d_2 \downarrow d_0)$ є точною нижньою гранню d_1 і d_2, і відповідно $(PUP(X), \preceq)$ є граткою. □

Висновки і майбутні дослідження

Побудовані операції, зокрема, граткові, є природними при агрегації часткових класифікацій. Звичайно, зроблено тільки перші кроки у цьому напрямку. Властивості побудованої гратки часткових ультрапсевдометрик стануть темою наступної публікації. Зокрема, будуть досліджені відношення апроксимації, описані у [7] для граток ультрапсевдометрик з фіксованою областю визначення. Вартими уваги є і функції, які можна подати у вигляді сум ультрапсевдометрик із застосуванням підходу [3].

Іншим напрямком стане розробка ефективних точних і наближених алгоритмів для запропонованих у даній статті операцій, у першу чергу для подання ультрапсевдометрик у вигляді еволюційних дерев [2].

Розроблені алгоритми будуть відтестовані на задачах ієрархічної кластеризації економічних даних.

Література

1. Bang Ye Wu, Kun-Mao Chao, Chuan Yi Tang. *Approximation and exact algorithms for constructing minimum ultrametric trees from distance matrices* // Journal of Combinatorial Optimization, 1999. – 3. – 199–211.
2. M. Farach, S. Kannant, T. Warnow. *Robust model for finding optimal evolutionary trees: extended abstract* // Algorithmica, 1995. – 13. – 155–179.

3. M.E. Mikhailov. *Decompositions of finite pseudometric spaces* // Mathematical Notes, 1998. – 63(2). – 225–234.
4. F. Murtagh. *On ultrametricity, data coding, and computation* // Journal of Classification, 2004. – 21. – 167-184.
5. F. Murtagh, P. Contreras. *Algorithms for hierarchical clustering: an overview* // Wiley Interdiscip Rev: Data Mining Knowl Discov, 2012. – 2. — 86–97.
6. F. Murtagh, P. Contreras. *Algorithms for hierarchical clustering: an overview II* // WIREs Data Mining Knowl Discov, 2017. – 7. – e1219.
7. S.I. Nykorovych, O.R. Nykyforchyn, A.V. Zagorodnyuk, *Approximation relations on the posets of pseudoultrametrics* // Axioms, 2023. – 12(5). – 438.
8. R. Rammal, J. C. Angles d'Auriac, B. Doucot. *On the degree of ultrametricity* // J. Physique Lett., 1985. – 46. – L-945–L-952.
9. M. Di Summa, D. Pritchard, L. Sanitá. *Finding the closest ultrametric* // Disc. Appl. Math., 2015. – 180. – 70–80.

Стаття надійшла до редакційної колегії 18.10.2024 р.

THE LATTICE OF PARTIAL ULTRAPSEUDOMETRICS

I.V. Nykyforchyn, O.R. Nykyforchyn, V.M. Penhryn

Vasyl' Stefanyk Precarpathian National University;
76018, Shevchenka 57, Ivano-Frankivsk, Ukraine;
e-mail: oleh.nykyforchyn@pnu.edu.ua, iryna.nykyforchyn@pnu.edu.ua, volodymyrpenhryn@gmail.com

A partial order is introduced on the set of all ultrapseudometrics defined in subsets of a fixed set. We propose operations of gluing and contraction for ultrapseudometrics with distinct domains. These operations are used to prove that the obtained poset is a lattice.

Key words: *lattice, ultrametric, pseudometric.*