Партиције Дискретних Скупова Тачака

Јовиша Жунић

Увод. Проблем партиције (поделе) скупа тачака, и генерално скупа објеката, јавља се као фундаментални проблем у многим областима математике (теорији бројева, комбинаторици, дискретној геометрији, Буловим алгебрама, итд.) и рачунарства (рачунарској геометрији, неуралним мрежама, машинском учењу, препознавању облика, анализи слика, и многим другим).

У овом истраживачком поглављу дискутоваћемо партиције дискретних скупова тачака. Тај проблем је математички проблем, у својој суштини, али остварени резултати имају утицај на развој више области рачунарства. У овом кратком поглављу, у више детаља, дискутоваћемо само неке од њих.

Неки од класичних математичких проблема, као што су *Gauss Circle Problem* (Johann Carl Friedrich Gauss, 1777 –1855) или *Dirichlet Divisor Problem,* (Johann Peter Gustav Lejeune Dirichlet, 1805–1859), могу бити схваћени као верзије посматраног проблема на дводимензионој целобројној мрежи (партиције реализоване помоћу кругова и хипербола). Теоретске основе многих алгоритама и процедура у различитим областима рачунанарства потичу из теорије партиција (целобројно програмирање, управљање роботима, клсификовање објеката, теорија кодирања). Овде ћемо посебно поменути и проблем *кодирања партиција*, који може бити схваћен као један од савременијих проблема у теорији партиција, мотивисан потрбама које потичу из теорије рачунарства, рачунарске индустрије, односно потребама за успостављање теоријских оквира тих дисциплина.

Проблем кодирања партиција, може бити схваћен као проблем проналажења погодне ознаке (лабеле) за сваку партицију, из посматраног скупа партиција, посебно. То је важан и применљив пробелм у свим димензијама. Ми ћемо посветити пажњу кодирању партиција, датог дискретног скупа тачака, у произвољно димензионом простору, са неколико површина (односно линија, у дво-димензионом простору). Посветићемо и пажњу партицијама d-димензионе целобројне мреже $\{0,1,\ldots,m\}^d$ и њиховом кодирању помоћу одређеног скупа дискретних момената, могло би се рећи, методом већ интензивно коришћеном у литературиi. Да би се обезбедило ефикасно кодирање, потребно је обезбедити да величина и број коришћених дискретних момената буду бирани да буду што мањи -- другим речима, потребно је балансирати иземеђу величине коришћених дискретних момената и њиховог броја, На тај начин се остварује што мањи меморијски простор, потребан за чување (складиштење) придруженог кода (лабеле) посматране партиције.

Ефикасно кодирање треба да омогући:

- Проналажење брзог одговора на питање: Да ли су две партиције једнаке? Да ли посматрана партиција већ постоји у коришћеној бази података?
- Ефикасно коришћење меморијског простора потребно за смештај кодова, придружених посматраним партицијама (мали број битова, по кодираној партицији, може бити један од захтева, а могући су и неки други критеријуми).

Да би било могуће одговорити колико је ефикасно предложено кодирање, потребно је урадити додатна истраживања. Најчешћи сценарио претпоставља прецизну оцену броја посматраних партиција. Пошто је таква оцена урађена, дужина ефикасног кода (при кодирању где је код константнае дужине придружен свакој партицији) се оцењује као логаритам процењене вредности броја посматране класе партиција. Наравно постоје и други поступци оцене ефикасности предложених поступка кодирања. Рецимо, просечна дужина кода (ако нису предложени кодови исте дужине) може бити коришћена као критеријум успешности, или пак просечна дужина кодова обрачуната уз претходно додавање тежинских коефицијента (предложениих кодова) одређених у зависности од учесталости појављивања придружене партиције у "реалним апликацијама".

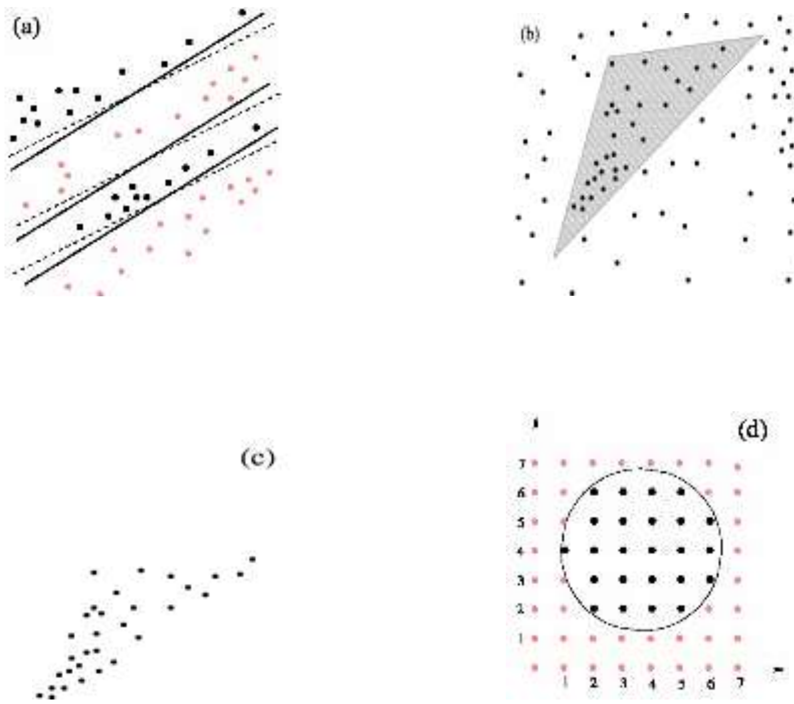
Оцена броја партиција, из посматране класе партиција, је математички проблем разматран од стране великог броја истраживача из различитих области математике (комбинаторике, теорије бројева, статистике, итд.), чак и ако метод за оцену броја партиција не води ка ефикасном поступку за кодирање партиција из посматране класе. Као пример за наведено могу се узети партиције целобројне квадратне мреже $\{0,1,\ldots,m\}^2$ коришћењем конвексних полигона, чија темена су тачке са те мреже. Познато је да је кодирање партиција, из тог скупа, могуће коришћењем $O(m^{2/3})$ битова по свакој кодираној партицији, али је поступак кодирања који би достигао такву успешност и даље непознат. Иако је такав поступак био тема истаживања многих истраживача који се баве дигиталном геометријом такав поступак још увек није пронађен. Напоменимо да постоји хипотеза да бии кодирање уз помоћ такозваних "дигиталних праволинијских сегменеата" достигло оптималну ефикасност, али та хипотеза није није доказана (ни побијена).

Осим теоретске (углавном математичке) оцене ефикасности датог поступка кодирања, прихватљиво је и експериментално оцењивање успешности коришћеног поступка. Познати су резултати, таквих оцењивања, урађени на скуповима *2D* and *3D* бинарних слика (односно партицијама $\{0,1,\ldots,m\}^2$ and $\{0,1,\ldots,m\}^3$), затим на поступцима кодирања "multi-level threshold" функција дефинисаних на бинарном улазу (односно партицијама $\{0,1\}^d$ коришћењем неколико површи), или, још општије, кодирање такозваних "multi-level multi-valued threshold" функција (које кореспондирају са парицијама скупа $\{1,2,\ldots,m\}^d$, при чему и m и d варирају) значајних у теорији неуралних мрежа.

Коначно, поменимо и партиције скупова $\{1,2,\ldots,m\}^d$, при чему су и m и d су јако велики, је од интереса у студијама везаним за хиперспектралне слике. Због величине таквих слика и количине информација које оне носе, свако успешно кодирање (не обавезно и оптимално) било би од велике користи у манипулисању са тако екстремно великим објектима, са којима је тешкп манипулисати због њихове величине. Истраживања на ову тему су савим нова, те је и недостатак значајнијих резултата на ово тему очекиван.

Партиције реализоване помоћу различитог броја површи. Партиције дискретних скупова тачака $S \subset R^d$ са више површи, са поснним освртом на случај $S = \{0,1,\ldots,m\}^d$ су од посеног значаја. Такве париције зовемо *multi-surface partitions*.

Очигледно је да партиције помоцу више површина генерализују париције које могу бити реализована коришћењем само једне површине (такве партиције су већ интензивно студиране у литератури). Постоји велики број верзија овог проблема, које су разматране у различитим областима математике и рачунарства, у зависности од димензионалности простора, кардиналности и структуре скупа на коме се врше партиције, и наравно од врсте и броја површина које су коришћене за поделу посматраног скупа (Слика 1 илуструје неке од већ студираних проблема). Партиције реализоване помоћу једне линије или помоћу једне кружнице (у две димензије), односно са једном хипер-равни или једном хипер-сфером (у више димензија) су највише студиране, али су студиране и партиције помоћу више површина, и то у различитим апликацијама.



Слика 1. (a) *Испрекидане и неиспрекидане линије производе идентичне партиције датог скупа тачака (класа 1: црне тачке; класа 2: црвене тачке). Овај једноставан пример илуструје да једначине коришћених линија не може бити коришћен за кодирање, јер би тада свака таква партиција имала безброј придружених кодова.* **(b)** *Реални троугао дели представљени дискретни скуп тачака на два дела: један део чине тачке унутар троугла и други део који чине тачке изван троугла. Може се рећи да је оваква подела остварена уз помоћ три линије.* **(c)** *Поставља се питање који скуп дискретних момената показује да ли се скуп тачака унутар троугала разликује од презентованог скупа тачака? Да ли такав скуп придружених дискретних момената сугерише да је разлика између поменутих скупова тачака мала?* **(d)** *Реалан круг дели целобројну мрежу на два дела: један део чине тачке унутар круга, док други део чине тачке ван њега. Познато да је скуп тачака унутар круга једини 24-орочлани подскуп од N^2 чији су дикретни моменти првог реда* $\mu_{1,0} = 89$ *и* $\mu_{1,0} = 96$.

На крају овог кратког прегледа, поменимо неке од до сада разматраних проблема:

• У облати препознавања облика (pattern recognition), објекти које треба класификовати су представљени помоћу више димензионалних вектора (тачака). Компоненте (координате) тих вектора су селектоване нумеричке карактеристике коришћене за разликовање објеката из различитих класа. Поделе у тако формираном вишедимензионом простору, уствари, омогућавају жељену класификацију у скупу посматраних објеката. У оваквим случајевима координате вектора, придружених објектима, не морају нужно бити цели бројеви.

• У областима рачунаркот процесирања слика, реални објекти су предтављени подскуповима скупа $\{0,1,\ldots,m\}^2$, при чему свака целобројна тачка представља одговарајући пиксел са посматране дигиталне слике. Број такозваних линеарних дихотомија (подела помоћу једне линије) скупа $\{0,1,\ldots,m\}^2$ је прво студиран, али су партиције уз помоћ кругова и конвексних полигона такође анализиране.

• Број различитих n-точланих скупова $S \subset N^d$ који могу бити раздвојени од њиховог комплемента $N^d \setminus S$ уз помоћ једне хипер равни је тагође разматран – мотивација се јавља у теорији група. Проблем је такође од интереса у теорији линеарног програмирања.

• Геометријска сепарабилност Булових функција је такође анализирана.

• Велика пажња је посвећена поделама скупа $S = \{0,1\}^d$ уз помоћ једне хипер-равни. Мотивација за решавање овог проблема произилази из области теорије неуралних мрежа и анализе линеарних "threshold" уређаја.

• Партиције скупа $\{0,1\}^d$ уз помоћ површи дефинисаних са $\sum_{(p_1,\dots,p_n)\in S} \alpha_{p_1,\dots,p_n} \cdot x_1^{p_1} \cdot \dots \cdot x_n^{p_n} =$ *const.*, где је $\alpha_{p_1,\dots,p_n} \in R$ и где је $S \subset \{0,1\}^d$ дати скуп, су разматране везано за такозване "S-threshold" функције.

• Партиције скупа $\{0,1\}^d$ pomoћу паралелних хипер равни су такође разматране у литератури.

Следећа листа садржи 10 селектованих публикација које садрже ауторске и ко-ауторске резултате аутора овог прегледа, а који се односе на дискутовану истраживачку тему.

Reference

1. Jovisa Zunic, Paul L. Rosin: "Measuring Shapes with Desired Convex Polygons," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 42, no. 6, pp. 1394-1407, 2020.

2. Martin N. Huxley, Jovisa Zunic: "The Number of Different Digital N-discs," *Journal Mathematical Imaging and Vision,* vol. 56, no. 3, pp. 403-408, 2016.

3. Martin N. Huxley, Jovisa Zunic: "The Number of Configurations in Lattice Point Counting II," Proceedings *of the London Mathematical Society,* vol. 107, no. 6, pp. 1331-1352. 2013.

4. Martin N. Huxley, Jovisa Zunic: " The Number of Configurations in Lattice Point Counting II," *Forum Matematicum,* vol. 22, no. 1, 127-152, 2010.

5. Martin N. Huxley, Jovisa Zunic: "The Number of n-Point Digital Discs," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, no. 1, pp. 159-161, 2007.

6. Martin N. Huxley, Jovisa Zunic: "Different Digitisations of Displaced Discs," *Foundations of Computational Mathematics,* vol. 6, no. 2, pp. 255-268, 2006.

7. Jovisa Zunic: "On Encoding and Enumerating Threshold Functions," *IEEE Transactions on Neural Networks,"* vol. 15, no. 2, pp. 261-267, 2004.

8. Jovisa Zunic: "Cutting Corners with Spheres in d-dimensions," *Advances in Applied Mathematics*, vol. 32, no. 3, pp. 609-614, 2004.

9. Jovisa Zunic, Natasa Sladoje: "Efficiency of Characterizing Ellipses and Ellipsoids by Discrete Moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, no. 4, pp. 407-414, 2000.

10. Aleksandar Ivic, Jack Koplowitz, Jovisa Zunic: "On the Number of Digital Convex Polygons Inscribed into an (m,m)-grid," *IEEE Transactions on Information Theory,* vol. 40, no. 5, pp. 1681-1686, 1994.