

# Razvrstavanje objekata s prostornom neodređenosti primijenjeno na modelu skladišta

---

**Kohler, Mirko**

**Doctoral thesis / Disertacija**

**2014**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **Josip Juraj Strossmayer University of Osijek, Faculty of Electrical Engineering / Sveučilište Josipa Jurja Strossmayera u Osijeku, Elektrotehnički fakultet**

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:200:277198>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-11-26**

*Repository / Repozitorij:*

[Faculty of Electrical Engineering, Computer Science and Information Technology Osijek](#)



SVEUČILIŠTE J. J. STROSSMAYERA  
ELEKTROTEHNIČKI FAKULTET OSIJEK

**Mirko Köhler**

**Razvrstavanje objekata s prostornom neodređenosti  
primijenjeno na modelu skladišta**

Doktorska disertacija

Osijek, 2014.

Doktorska disertacija izrađena je na Elektrotehničkom fakultetu Osijek, Sveučilišta J. J. Strossmayera u Osijeku.

**Mentor:** doc.dr.sc. Krešimir Nenadić

Disertacija ima 108 stranica

Disertacija broj: 46

## **Povjerenstvo za ocjenu doktorske disertacije**

1. Prof.dr.sc. Željko Hocenski, redoviti profesor,  
Elektrotehnički fakultet Sveučilišta J. J. Strossmayera u Osijeku
2. Doc.dr.sc. Krešimir Nenadić,  
Elektrotehnički fakultet Sveučilišta J. J. Strossmayera u Osijeku
3. Prof.dr.sc. Mario Žagar, redoviti profesor,  
Fakultet elektronike i računarstva Sveučilišta u Zagrebu

## **Povjerenstvo za obranu doktorske disertacije**

1. Prof.dr.sc. Željko Hocenski, redoviti profesor,  
Elektrotehnički fakultet Sveučilišta J. J. Strossmayera u Osijeku
2. Doc.dr.sc. Krešimir Nenadić,  
Elektrotehnički fakultet Sveučilišta J. J. Strossmayera u Osijeku
3. Prof.dr.sc. Mario Žagar, redoviti profesor,  
Fakultet elektronike i računarstva Sveučilišta u Zagrebu
4. Doc.dr.sc. Damir Blažević,  
Elektrotehnički fakultet Sveučilišta J. J. Strossmayera u Osijeku
5. Doc.dr.sc. Alfonzo Baumgartner,  
Elektrotehnički fakultet Sveučilišta J. J. Strossmayera u Osijeku

Datum obrane disertacije: 13.11.2014.

## **Zahvale**

*Zahvaljujem se mentoru, doc.dr.sc. Krešimiru Nenadiću, na odličnoj suradnji i podršci pri izradi ove doktorske disertacije. Hvala na vremenu, savjetima, konstruktivnim kritikama i svim prijedlozima.*

*Zahvaljujem se prof.dr.sc. Franji Joviću, svom dugogodišnjem voditelju na poslijediplomskom doktorskom studiju, na velikoj pomoći tijekom studija i vjeri da ću uspješno završiti upisani studij.*

*Zahvaljujem se svom nastavnom i nenastavnom osoblju Elektrotehničkog fakulteta na podršci i pomoći koju su mi pružili, ne samo pri izradi disertacije, nego i tijekom rada u nastavi i znanstvenim istraživanjima.*

*Posebno se zahvaljujem prijatelju dr. sc. Ivici Lukiću na svakodnevnoj suradnji u znanstvenom istraživanju i akademskim obvezama.*

*Velika zahvala majci Lidiji i ocu Zlatku na bezuvjetnoj potpori i pomoći koju su mi pružili tijekom cijelog životnog i obrazovnog puta. Također, hvala i ostaloj obitelji na potpori koju su mi pružali tijekom mojeg akademskog obrazovanja.*

*Mojoj supruzi Luciji najveća hvala. Bila mi je motiv i najveća podrška u trenucima kada se problem činio prevelik i nerješiv.*

# Sadržaj

<b>SADRŽAJ</b> .....	<b>5</b>
<b>1. UVOD</b> .....	<b>1</b>
<b>2. ANALIZA I RAZVRSTAVANJE PODATAKA</b> .....	<b>5</b>
<b>2.1. Dubinska analiza podataka</b> .....	<b>5</b>
<b>2.2. Razvrstavanje</b> .....	<b>6</b>
2.2.1. Proces razvrstavanja .....	8
2.2.2. Podjela razvrstavanja po modelima .....	9
2.2.3. Podaci s nedostacima .....	11
<b>2.3. Sličnost i različitost kao mjere pri razvrstavanju objekata</b> .....	<b>11</b>
2.3.1. Matrica blizine .....	12
2.3.2. Graf blizine .....	13
2.3.3. Matrica razasutosti .....	13
2.3.4. Matrica kovarijanci .....	14
2.3.5. Udaljenost kao mjera brojčanih podataka .....	15
2.3.6. Sličnost kao mjera brojčanih podataka .....	18
<b>2.4. Sličnost i različitost kao mjera između grozdova</b> .....	<b>19</b>
<b>2.5. Metode za razvrstavanje objekata oko središnje točke</b> .....	<b>21</b>
2.5.1. <i>k-means</i> metoda.....	21
2.5.2. Kontinuirana <i>k-means</i> metoda .....	24
2.5.3. Usporedna <i>k-means</i> metoda .....	24
2.5.4. Razvrstana <i>k-means</i> metoda .....	25
2.5.5. <i>x-means</i> metoda .....	25
2.5.6. <i>k</i> -harmonijska <i>means</i> metoda .....	26
2.5.7. Ostale metode razvrstavanja oko središnje točke.....	27
<b>2.6. Metode za razvrstavanje objekata prema gustoći</b> .....	<b>28</b>
<b>2.7. Ostali modeli jednoznačnog raščlanjujućeg razvrstavanja objekata</b> .....	<b>29</b>

<b>3. METODE RAZVRSTAVANJA I ODBACIVANJA SREDIŠTA GROZDOVA ZA OBJEKTE S PROSTORNOM NEODREĐENOSTI .....</b>	<b>31</b>
3.1. Razvrstavanje objekata s prostornom neodređenosti pomoću <i>uk-means</i> metode .....	35
3.2. MinMax metoda za odbacivanje središta grozdova .....	37
3.3. Metoda za odbacivanje središta grozdova pomoću Voronojevih dijagrama .....	41
3.4. SPP metoda za odbacivanje središta grozdova .....	46
3.5. Usporedba postojećih metoda .....	50
<b>4. RAZVRSTAVANJE OBJEKATA KOMBINIRANOM METODOM.....</b>	<b>52</b>
4.1. Prilagođavanje postojećih metoda trodimenzionalnom prostoru.....	53
4.1.1. MinMax metoda u $\mathbb{R}^3$ geometrijskom prostoru .....	53
4.1.2. SPP metoda kao temelj za metodu simetralnih ravnina .....	54
4.2. Kombinirana metoda razvrstavanja objekata s trodimenzionalnom prostornom neodređenosti ..	58
4.3. Usporedba i analiza rezultata pokusa .....	59
4.3.1. Analiza rezultata pokusa s osnovnim skupom parametara .....	60
4.3.2. Analiza rezultata pokusa s promjenom broja objekata .....	62
4.3.3. Analiza rezultata pokusa s promjenom maksimalne duljine stranice MBB-a.....	64
4.3.4. Analiza rezultata pokusa s promjenom broja uzoraka u MBB-u .....	67
4.3.5. Analiza rezultata pokusa s promjenom broja grozdova .....	69
4.4. Zaključak o provedenim pokusima .....	73
<b>5. MODEL SKLADIŠTA ZA VERIFIKACIJU RAZVIJENE NOVE METODE RAZVRSTAVANJA .....</b>	<b>74</b>
5.1. Logistika skladišta .....	75
5.1.1. Razne podjele skladišta .....	77
5.1.2. Skladišni prostor .....	79
5.1.3. Pozicioniranje pojedine robe u skladištu .....	80
5.2. O modelu skladišta .....	82
5.2.1. Dimenzije i kapacitet skladišta .....	82
5.2.2. Proizvodi i poslužitelji .....	83

<b>5.3. Verifikacija metoda u modelu skladišta .....</b>	<b>85</b>
5.3.1. Razvrstavanje početnih pozicija poslužitelja .....	85
5.3.2. Proces razvrstavanja pri normalnom opterećenju .....	87
5.3.3. Proces razvrstavanja pri povećanom opterećenju .....	89
<b>6. PRILAGOĐAVANJE MODELA SKLADIŠTA S CILJEM OPTIMIZACIJE POSLUŽIVANJA .....</b>	<b>92</b>
<b>6.1. Prilagođeni model skladišta .....</b>	<b>92</b>
<b>6.2. Proces razvrstavanja u prilagođenom modelu skladišta .....</b>	<b>94</b>
6.2.1. Proces razvrstavanja pri normalnom opterećenju .....	94
6.2.2. Proces razvrstavanja pri povećanom opterećenju .....	96
<b>6.3. Usporedba rezultata klasičnom i prilagođenog modela skladišta .....</b>	<b>97</b>
6.3.1. Usporedba rezultata pri normalnom opterećenju .....	97
6.3.2. Usporedba rezultata pri povećanom opterećenju .....	99
<b>7. ZAKLJUČAK.....</b>	<b>100</b>
<b>8. LITERATURA.....</b>	<b>102</b>
<b>SAŽETAK.....</b>	<b>106</b>
<b>ABSTRACT .....</b>	<b>107</b>
<b>ŽIVOTOPIS .....</b>	<b>108</b>



## Popis oznaka

<i>ED</i>	očekivana udaljenost
<i>MBB</i>	minimalna granična kutija
<i>MBR</i>	minimalni granični pravokutnik
<i>NED</i>	broj računanja očekivane udaljenosti po objektu u svakoj iteraciji
<i>PDF</i>	funkcija gustoće vjerojatnosti
<i>PS</i>	početno stanje
<i>SPP</i>	simetralna metoda razvrstavanja
<i>TED</i>	ukupna očekivana udaljenost

## Popis slika

Slika 2.1: Grozdovi dvodimenzionalnih objekata.

Slika 2.2: Proces razvrstavanja podataka.

Slika 2.3: Modeli razvrstavanja podataka.

Slika 2.4: Raščlanjujući modeli razvrstavanja podataka (a) oko središta (b) prema gustoći.

Slika 2.5: Hijerarhiski modeli razvrstavanja podataka.

Slika 2.6: Skup točaka u dvodimenzionalnom prostoru za različite vrijednosti varijable  $p$ .

Slika 2.7: Udaljenost između najbližih (a) i najudaljenijih (b) objekata dvaju susjednih grozdova.

Slika 3.1: Objekti s prostornom neodređenosti koji su opisani minimalnim graničnim područjem: (a) u 2D prostoru (b) u 3D prostoru.

Slika 3.2: Objekt s dvodimenzionalnom prostornom neodređenosti opisan MBR-om i podijeljen na 400 uzoraka.

Slika 3.3: Minimalna i maksimalna udaljenost između središta grozda  $c_j$  i objekt s prostornom neodređenosti  $o_i$  prikazanog pomoću minimalnog graničnog pravokutnika.

Slika 3.4: Objekt s prostornom neodređenosti  $o_i$  prikazan pomoću modela  $s_a$ : (a) jednom, (b) pet, (c) devet (d) jednom, (e) trinaest i (f) dvadestet jednom sidrišnom točkom.

Slika 3.5: Konstrukcija  $V_{c_p}$  ćelije u  $\mathbb{R}^2$  geometrijskom prostoru.

Slika 3.6: Prikaz odbacivanja središta grozdova pomoću Voronojevih dijagrama 2D prostoru.

Slika 3.7: Prikaz uspješnog odbacivanja središta grozdacj pomoću algoritma Voronojevih dijagrama za (a) prvi podslučaj i (b) drugi podslučaj kada su  $c_p \neq c_j$ .

Slika 3.8: Podjela područja neodređenosti na dva dijela.

Slika 3.9: Podjela  $\mathbb{R}^2$  prostora simetralama između središta grozdova  $c_p$  i  $c_q$  gdje se objekt  $o_i$  nalazi (a) iznad i (b) ispod simetrale.

Slika 4.1: Voronojeve ćelije u 3D prostoru s (a) tri i (b) deset središta grozdova.

Slika 4.2: Minimalna i maksimalna udaljenost između središta grozda  $c_j$  i objekt s prostornom neodređenosti  $o_i$  prikazanog pomoću minimalne granične kutije.

Slika 4.3: Prikaz  $\mathbb{R}^3$  prostora podjeljenog simetralnom ravninom između središta grozdova  $c_p$  i  $c_q$ .

Slika 4.4: Grafički prikaz vremena trajanja razvrstavanja za sve tri metode pri promjeni broja objekata.

Slika 4.5: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode pri promjeni broja objekata.

Slika 4.6: Grafički prikaz omjera vremena izvođenja metode simetralnih ravnina i Kombinirane metode pri promjeni broja objekata.

Slika 4.7: Grafički prikaz vremena trajanja razvrstavanja za sve tri metode pri promjeni maksimalne duljine stranice  $a$  MBB-a.

Slika 4.8: Prikaz promjene NED-a za sve tri metode pri promjeni maksimalne duljine stranice MBB-a.

Slika 4.9: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode pri promjeni maksimalne duljine stranice MBB-a.

Slika 4.10: Grafički prikaz omjera vremena izvođenja metode simetralnih ravnina i Kombinirane metode pri promjeni maksimalne duljine stranice MBB-a.

*Slika 4.11: Grafički prikaz omjera NED-a za MinMax i Kombiniranu metodu pri promjeni maksimalne duljine stranice MBB-a.*

*Slika 4.12: Grafički prikaz vremena trajanja razvrstavanja za sve tri metode pri promjeni broja uzoraka s koji se nalaze unutar MBB-a.*

*Slika 4.13: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode pri promjeni broja uzoraka unutar MBB-a.*

*Slika 4.14: Grafički prikaz omjera vremena izvođenja metode simetralnih ravnina i Kombinirane metode pri promjeni broja uzoraka unutar MBB-a.*

*Slika 4.15: Grafički prikaz vremena trajanja razvrstavanja za sve tri metode pri promjeni broja grozdova.*

*Slika 4.16: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode pri promjeni broja grozdova.*

*Slika 4.17: Grafički prikaz omjera vremena izvođenja metode simetralnih ravnina i Kombinirane metode pri promjeni broja grozdova.*

*Slika 4.18: Prikaz promjene NED-a za sve tri metode pri promjeni broja grozdova.*

*Slika 5.1: Tri tipa razmještaja u skladištu.*

*Slika 5.2: Uloga skladišta u logističkom lancu.*

*Slika 5.3: Euro paleta.*

*Slika 5.4: Tlocrtni prikaz i dimenzije skladišta.*

*Slika 5.5: Skica jednog paletnog regala s tri nivoa.*

*Slika 5.6: Primjer palete s 36 i 1728 paketa.*

*Slika 5.7: Početni razmještaj poslužitelja.*

*Slika 5.8: Grafički prikaz vremena trajanja razvrstavanja za PS 2 do 12 poslužitelja.*

*Slika 5.9: Omjer vremena trajanja razvrstavanja za MinMax i Kombiniranu metodu.*

*Slika 5.10: Grafički prikaz NED-a za PS 2 do 12 poslužitelja.*

*Slika 5.11: Razmještaj poslužitelja u skladištu prije prve narudžbe.*

*Slika 5.12: Grafički prikaz vremena trajanja razvrstavanja za MinMax i Kombiniranu metodu pri promjeni poslužitelja za prvi scenarij.*

*Slika 5.13: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode za prvi scenarij.*

*Slika 5.14: Grafički prikaz NED-a za MinMax i Kombiniranu metodu pri promjeni poslužitelja za prvi scenarij.*

*Slika 5.15: Grafički prikaz omjera NED-a MinMax i Kombinirane metode za prvi scenarij.*

*Slika 5.16: Grafički prikaz vremena trajanja razvrstavanja za MinMax i Kombiniranu metodu pri promjeni poslužitelja za drugi scenarij.*

*Slika 5.17: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode za drugi scenarij.*

*Slika 5.18: Grafički prikaz NED-a za MinMax i Kombiniranu metodu pri promjeni poslužitelja za drugi scenarij.*

*Slika 5.19: Grafički prikaz omjera NED-a MinMax i Kombinirane metode za drugi scenarij.*

*Slika 6.1: Skica jednog paletnog regala s tri nivoa odmaknutog od poda na visinu 0,6 m.*

*Slika 6.2: Tlocrtni prikaz moguće prilagodbe skladišta smanjenjem širine prolaza između redova.*

*Slika 6.3: Grafički prikaz vremena trajanja razvrstavanja za MinMax i Kombiniranu metodu pri promjeni poslužitelja za prvi scenarij u prilagođenom modelu skladišta.*

*Slika 6.4: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode za prvi scenarij u prilagođenom modelu skladišta.*

*Slika 6.5: Grafički prikaz NED-a za MinMax i Kombiniranu metodu pri promjeni poslužitelja za prvi scenarij u prilagođenom modelu skladišta.*

*Slika 6.6: Grafički prikaz vremena trajanja razvrstavanja za MinMax i Kombiniranu metodu pri promjeni poslužitelja za drugi scenarij u prilagođenom modelu skladišta.*

*Slika 6.7: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode za drugi scenarij u prilagođenom modelu skladišta.*

*Slika 6.8: Grafički prikaz NED-a za MinMax i Kombiniranu metodu pri promjeni poslužitelja za drugi scenarij u prilagođenom modelu skladišta.*

*Slika 6.9: Grafički prikaz omjera vremena, izvođenja Kombinirane metode za prvi scenarij, klasičnog i prilagođenog modela skladišta.*

*Slika 6.10: Grafički prikaz omjera vremena, izvođenja Kombinirane metode za drugi scenarij, klasičnog i prilagođenog modela skladišta.*

## Popis tablica

*Tablica 2.1: Udaljenosti između objekta  $x = (6,2,3)$  i  $y = (4,2,5)$*

*Tablica 4.1: Vrijednosti osnovnog skupa parametara*

*Tablica 4.2: Vremena trajanja procesa razvrstavanja i NED za pojedinu metodu koristeći vrijednosti osnovnog skupa parametara*

*Tablica 4.3: Vremena trajanja procesa razvrstavanja i NED za pojedinu metodu koristeći sljedeće parametre ( $n=500$ ,  $a=6$ ,  $s=19683$ ,  $k=49$ )*

*Tablica 4.4: Vremena trajanja procesa razvrstavanja i NED za pojedinu metodu koristeći različit broj objekata  $n$  i ostale varijable iz osnovnog skupa parametara*

*Tablica 4.5: Vremena trajanja procesa razvrstavanja i NED za pojedinu metodu koristeći različite maksimalne duljine stranice  $a$  i ostale varijable iz osnovnog skupa parametara*

*Tablica 4.6: Vremena trajanja procesa razvrstavanja i NED za pojedinu metodu koristeći različit broj uzoraka u MBB-u i ostale varijable iz osnovnog skupa parametara*

*Tablica 4.7: Vremena trajanja procesa razvrstavanja i NED za pojedinu metodu koristeći različit broj grozdova  $k$  i ostale varijable iz osnovnog skupa parametara*

*Tablica 6.1: Vremena trajanja procesa razvrstavanja Kombiniranom metodom za prvi scenarij u klasičnom i prilagođenom modelu skladišta*

*Tablica 6.2: Vremena trajanja procesa razvrstavanja Kombiniranom metodom za drugi scenarij u klasičnom i prilagođenom modelu skladišta*

## 1. UVOD

Dubinska analiza podataka (eng. Data Mining) ciljani je postupak analize podataka s različitim perspektivom, njihovo sažimanje u korisne informacije koje se mogu tada koristiti u svrhu povećanja prihoda, smanjenja troškova itd. Pronalaze se ovisnosti među podacima, uči novo znanje iz podataka i pomoću tog novog znanja predviđa se buduće ponašanje sustava, i to sve na osnovi prethodnih podataka. Dubinska analiza podataka radi se u cilju pronalaska korisnih podataka i njihove prezentacije u ljudima pristupačnom obliku. Razvrstavanje (eng. Clustering) je jedan od najvažnijih dijelova dubinske analize podataka.

Podjela podataka u grupe jedan je od fundamentalnih načina razumijevanja i učenja. Postupak podjele skupa podataka u homogene skupove, grozdove (eng. Cluster), na takav način da su unutar istog grozda podaci međusobno sličniji, nego izvan tog grozda naziva se razvrstavanje. Kada su podaci koji se razvrstavaju prostorne koordinate objekta, tada se može govoriti o razvrstavanju objekata. Objekti koji se nalaze u istom grozdu međusobno su sličniji jedni drugima i na isti način različiti od objekata koji se nalaze u ostalim grozdovima. Za opisivanje sličnost ili različitost među objektima upotrebljavaju se mjere sličnosti. Objekti koji čine jedan grozd trebaju imati veliku mjeru sličnosti i razlikovati se od objekata u drugim grozdovima. U ovoj disertaciji objekti su zadani s trodimenzionalnim koordinatama i mjera sličnost među objektima određuje se pomoću funkcije udaljenosti. Objekti su sličniji što im je manja međusobna udaljenost.

Razvrstavanje podataka se upotrebljava u različitim znanstvenim granama: informatika, ekonomija, biologija, kemija, medicina itd. U računarstvu svoju je primjenu pronašlo u strojnom učenju, prepoznavanju uzoraka, analizi slika, lokacijskim uslugama i sl. Razvrstavanje sadrži važan informacijski alat – odabir (eng. Selection). Zbog široke primjene razvrstavanja nastale su brojne metode i algoritmi koji koriste različite modele kreiranja grozdova. Postoje različiti parametri u odabiru. To su, već spomenuta, funkcija udaljenosti, maksimalan broj objekata u grozdu, gustoća objekata u grozdu i sl. Parametri su ovisni o skupu objekata koje se razvrstava i mijenjaju se za svaki novi skup.

Razvrstavanje je iterativni postupak. Postupak se ponavlja dok se ne zadovolji ciljna funkcija ili dok rješenje ne konvergira. Ciljna funkcija najčešće se predstavlja minimalnim pomakom središta grozdova u odnosu na središta iz prethodne iteracije, to jest, to je minimalna udaljenost za koju se središta grozdova mogu pomaknuti u odnosu na središta u prethodnoj iteraciji. S druge strane uvjet konvergiranja može biti prestanak prelaženja

objekata iz jednog grozda u drugi. Proces razvrstavanja završava ispunjenjem ciljne funkcije. Trajanje procesa razvrstavanja proporcionalno je broju iteracija, a dodatan problem može predstavljati i neodređenost objekata.

Neodređenost objekata nastaje zbog ograničenja resursa, mjernih uređaja, baza podataka i sl., a najčešće je posljedica nepreciznosti ili šuma pri mjerenju te zbog neredovitog periodičkog izvještavanja podataka o objektima. U ovoj se disertaciji daje prijedlog nove metode za razvrstavanje objekata koji sadrže neodređenost u svojim prostornim podacima. Kao primjer uzimaju se pokretni objekti koji periodički izvještavaju svoj položaj, a određivanje položaja između dva izvještaja procjenjuje se, s određenom neodređenosti, koristeći pretpostavke o smjeru i brzini kretanja tog objekta. Tada se takvi objekti s neodređenosti ne mogu predstaviti samo sa svojim prostornim koordinatama, kao točka u prostoru, nego se predstavljaju pomoću minimalnog područja neodređenosti. Objekt se mora nalaziti unutar minimalnog područja neodređenosti i vjerojatnost da se objekt nalazi izvan tog područja jednaka je nula. To se područje neodređenosti, u većini slučajeva, predstavlja funkcijom gustoće vjerojatnosti (eng. Probability Density Function - PDF). PDF se može modelirati na više načina, a u ovoj disertaciji predstavlja se pomoću niza uzoraka unutar minimalnog graničnog pravokutnika (eng. Minimum Bounding Rectangle - MBR) ili minimalne granične kutije (eng. Minimum Bounding Box - MBB). Svakome uzorku se dodjeljuje vjerojatnost da se baš u njemu nalazi objekt. Zbroj svih vjerojatnosti uzoraka, unutar područja neodređenosti, mora biti jednak jedan.

Mjera sličnost kod objekata s neodređenosti određuje se pomoću funkcije očekivane udaljenosti (eng. Expected distance - ED). Računanje ove udaljenosti između središta grozda i objekta zadanog s neodređenosti puno je zahtjevnije nego jednostavno računanje udaljenosti kod jednoznačno zadanih objekata. Za računanje očekivane udaljenosti nekog objekta do središta grozda mora se izračunati udaljenost od svakog uzorka tog objekta do središta grozda, tj. računa se integral svih umnožaka vjerojatnosti i udaljenosti za sve uzorke objekta od središta grozda. Kako bi se funkcije gustoće vjerojatnosti što bolje opisala potreban je velik broj uzoraka te je računanje očekivane udaljenosti i po nekoliko tisuća puta duže nego računanje udaljenosti jednoznačno zadanih objekata. Kao posljedica toga, razvrstavanje objekata s neodređenosti je i po nekoliko tisuća puta dugotrajnije od razvrstavanja objekata koji ne sadrže neodređenost. Može se zaključiti da je u procesu razvrstavanja objekata s neodređenosti računanje očekivane udaljenosti vremenski najzahtjevnija računalna operacija. Kako bi se smanjio utjecaj koji ima računanje očekivane udaljenosti na vrijeme izvođenja

razvrstavanja razvijene su različite metode koje odbacuju grozdove kao kandidate za pojedini objekt s neodređenosti bez računanja očekivane udaljenosti.

U ovoj disertaciji objašnjene su trenutno korištene metode za odbacivanje grozdova kao kandidata (eng. Pruning Methods) i predstavljena je nova metoda za razvrstavanje objekata koji sadrže neodređenost u sve tri dimenzije geometrijskog prostora  $\mathbb{R}^3$ . Ta nova metoda koristi prednosti dvije različite metode koje su prilagođene  $\mathbb{R}^3$  prostoru. Uklanjanjem nedostataka pojedine metode, Kombinirana metoda odbacuje veći ili jednak broj grozdova kao i ostale metode. Prednost je Kombinirane metode, u odnosu na ostale metode, da joj je potrebno manje vremena za odbacivanje grozdova kao kandidata. U ovoj su disertaciji provedeni i pokusi koji dokazuju kako Kombinirana metoda odbacuje više grozdova od postojećih metoda ili je sam postupak odbacivanja grozdova brži nego kod postojećih metoda.

Metode razvrstavanja objekata s neodređenosti koje su opisane i predstavljene u ovoj disertaciji primijenjene su na modelu skladišta. Model skladišta je napravljen s ciljem ispitivanja metoda razvrstavanja i uspoređivanje njihovih svojstava. Model skladišta sadrži bazu podataka u koju su spremljeni podaci o proizvodima, narudžbama i zadacima sustava koje se mogu analizirati u svrhu smanjenja vremena potrebnog za skupljanje narudžbe. Na temelju postojećih podataka razvrstavanjem se može dobiti informacija koji je poslužitelj najpogodniji za preuzimanje narudžbe i napraviti predviđanje stanja sustava nakon njezinog izvršenja. Ta predviđanja mogu kao rezultat imati znatnu uštedu sredstva za održavanje sustava i vremena potrebnog za obavljanje narudžbe.

Na kraju ove disertacije dane su preporuke kako prilagoditi skladište u svrhu optimiranja posluživanja. No osnovu tih preporuka napravljen je model skladišta prilagođen novoj metodi razvrstavanja objekata koji sadrže neodređenost u svojim prostornim koordinatama. Cilj ovih prilagodbi je ubrzati sam proces razvrstavanja, tj. dodjeljivanja narudžbi pojedinom poslužitelju kao i omogućavanje što kraćeg komisioniranja narudžbe.

Disertacija je organizirana na sljedeći način. U drugom poglavlju opisani su pojmovi dubinske analize, razvrstavanja, mjera sličnosti i različitosti te dan pregled modela za razvrstavanje objekata. Za svaki model predstavljeni su najpoznatiji algoritmi za razvrstavanje i navedene njihove prednosti i nedostaci.

U trećem poglavlju predstavljene su metode razvrstavanja objekata koji sadrže neodređenost u dvije dimenzije i načini razvrstavanja takvih objekata klasičnim metodama iz drugog poglavlja. Uveden je pojam odbacivanja grozdova i objašnjene su metode koje koriste ovaj postupak za ubrzanje postupka razvrstavanja.



U četvrtom poglavlju dan je pregled prilagođenih metoda za razvrstavanje objekata koji sadrže neodređenost u sve tri dimenzije geometrijskog prostora  $\mathbb{R}^3$  i opisana je metoda razvijena u ovoj doktorskoj disertaciji. Napravljeni su pokusi i analize koje uspoređuju opisane metode za razvrstavanje objekata s neodređenosti pri različitim vrijednostima parametara.

U petom poglavlju objašnjeni su osnovni pojmovi vezani uz logistiku skladišta i dan je opis razvijenog modela skladišta. Napravljeni su pokusi i analiza koja uspoređuju opisane metode za razvrstavanje i novo razvijenu metodu pri odabiru poslužitelja i određivanju optimalne pozicije poslužitelja između narudžbi.

U šestom poglavlju opisani su nedostaci modela skladišta iz prethodnog poglavlja i dane su preporuke kako prilagoditi skladište u svrhu optimiranja posluživanja. Napravljene su izmjene u postojećem modelu i ponovno su napravljeni pokusi i analiza te uspoređeni s ranije dobivenim rezultatima. Na kraju doktorske disertacije izneseni su zaključci i dan je sažetak iste.

## 2. ANALIZA I RAZVRSTAVANJE PODATAKA

Ovo poglavlje predstavlja neke osnovne koncepte. Prvo je dan pregled ciljeva i razloga za provođenje analize i razvrstavanja podataka. Nakon toga, objašnjeni su pojmovi kao što su mjere sličnosti i udaljenosti, najčešći modeli i različite metode razvrstavanja.

### 2.1. Dubinska analiza podataka

Informacijski sustavi, u današnje vrijeme, prikupljaju velike količine raznovrsnih podataka koji dolaze iz različitih izvora, različite su forme, sadržaja i vrijednosti. Skup metoda, tehnika i alata koji pretražuju, analiziraju, klasificiraju, procjenjuju i uspoređuju velike količine podataka nazivaju se dubinska analiza podataka. Dubinska analiza podataka obuhvaća istraživanje i analizu velikih količina podataka u svrhu otkrivanja smislenih uzoraka i pravila. To je analitički proces koji rezultira novim znanjem [1], koje se može prikazati u obliku izvještaja. Dubinska analiza podataka još se naziva i dubinska pretraga podataka, inteligentna analiza te rudarenje podataka.

Još jedna definicija dubinske analize podataka navodi da je to analiza opservacijskih podatkovnih setova s ciljem pronalaženja neočekivanih veza ili prikaza podataka koji su za vlasnika podataka novi i korisni [2]. Pod opservacijskim podacima misli se na one podatke koji nisu prikupljeni s ciljem analiziranja, već podacima koji su prikupljeni praćenjem, recimo transakcija. Zbog toga se dubinska analiza podataka često naziva sekundarnom analizom te se na taj način razlikuje od statistike [3].

Većina tehnika dubinske analize podataka preuzeta je iz statistike i računalnih znanosti u svrhu otkrivanja odnosa među varijablama te kreiranju novog zvanja [1]. Izbor određene tehnike, tj. kombinacije tehnika, koje se primjenjuju u određenoj situaciji ovisi o vrsti problema, dostupnim podacima te iskustvenom odabiru istraživača. Dubinska analiza podataka u velikoj mjeri bavi se izgradnjom modela koji se predstavljaju različitim algoritmima. Većina algoritama za dubinsku analizu podataka zahtijeva velike količine podataka kako bi se napravili i trenirali modeli koji će se zatim koristiti u svrhu klasifikacije, predikcije, procjene, razvrstavanja ili nekog drugog zadatka dubinske analize.

Različite metodologije dubinske analize podataka pokazuju kako je moguće niz aktivnosti, koje radi analitičar, predstaviti kao niz logičnih koraka ili zadataka. Danas u dubinskoj analizi podataka dominiraju dvije metodologije: CRISP-DM i SEMMA. Prema [4] najzastupljenija metodologija za rudarenje podataka je CRISP-DM (eng. Cross Industry

Standard Proces for data mining) metodologija, na drugom mjestu su vlastito razvijene metodologije, a na trećem se nalazi SEMMA (eng. Sample, Explore, Modify, Model, Assess). CRISP-DM, opisan u [5], sastoji se od šest osnovnih koraka: razumijevanje problema, razumijevanje podatka, priprema podatka, modeliranje, evaluacija modela i primjena. Ova metodologija definira dubinsku analizu podataka kao iterativni proces koji prolazi kroz nabrojane korake, kako bi riješio određeni problem. SEEMA-u je razvio SAS instituta i sastoji se od 5 koraka iz naziva: uzrokovanje, istraživanje, modificiranje, modeliranje i procjena [6].

Primjena dubinske analize podataka može se pronaći u mnogim područjima, kao što su: transport, marketing, farmacija, proizvodnja, upravljanje, medicinska informatika, komunikacije, planiranje vojnih operacija itd [7]. Dubinska analiza podataka koristi se za:

- otkrivanje anomalija,
- uočavanje ovisnosti podataka,
- razvrstavanje,
- klasificiranje,
- regresiju,
- predstavljanje podataka, njihovu vizualizaciju i pisanje izvještaja.

U ovoj disertaciji analiziraju se podaci koji su zadani s neodređenosti. Dubinska analiza nad takvim skupom podataka tim je teža jer se osim informacije o samim podacima daje i njihova neodređenost. Obrada velikih skupova podataka u prihvatljivom vremenu vrlo je važna te je osim velike računalne moći potrebno napraviti i metode koje skraćuju vrijeme izvođenja analize.

## 2.2. Razvrstavanje

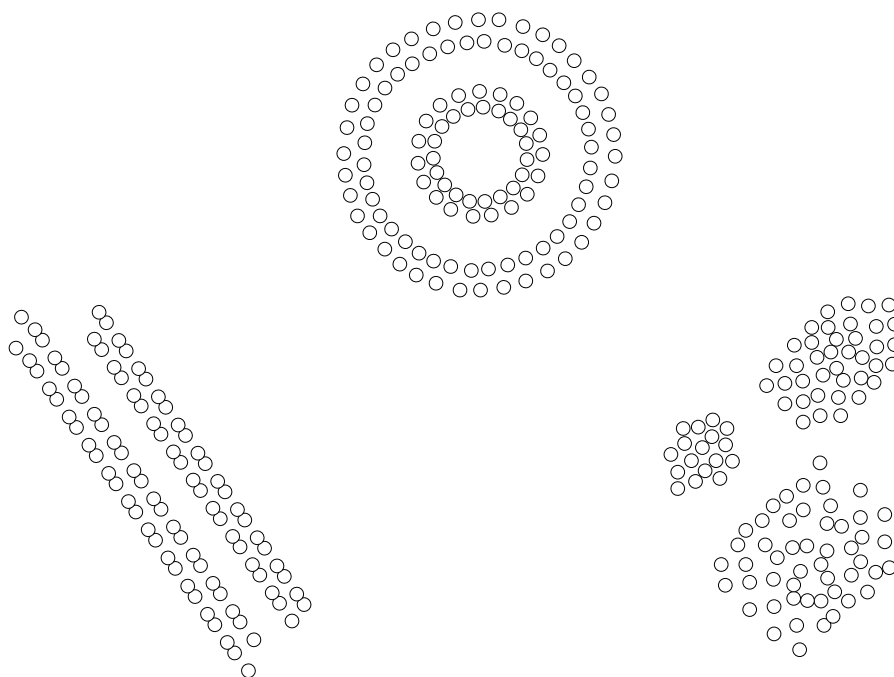
Kako je navedeno u prethodnom potpoglavlju, razvrstavanje podataka jedan je od najvažnijih zadataka dubinske analize. To je postupak podjele koji skup podataka raspodjeljuje u homogene skupove, grozdove, na takav način da su podaci unutar istog grozda međusobno sličniji (na osnovu jednog ili više kriterija), nego podaci izvan tog grozda. Razvrstavanje podataka upotrebljava se u svim granama računarstva (strojnom učenju, prepoznavanju uzoraka, analizi slika, lokacijskim uslugama) kao i u biologiji, medicini, ekonomiji itd. Kada je podatak koji se razvrstava prostorna lokacija nekog objekta, onda se govori o razvrstavanju objekata u grozdove. Objekti koji se nalaze u istom grozdu međusobno

su slični jedni drugima i na isti način različiti od objekata koji se nalaze u svim ostalim grozdovima. Za opisivanje sličnost ili različitost među objektima upotrebljava se mjera sličnosti. Sličnost je predefiniрани kriterij koji se računa iz opažanja nad objektima. Prema [8] mjera sličnosti koristi se kako bi se opisala sličnost između dva objekta na kvantitativan način. Što je mjera sličnost između dva objekta veća ti su objekti sličniji, a vrijedi i obrnuto. Objekti koji čine jedan grozd trebaju imati:

- što veću mjeru sličnosti,
- što manju međusobnu udaljenost i
- što veću razliku od objekata u drugim grozdovima.

Ako ne postoji mjera sličnosti između dva objekta tada nije moguće provesti ispravno razvrstavanje, jer ne postoje kvantitativni podaci prema kojima bi se razvrstavanje napravilo. Postoji više različitih mjera sličnosti koje će biti obrađene u kasnijim pod poglavljima.

Određeni broj sličnih objekata razvrstanih ili grupiranih zajedno čini jedan grozd. U literaturi se može pronaći mnoštvo definicija grozda, no u [9] dana je sljedeća definicija: „Grozdovi se mogu opisati kao povezane regije multi-dimenzionalnog prostora koje sadrže relativno visoku gustoću točaka, odvojene od istih takvih regija s regijom koja sadrži relativno nisku gustoću točaka.“ Ova definicija pretpostavlja da su objekti, koje se razvrstava u grozdove, predstavljeni kao točke koje sadrže koordinate koje su mjerljive. Sa slike 2.1 vidljiva su područja s visokom i niskom gustoćom točaka. Iako postoje različite definicije grozda, teško je definirati što je u nekoj realnoj situaciji zapravo grozd. Razlog tome može se



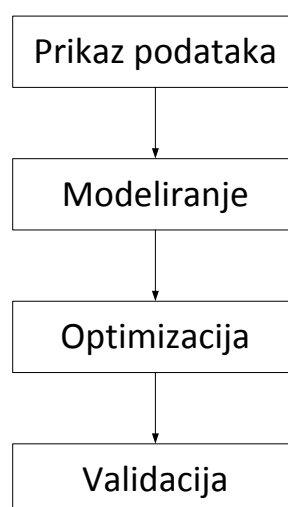
Slika 2.1: Grozdovi dvodimenzionalnih objekata.

pronaći u slici 2.1. Dvodimenzionalni grozdovi prikazani su na spomenutoj slici. Ako se promatra cijeli prostor slike vidljivo je da su objekti podijeljeni u tri grozda, no ako se lokalno promatra prostor može se izbrojati sedam jedinstvenih grozdova.

U [10] grozdove se dijeli u dvije skupine: kompaktni i ulančani grozdovi. Kompaktni grozd je skup objekata koji imaju visoku mjeru sličnosti i ima karakteristiku da se grozd može predstaviti sa središnjom točkom. Na slici 2.1, u donjem desnom kutu, prikazano je nekoliko kompaktnih grozdova. Ulančani grozd skup je objekata koji su međusobno slični i svaki objekt povezan je s njemu susjednim objektima. Znači da mora postojati putanja u grozdu koja međusobno povezuje bilo koja dva objekta unutar grozda. Na slici 2.1 prikazana su četiri ulančana grozda kružnog i ravnog oblika. Vidljivo je da između bilo koja dva objekta unutar jednog grozda postoji putanja koja ih povezuje.

### 2.2.1. Proces razvrstavanja

Proces razvrstavanja prolazi kroz četiri koraka, kao što je prikazano na slici 2.2. Prvi korak je prikaz podataka (eng. Data Representation), u kojoj se određuju moguće strukture koje se mogu pronaći među podacima. U koraku modeliranja se na osnovi prikaza podataka o objektima određuje način i kriteriji prema kojima će se objekti razvrstavati u grozdove, tj. odabire se model razvrstavanja koji će se koristiti za stvaranje grozdova. U trećem koraku vrši se optimizacija korištenog prikaza podataka i odabranog modela kako bi se dobili što bolji rezultati. Kao zadnji korak u procesu razvrstavanja obavlja se validacija rezultata razvrstavanja u smislu provjere dobivenog i očekivanog rezultata.



Slika 2.2: Proces razvrstavanja podataka.

### 2.2.2. Podjela razvrstavanja po modelima

Kada se govori o razvrstavanju, u ovoj disertaciji, misli se na jednoznačno razvrstavanje. U ovom tipu modela razvrstavanja objekt  $o_i \in O = \{o_1, o_2, \dots, o_n\}$  može pripadati samo jednom grozdu iz skupa  $C = \{c_1, c_2, \dots, c_k\}$ . Rezultat ovog tipa razvrstavanja dan je  $k \times n$  matricom, gdje je  $k$  broj grozdova, a  $n$  broj objekata, kako je prikazano jednadžbom (2.1).

$$O = \begin{pmatrix} o_{11} & o_{12} & \dots & o_{1n} \\ o_{21} & o_{22} & \dots & o_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ o_{k1} & o_{k2} & \dots & o_{kn} \end{pmatrix} \quad (2.1)$$

Vrijednosti za svaki pojedinačni  $o_{ji}$  zadovoljavaju jednadžbu (2.2).

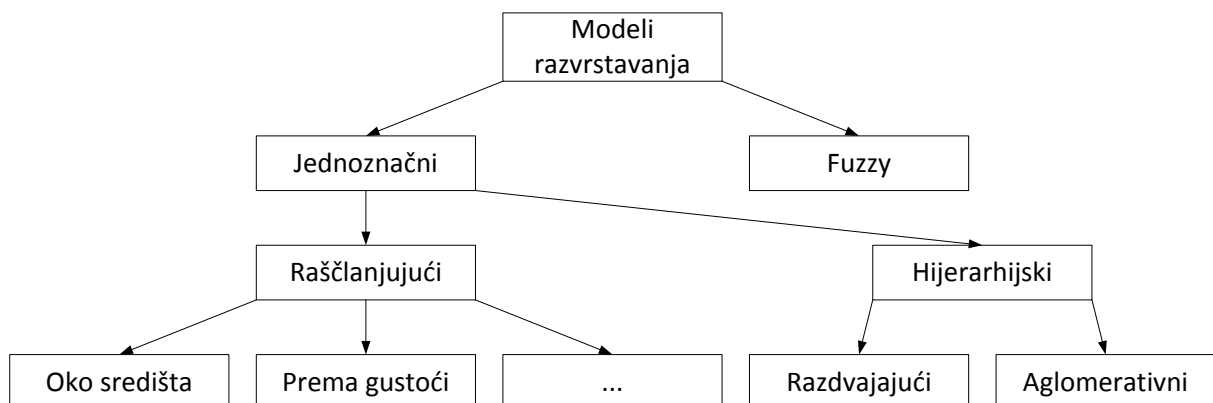
$$o_{ji} \in \{0,1\}, \quad 1 \leq j \leq k, 1 < i < n$$

$$\sum_{j=1}^k o_{ji} = 1, \quad 1 \leq i \leq n \quad (2.2)$$

$$\sum_{i=1}^n o_{ji} > 0, \quad 1 \leq j \leq k$$

Nasuprot jednoznačnom razvrstavanju postoji i višeznačno, tj. *fuzzy* razvrstavanje. Prema ovom modelu razvrstavanja objekt može pripadati jednom ili više grozdova s odgovarajućim vjerojatnostima pripadnosti pojedinom grozdu, po pravilima neizrazite (eng. fuzzy) logike.

Na slici 2.3 prikazana je detaljna podjela modela razvrstavanja. Jednoznačni algoritmi mogu se podijeliti u dvije grupe: raščlanjujući i hijerarhijski model. Raščlanjujući model dijeli se na brojne modele od kojih su najpoznatiji model razvrstavanja oko središnje točke (eng. Centroid Model) i model razvrstavanja prema gustoći groza (eng. Density-based Model),

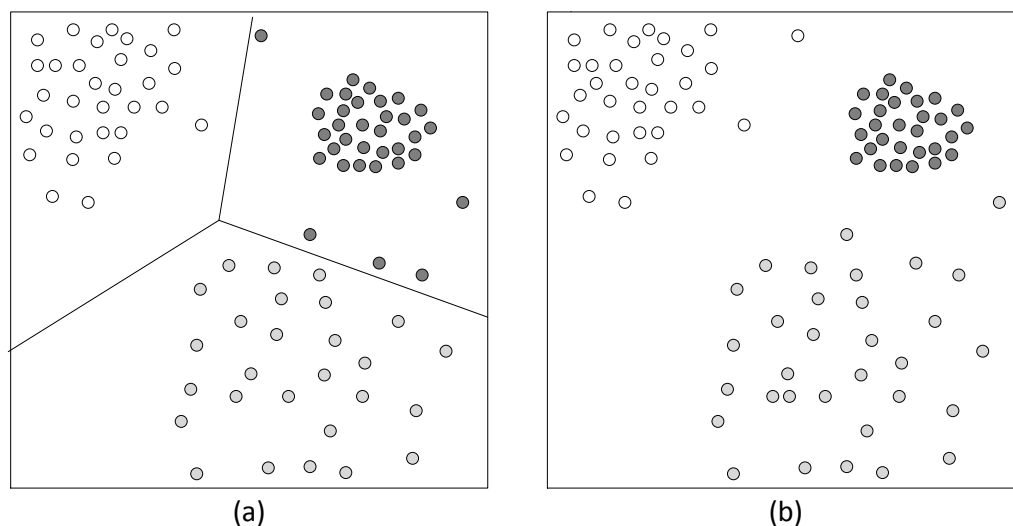


Slika 2.3: Modeli razvrstavanja podataka.

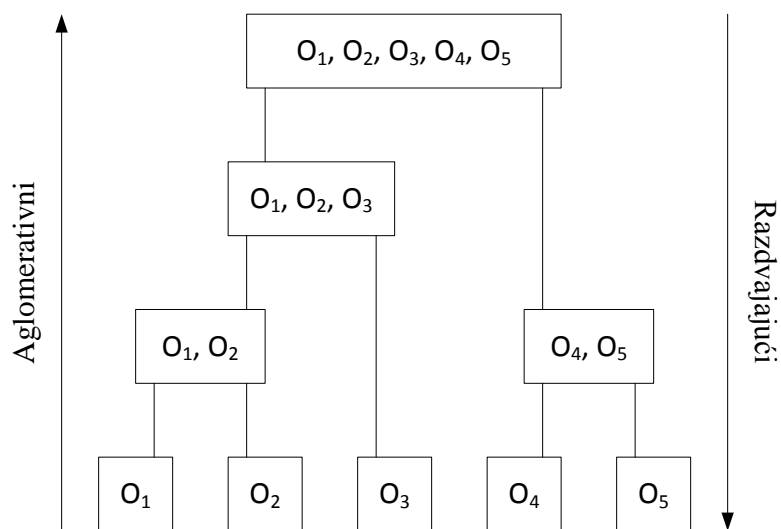
prikazani na slici 2.4. Hijerarhijski algoritmi dijele se na razdvajajuće hijerarhijske algoritme i aglomerativne hijerarhijske algoritme.

Hijerarhijski se modeli ne upotrebljavaju kod velikih skupova podataka, osim u slučajevima kada su jedini izbor. Razlog tome je što zahtijevaju  $n^2$  memorijskog prostora i  $n^3$  procesorskog vremena, gdje je  $n$  broj objekata [10]. U razdvajajućem hijerarhijskom modelu kreće se s jednim velikim grozdom u kojem se nalaze svi objekti i on se postupno razdvaja u manje grozdove, po principu od vrha prema dnu. U aglomerativnom hijerarhijskom modelu odvija se obrnuti proces, tj. na početku se u svakom grozdu nalazi samo jedan objekt i grozdovi se postupno spajaju u veće grozdove dok ne ostane jedan grozd u kojem se nalaze svi objekti, po princip od dna prema vrhu, kao na slici 2.5.

Postoji veći broj modela razvrstavanja podataka i napravljena su brojna istraživanja kako bi se riješili općeniti problemi, no većina problema koji se susreću u praksi ima



Slika 2.4: Raščlanjujući modeli razvrstavanja podataka (a) oko središta (b) prema gustoći.



Slika 2.5: Hijerarhiski modeli razvrstavanja podataka

specifične zadatke s problemima koji se ne mogu riješiti općenitim modelima te zahtijevaju specifične metode razvrstavanja. U literaturi se mogu pronaći općeniti modeli i različita istraživanja gdje se ti modeli prilagođavaju svakom specifičnom problemu iz prakse.

### 2.2.3. Podaci s nedostacima

Primjenom razvrstavanja u realnim uvjetima pojavljuju se dva dodatna problema. Prvi je da neki podaci, koje je potrebno razvrstati, jednostavno nisu dostupni i drugi problem da imaju neku vrstu greške. Ako u nekom skupu podataka nedostaju sva mjerenja za jedan zapis ili sva mjerenja za određenu varijablu tog zapisa tada se taj zapis mora ukloniti iz skupa objekata, a ako nedostaju samo neki podaci o varijabli tada se može primijeniti jedan od dva postupka. Prvi se postupak obavlja prije procesa razvrstavanja tako da se varijabli kojoj nedostaju podaci dodijeli srednja vrijednost ostalih podataka te varijable iz skupa podataka. Drugi postupak određuje vrijednost varijable tijekom samog procesa razvrstavanja koristeći jednu od metoda opisanih u [11].

Podaci s nedostacima i postupci za njihovo nadomještanje unose i pojam neodređenosti u razvrstavanje. O razvrstavanju objekata koji sadrže neodređenost u svojim prostornim podacima bit će više riječi u trećem poglavlju.

## 2.3. Sličnost i različitost kao mjere pri razvrstavanju objekata

U jednom grozdu nalaze se objekti međusobno sličniji od objekata izvan grozda. Svrha razvrstavanja je međusobno slične objekte objediniti u jedan grozd, ali jednako je važno da međusobno različiti objekti ne budu razvrstani u isti grozd. Zato su sličnosti i različitosti jednako važne pri razvrstavanju. U literaturi se pronalaze termini: mjera sličnosti (eng. Similarity Measures), koeficijent sličnosti (eng. Similarity Coefficients), mjera različitosti (eng. Dissimilarity Measure) i udaljenost (eng. Distance), kojima se kvantitativno opisuju sličnost i različitost dva objekta ili dva grozda. Udaljenost i sličnost su dva suprotstavljena koncepta. Mjera sličnosti i koeficijent sličnosti upotrebljavaju se kada se želi kvantitativno opisati koliko su dva objekta slična. Što je mjera sličnosti veća objekti su sličniji jedan drugome. S druge strane, ako se želi istaknuti razlika između dva objekta upotrijebiti će se mjera različitosti ili udaljenost, gdje će se istaknuti da što je udaljenost veća je i različitost među objektima. Postoji osnovna razlika između mjera sličnosti i mjera različitosti, gdje se termin sličnost često u literaturi koristi kao sinonim za povezanost ili korelaciju.



Ako se promotre dva objekta  $o = (o_1, o_2, \dots, o_m)$  i  $p = (p_1, p_2, \dots, p_m)$ , opisana koordinatama u  $m$ -dimenzionalnom prostoru, tada se koeficijent sličnosti između njih može opisati kao funkcija vrijednosti njihovih atributa. U ovom slučaju atributi su koordinate objekata u prostoru, kako je prikazano jednadžbom (2.3).

$$s(o, p) = s(o_1, o_2, \dots, o_m, p_1, p_2, \dots, p_m) \quad (2.3)$$

Koeficijent sličnosti ima svojstvo komutativnosti, tj. vrijedi da je  $s(o, p) = s(p, o)$ . Vrijednosti kojima se opisuje koeficijent sličnosti su brojevi između  $0 \leq s(o, p) \leq 1$  i ne može poprimiti negativnu vrijednost. Također vrijedi da je  $s(o, p) = 1$  ako je  $o = p$ , tj. za dva ista objekta koeficijent sličnosti jednak je jedan.

Ako se kao metrika različitosti između dvije točke koristi udaljenost, ona mora zadovoljavati sljedeća svojstva:

- $d(o, p) \geq 0$ ,
- $d(o, p) = 0$  ako je  $o = p$ ,
- $d(o, p) = d(p, o)$ ,
- $d(o, p) \leq d(o, z) + d(p, z)$ .

gdje su  $o$ ,  $p$  i  $z$  proizvoljne točke u  $n$ -dimenzionalnom prostoru. Prvo svojstvo govori da udaljenost između dva objekta ne može biti negativna. Drugo svojstvo je svojstvo refleksivnosti, a treće je komutativnost. Četvrto svojstvo je svojstvo nejednakosti trokuta.

Prema [12] postoji 12 različitih struktura za opis mjere sličnosti. U sljedećim poglavljima bit će opisane neke od njih.

### 2.3.1. Matrica blizine

Matrica blizine, objašnjena u [8], sadrži podatke o međusobnoj blizini indeksiranih parova skupa objekata. To je kvadratna simetrična matrica u kojoj se indeks blizine odnosi na mjeru sličnosti ili mjeru različitosti. Matrica blizine sadrži podatke o bliskosti između svih parova objekata i opisuje koliko je objekt sličan s bilo kojim drugim objektom u skupu objekata u ovisnosti o njihovoj udaljenosti. Za  $m$ -dimenzionalni prostorni skup objekata  $D = \{o_1, o_2, \dots, o_n\}$  dana je udaljenosti od objekta  $o_i$  do ostalih objekata u promatranom prostoru koja predstavlja jedan red u matrici udaljenosti  $M_u(D)$ , koja je opisana jednadžbom (2.4).

$$M_u(D) = \begin{pmatrix} 0 & d_{12} & \dots & d_{1n} \\ d_{21} & 0 & \dots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & 0 \end{pmatrix} \quad (2.4)$$

gdje je  $d_{ij} = d(o_i, o_j)$ , a  $d$  predstavlja neku od funkcija udaljenosti kao što su minowski, euklidsku i sl. Funkcije udaljenosti detaljnije su objašnjene u sljedećim poglavljima. Iz matrice (2.4) vidi se da sve udaljenosti  $d_{ii} = 0$  što zadovoljava drugo svojstvo kada se udaljenost koristi kao metrika različitosti između dvije točke.

Matrica sličnosti za isti skup objekata  $D = \{o_1, o_2, \dots, o_n\}$  opisana je jednadžbom (2.5).

$$M_s(D) = \begin{pmatrix} 1 & s_{12} & \dots & s_{1n} \\ s_{21} & 1 & \dots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \dots & 1 \end{pmatrix} \quad (2.5)$$

gdje je  $s_{ij} = s(x_i, x_j)$ , a  $s$  predstavlja neku od mjera sličnosti.

Matrica udaljenosti  $M_u(D)$  i matrica sličnosti  $M_s(D)$  nad skupom objekata  $D$  su primjeri matrica blizine. Ako su funkcija udaljenosti i mjera sličnosti simetrične, tada će i dvije matrice blizine također biti simetrične.

### 2.3.2. Graf blizine

Uz matricu blizine kao mjera sličnosti može se koristiti i graf blizine. Graf blizine je težinski graf u kojem čvorove čine razvrstani objekti, a grane koje ih spajaju predstavljaju koeficijente sličnosti. Za isti skup  $D$  vrijednosti koeficijenta sličnosti iste su u oba slučaja, ali su predstavljene na različit način. Usmjereni grafovi blizine jedanki su asimetričnim matricama blizine, a neusmjereni grafovi jednaki su simetričnim matricama blizine.

### 2.3.3. Matrica razasutosti

Matrica razasutosti za  $m$ -dimenzionalni prostorni skup objekata  $D = \{o_1, o_2, \dots, o_n\}$  u kojem je svaki objekt  $o_i$  opisan vektorom  $o_i = (o_{i1}, o_{i2}, \dots, o_{im})$  računa se pomoću jednadžbe (2.6),

$$M_r(D) = \sum_{i=1}^n (o_i - \bar{o})^T (o_i - \bar{o}) \quad (2.6)$$

gdje je  $\bar{o}$  aritmetička srednja vrijednost koja se računa po jednadžbi (2.7), iz [10].

$$\bar{o} = \frac{1}{n} \sum_{i=1}^n o_i \quad (2.7)$$

Iz jednadžbe (2.6) vidljivo je da se radi o sumi kvadrata.

Trag matrice računa se jednadžbom (2.8) i predstavlja statističku razasutost skupa objekata  $D$ .

$$Tr(M_r(D)) = \sum_{i=1}^n (o_i - \bar{o}) (o_i - \bar{o})^T \quad (2.8)$$

Za neki grozd  $c_i \in D$  matrica razasutosti  $M_r(c_i)$  naziva se matrica unutrašnje razasutosti od  $c_i$  i označava s  $M_{ur}(c_i)$ . Za skup grozdova  $C = \{c_1, c_2, \dots, c_k\}$  koji je podskup skupa  $D$  matrica unutrašnje razasutosti grozdova računa se po jednadžbi (2.9),

$$M_{ur}(C) = \sum_{i=1}^k \sum_{o \in c_i} (o - z_i)^T (o - z_i) \quad (2.9)$$

gdje je  $z_i$  srednja vrijednost grozda  $c_i$  i računa se po jednadžbi (2.10)

$$z_i = \frac{1}{|c_i|} \sum_{o \in c_i} o \quad (2.10)$$

Dok se matrica „vanjske“ razasutosti grozdova (razasutost između grozdova) može izračunati jednadžbom (2.11).

$$M_{vr}(C) = M_r(D) - M_{ur}(C) \quad (2.11)$$

### 2.3.4. Matrica kovarijanci

Kovarijanca je poznati pojam u statistici i može služiti kao mjera sličnosti. To je mjera koja kazuje koliko se dvije slučajne varijable mijenjaju zajedno. Ako je  $D = \{o_1, o_2, \dots, o_n\}$  skup objekata, gdje je svaki objekt opisan s  $m$  atributa  $o_i = (o_{i1}, o_{i2}, \dots, o_{im})$ , onda je kovarijanca između takva dva takva atributa  $o_{ip}$  i  $o_{iq}$  definirana kao omjer sume produkata njihovih standardnih devijacija od srednjeg broja i računa se jednadžbom (2.12).

$$c_{pq} = \frac{1}{n} \sum_{i=1}^n (o_{ip} - \bar{o}_p)(o_{iq} - \bar{o}_q) \quad (2.12)$$

gdje je  $o_{ij}$   $j$ -ti atribut objekata  $o_i$  i  $\bar{o}_j$  je srednja vrijednosti svih podataka na  $j$ -tom atributu. Srednja vrijednost svih podataka na  $j$ -tom atributu računa se jednadžbom (2.13).

$$\bar{o}_j = \frac{1}{n} \sum_{i=1}^n o_{ij}, j = 1, 2, \dots, m \quad (2.13)$$

Nakon što se izračunaju svi elementi dobije se  $m \times m$  matrica, u kojoj svaki unos  $(p, q)$  predstavlja kovarijancu između atributa  $o_p$  i  $o_q$  i ta se matrica naziva matrica kovarijanci, prikazana je jednadžbom (2.14).

$$\Sigma = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mm} \end{pmatrix} \quad (2.14)$$

### 2.3.5. Udaljenost kao mjera brojčanih podataka

U ovom poglavlju opisane su različite vrste funkcija računanja udaljenosti koje se koriste pri razvrstavanju objekata. Koju od predstavljenih udaljenosti odabrati kao mjeru u pojedinoj primjeni najviše ovisi o načinu kako je opisan skup objekata. Kako će uskoro biti detaljno objašnjeno različite udaljenosti najbolje odgovaraju različitim slučajima. Odabir najbolje mjere najčešće je rezultat iskustva, vještine i znanja. U svim opisima izračunavati će se udaljenost između dva objekta  $x = (x_1, x_2, \dots, x_m)$  i  $y = (y_1, y_2, \dots, y_m)$ .

#### Udaljenost Minkowskog

Udaljenost Minkowskog je metrika za udaljenost između dva objekta  $x$  i  $y$ . Računa se kao  $p$ -ti korijen sume apsolutnih razlika između svih parova vrijednosti potenciranih na  $p$ -tu potenciju te je opisana jednadžbom (2.15),

$$d(x, y) = \sqrt[p]{\sum_{i=1}^m |x_i - y_i|^p}, p \geq 1 \quad (2.15)$$

gdje su  $x_i$  i  $y_i$  vrijednosti  $i$ -tog atributa objekata  $x$  i  $y$ . Varijabla  $p$  mora biti veća ili jednaka jedan. Za različite vrijednosti varijable  $p$  dobiju se različiti oblici ove jednadžbe. Kada je  $p = 1$  udaljenost minkowskog opisuje udaljenost gradskih blokova (Manhattan udaljenost), kada je  $p = 2$  onda se govori o euklidskoj udaljenosti, a kada  $p \rightarrow \infty$  govori se o Čebiševljevoj udaljenosti (maksimalna udaljenost).

Na slici 2.6 prikazan je skup točaka u dvodimenzionalnom prostoru koji je od neke čvrste točke središta (u ovom primjeru ishodišta) udaljen za odabranu konstantnu vrijednost  $p$  za udaljenost minkowskog.

### Udaljenost gradskih blokova

Udaljenost gradskih blokova ili Manhattan udaljenost je specijalni slučaj udaljenosti minkowskog kada je  $p = 1$  i upotrebljava se za računanje udaljenosti između gradskih blokova. Opisana je kao suma udaljenosti svih atributa. Za dva objekta  $x$  i  $y$  u  $m$ -dimenzionalnom prostoru Manhattan udaljenost opisana je jednadžbom (2.16).

$$d(x, y) = \sum_{i=1}^m |x_i - y_i| \quad (2.16)$$

U nekim slučajevima nije potrebno računati sa svim atributima dimenzija [13]. U tom se slučaju koristi djelomična verzija Manhattan udaljenosti u kojoj se pri računanju udaljenosti koristi samo podskup  $P$  iz ukupnog skupa objekata  $D$ , gdje  $P$  ne smije biti prazan skup, kako je prikazano jednadžbom (2.17).

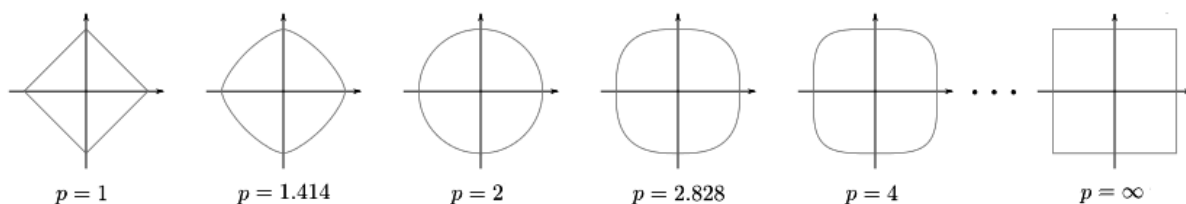
$$d(x, y) = \sum_{i \in P} \frac{|x_i - y_i|}{|P|} \quad (2.17)$$

### Euklidska udaljenost

Euklidska udaljenost je specijalna slučajeva udaljenosti Minkowskog kada je  $p = 2$  i ona je najčešće korištena mjera za udaljenost broječnih podataka. Za dva objekta  $x$  i  $y$  u  $m$ -dimenzionalnom prostoru euklidska udaljenost računa se jednadžbom (2.18).

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} = \sqrt{(x - y)(x - y)^T} \quad (2.18)$$

Postoji još jedan oblik euklidske udaljenosti koji se naziva kvadratna euklidska udaljenost i računa se po jednadžbi (2.19). Potrebno je napomenuti da kvadratna euklidska udaljenost nije zapravo udaljenost nego mjera.



Slika 2.6: Skup točaka u dvodimenzionalnom prostoru za različite vrijednosti varijable  $p$ .

$$d(x, y)^2 = \sum_{i=1}^m (x_i - y_i)^2 = (x - y)(x - y)^T \quad (2.19)$$

### Čebiševljeva udaljenost

Čebiševljeva udaljenost ili maksimalna udaljenost između dva objekta izražena je kao maksimalna apsolutna razlika između dva atributa u bilo kojoj od varijabli. Za dva objekta  $x$  i  $y$  u  $m$ -dimenzionalnom prostoru euklidska udaljenost računa se jednadžbom (2.20).

$$d(x, y) = \max_{1 \leq i \leq m} |x_i - y_i| \quad (2.20)$$

U tablici 2.1 dan je primjer rezultata udaljenosti dva objekta za sve prikazane metrike.

Tablica 2.1: Udaljenosti između objekta  $x = (6, 2, 3)$  i  $y = (4, 2, 5)$

$i$	1	2	3	
$x(i)$	6	2	3	Udaljenost
$y(i)$	4	2	5	
udaljenost minkowskog ( $p=2$ )	4	0	4	2.828
udaljenost gradskih blokova	2	0	2	4
euklidska udaljenost	4	0	4	2.828
kvadratna euklidska udaljenost	4	0	4	8
čebiševljeva udaljenost	2	0	2	2

### Udaljenosti bazirane na euklidskoj udaljenosti

Postoji slučaj kada euklidska udaljenost između dva objekta koji imaju potpuno različite atribute može biti manja od udaljenosti između dva objekta kod kojih su neki od atributa identični. U takvim slučajevima euklidsku udaljenost potrebno je izmijeniti pa se dobije nova udaljenost koja se naziva prosječna udaljenost. Za dva objekta  $x$  i  $y$  u  $m$ -dimenzionalnom prostoru prosječna udaljenost opisana je jednadžbom (2.21).

$$d(x, y) = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_i - y_i)^2} \quad (2.21)$$

Još jedna udaljenost koja je nastala izmjenom euklidske udaljenosti je tetivna udaljenost. Ona se opisuje kao duljina tetive koja spaja dvije normalizirane točke unutar kruga ili sfere s polumjerom jednakim jedan [14]. Za dva objekta  $x$  i  $y$  u  $m$ -dimenzionalnom prostoru tetivna udaljenost opisana je jednadžbom (2.22),

$$d(x, y) = \sqrt{2 - 2 \frac{\sum_{i=1}^m x_i y_i}{\|x\|_2 \|y\|_2}} \quad (2.22)$$

gdje  $\|x\|_2$  i  $\|y\|_2$  predstavljaju  $L_2$  normu i računaju se jednažbom (2.23).

$$\|x\|_2 = \sqrt{\sum_{i=1}^m x_i^2}, \quad \|y\|_2 = \sqrt{\sum_{i=1}^m y_i^2} \quad (2.23)$$

Tetivna udaljenost rješava problem sa skaliranjem mjera udaljenosti i rješavanja, gore navedene, nedostatke euklidske udaljenosti. Modifikacija tetivne udaljenosti koja se zove geodetska udaljenost objašnjena je u [10].

### 2.3.6. Sličnost kao mjera brojčanih podataka

U ovom poglavlju opisane su dvije vrste računanja sličnosti koje se koriste pri razvrstavanju objekata. U svim opisima izračunavati će se udaljenost između dva objekta  $x = (x_1, x_2, \dots, x_m)$  i  $y = (y_1, y_2, \dots, y_m)$ .

#### Kosinusna sličnost

Kosinus kuta između dva vektora može se koristiti za predstavljanje mjere sličnosti. Kosinus kuta jednak je koeficijentu korelacije i računa se po jednažbi (2.24). Kada su dva vektora iste orijentacije sličnost je 1, ako je kut između njih  $90^\circ$  mjera sličnosti je 0 [15].

$$s(x, y) = \frac{\sum_{i=1}^m (x_i y_i)}{\sqrt{\sum_{i=1}^m x_i^2 \sum_{i=1}^m y_i^2}} \quad (2.24)$$

#### Pearsonova korelacija

Pearsonov koeficijent korelacije između dva objekta koristi se kao mjera sličnosti između ta dva objekta i dan je jednažbom (2.25),

$$s(x, y) = \frac{1}{m-1} \sum_{i=1}^m \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right) \quad (2.25)$$

gdje su  $\bar{x}$  i  $\bar{y}$  aritmetičke sredine svih atributa objekta  $x$  i  $y$ , a  $s_x$  i  $s_y$  računaju se jednažbom (2.26).

$$s_x = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})^2}, \quad s_y = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{y})^2} \quad (2.26)$$

## 2.4. Sličnost i različitost kao mjera između grozdova

Kod razvrstavanja hijerarhijskim aglomerativnim algoritmima u svakoj se iteraciji dvije najbliže grupe grozdova spajaju kako bi se od njih stvorio novi grozd. Ovaj proces nastavlja se dok se ne dobije željeni broj grozdova ili dok se svi grozdovi ne spoje u jedan grozd. Suprotan proces odvija se kod upotrebe hijerarhijskih razdvajajućih algoritama razvrstavanja. U ovom procesu svi se objekti nalaze u jednom grozdu koji se postupno dijeli u manje grozdove. Kako bi se mogla dva grozda spojiti u jedan ili jedan grozd podijeliti na dva potrebno je znati mjeru sličnosti ili različitosti. To se postiže izračunom udaljenosti između središta grozda i objekata te udaljenosti između dva grozda. U primjeru koji slijedi računaju se mjere sličnosti između dva grozda  $c_1 = \{o_{11}, o_{12}, \dots, o_{1n}\}$  i  $c_2 = \{o_{21}, o_{22}, \dots, o_{2m}\}$ , gdje su  $n$  i  $m$  broj objekata u pojedinom grozdu.

### Udaljenost između središta grozdova

Kao i u prethodnom potpoglavlju, način na koji se mogu prikazati razlike između grozdovima mjerenje je udaljenosti između središta dva grozda. Ako grozdovi sadrže objekte opisane brojčanim podacima tada se udaljenost između središta grozdova može izračunati prema jednadžbi (2.27),

$$D(c_1, c_2) = d(\mu(c_1), \mu(c_2)) \quad (2.27)$$

gdje je  $d$  neka od funkcija za računanje udaljenosti, a  $\mu(c_1)$  i  $\mu(c_2)$  su središta grozdova izračunata prema jednadžbi (2.28).

$$\mu(c_1) = \frac{1}{|c_1|} \sum_{x \in c_1} x, \quad \mu(c_2) = \frac{1}{|c_2|} \sum_{x \in c_2} x \quad (2.28)$$

### Najmanja udaljenost među grozdovima

Za mjeru sličnosti između dva grozda može se koristiti i udaljenost između dva najbliža objekta iz susjednih grozdova. Ova mjera računa se tako da se odabere jedna od funkcija udaljenosti  $d$  između dva grozda  $c_1$  i  $c_2$  i računa se po jednadžbi (2.29).

$$D_{min}(c_1, c_2) = \min_{1 \leq i \leq n; 1 \leq j \leq m} d(o_{1i}, o_{2j}) \quad (2.29)$$

Na slici 2.7(a) prikazan je način računanja najmanje udaljenosti između grozdova u dvodimenzionalnom prostoru.

### Najveća udaljenost među grozdovima

Kao mjera sličnosti može se koristiti i najveća udaljenost između grozdova  $c_1$  i  $c_2$ , tj. udaljenost između dva najudaljenija objekta iz ta dva grozda. Ova mjera računa se tako da se



odabere jedna od funkcija udaljenosti  $d$  između dva grozda  $c_1$  i  $c_2$  i računa se po jednadžbi (2.30).

$$D_{max}(c_1, c_2) = \max_{1 \leq i \leq n; 1 \leq j \leq m} d(o_{1i}, o_{2j}) \quad (2.30)$$

Na slici 2.7(b) prikazan je način računanja najveće udaljenosti među grozdovima u dvo-dimenzionalnom prostoru.

### Prosječna udaljenost među grozdovima

Za mjeru sličnosti dva grozda može se izabrati i prosječna udaljenosti između grozdova [16]. Prosječna udaljenost između dva susjedna grozda za danu funkciju udaljenosti  $d$  između dva grozda  $C_1$  i  $C_2$  može se izračunati prema sljedećoj jednadžbi (2.31).

$$D_{avg}(c_1, c_2) = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m d(o_{1i}, o_{2j}) \quad (2.31)$$

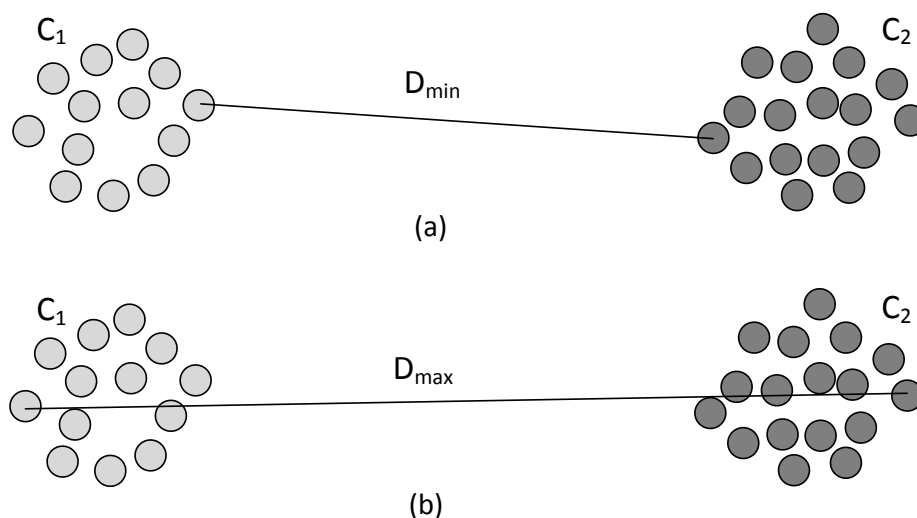
U istoj literaturi se navodi i statistička udaljenost koja se računa po jednadžbi (2.32),

$$D_{avg}(c_1, c_2) = \frac{nm}{n+m} (\bar{o}_1 - \bar{o}_2)(\bar{o}_1 - \bar{o}_2)^T \quad (2.32)$$

gdje su  $\bar{o}_1$  i  $\bar{o}_2$  srednje vrijednosti koje se računaju koristeći jednadžbu (2.33).

$$\bar{o}_1 = \frac{1}{n} \sum_{i=1}^n o_{1i}, \quad \bar{o}_2 = \frac{1}{m} \sum_{j=1}^m o_{2j} \quad (2.33)$$

Ako je  $c$  grozd koji nastaje spajanjem grozdova  $c_1$  i  $c_2$ , i ako su  $M_r(c)$ ,  $M_r(c_1)$  i  $M_r(c_2)$  matrice razasutosti grozdova  $c$ ,  $c_1$  i  $c_2$  onda vrijedi jednadžba (2.34).



Slika 2.7: Udaljenost između najbližih (a) i najudaljenijih (b) objekata dvaju susjednih grozdova.

$$M_r(c) = M_r(c_1) + M_r(c_2) + \frac{nm}{n+m} (\bar{o}_1 - \bar{o}_2)^T (\bar{o}_1 - \bar{o}_2) \quad (2.34)$$

Matrica  $\frac{nm}{n+m} (\bar{o}_1 - \bar{o}_2)^T (\bar{o}_1 - \bar{o}_2)$  naziva se matrica među-razasutosti i trag ove matrice je jednak statističkoj udaljenosti između grozdova  $c_1$  i  $c_2$ .

## 2.5. Metode za razvrstavanje objekata oko središnje točke

U ovoj se disertaciji najviše govori o jednoznačnom raščlanjujućem razvrstavanju objekata oko središnje točke. U disertaciji se upotrebljavaju veliki skupovi  $m$ -dimenzionalnih podataka koje je potrebno razvrstati te se zbog toga koriste metode ovog tipa razvrstavanja jer imaju veću učinkovitost od metoda hijerarhijskog razvrstavanja. U ovom poglavlju bit će objašnjene najčešće korištene metode kao i njihove prednosti i nedostaci.

Ovaj tip algoritama nije pogodan za razvrstavanje objekata u grozdove nepravilnih oblika. Razlog tome je što su grozdovi koji nastaju ovim postupkom konveksnog oblika i prikazuju se pomoću središnje točke koja se naziva središte grozda. Algoritmi ovog tipa imaju kao početni parametar, osim podataka o objektima i broj grozdova, ciljnu funkciju koja određuje točnost razvrstavanja. Cilj algoritma je što više smanjiti ciljnu funkciju i on se ponavlja sve dok se ciljna funkcija ne zadovolji. Ciljna funkcija je mjera koja se najčešće predstavljena minimalnim pomakom središta grozda u odnosu na središte iz prethodne iteracije.

### 2.5.1. *k-means* metoda

*k-means* metoda jedna je od najčešće upotrebljivanih metoda za razvrstavanje objekata s brojčanim podacima [10]. Ovu metodu razvilo je nekoliko znanstvenika kroz nekoliko znanstvenih polja, od kojih treba izdvojiti Lloyda (1957), Forgeya (1965), Friedman i Rubin (1967) i McQueen (1967) [17]. Detaljan povijesni pregled razvoja *k-means* metode i nekoliko njezinih inačica dan je u [8]. Razvijena je u svrhu razvrstavanja brojčanih podataka u  $k$  grozdova, od kojih je svaki opisan središnjom točkom (eng. Mean). Broj grozdova  $k$  je konstantan i kao takav ne mijenja se kroz sve iteracije postupka razvrstavanja. On je zadan na samom početku procesa zajedno s ciljnom funkcijom.

Postupak razvrstavanja počinjem odabirom početnih koordinata središta  $k$  grozdova i pridruživanjem objekata najbližem grozdu. Nakon ovog dijela postupka računaju se nova središta grozdova na osnovu srednje vrijednosti svih objekata koji su pridruženi grozdu. U

sljedećoj iteraciji ponovno se obavlja dodjeljivanje objekata novim središtima grozdova i proces se ponavlja dok se ne zadovolji ciljna funkcija ili dok objekti ne prestanu mijenjati pripadnost pojedinom grozdu, tj. dok ne konvergiraju. Ciljna funkcija je funkcija pogreške i određuje kraj razvrstavanja.

Neka je skup objekata  $O = \{o_1, o_2, \dots, o_n\}$  i neka je  $C = \{c_1, c_2, \dots, c_k\}$  skupa grozdova skupa  $O$  tada se funkcija pogreške  $E$  računa jednadžbom (2.35),

$$E = \sum_{i=1}^k \sum_{o \in c_i} d(o, \mu(c_i)) \quad (2.35)$$

gdje je  $\mu(c_i)$  središte grozda  $c_i$ , a  $d(o, \mu(c))$  udaljenost između objekta  $o$  i središta grozda. Funkcija udaljenosti  $d$  može biti bilo koja udaljenost objašnjena u 2.3.5 poglavlju, a najčešće se koristi euklidska udaljenost.

Pseudo kôd *k-means* metode prikazan je algoritmom 2.1. Algoritam se može podijeliti na dva dijela. Prvi dio naziva se inicijalizacijski i prikazan je drugom linijom kôda. U inicijalizacijskom dijelu algoritam slučajnim odabirom raspodjeljuje objekte u  $k$  grozdova. U drugom dijelu, od 3 do 11 linije kôda, koji se naziva iterativni dio računaju se udaljenosti između objekata i grozdova te se objekti se dodjeljuju najbližem grozdu. Ovaj postupak se radi za sve objekte i računa se njihova udaljenost prema svim grozdovima.

Ako se *k-means* metoda promatra kao optimizacijski problem s ciljem minimiziranja ciljne funkcije prema određenim uvjetima, onda se ciljna funkcija može definirati jednadžbom (2.36),

$$P(W, Q) = \sum_{j=1}^k \sum_{i=1}^n w_{ij} d(o_i, q_j) \quad (2.36)$$

**Algoritam 2.1:** Pseudo kôd *k-means* algoritma za razvrstavanje.

- 
1. **Ulaz:** skup grozdova  $C$  i skup objekata  $O$
  2. grozdovima na početnim pozicijama dodjeli najbliže objekte iz skupa  $O$
  3. **do**
  4.     **for all**  $o_i \in O$
  5.         izračunati  $d(o_i, c_j)$
  6.         pridružiti objek grozdu s najmanjim  $d$
  7.     **end for all**
  8.     **for all**  $c_j \in C$
  9.         izračunati nove pozicije središta grozdova
  10.     **end for all**
  11. **while** središta grozdova ne konvergiraju
-

gdje je  $Q = \{q_1, q_2, \dots, q_k\}$  skup grozdova, funkcija udaljenosti  $d$  je euklidska, a  $W$  je matrica veličine  $n \times k$  koja zadovoljava uvjete dane jednadžbom (2.37).

$$\begin{aligned} 1. \quad & w_{ij} \in \{0,1\} \quad i = 1,2, \dots, n, \quad j = 1,2, \dots, k \\ 2. \quad & \sum_{j=1}^k w_{ij} = 1 \quad i = 1,2, \dots, n \end{aligned} \quad (2.37)$$

Ovaj optimizacijski problem rješava se iterativno tako da se podijeli na dva potproblema:

1. Postaviti  $Q = \hat{Q}$  i riješiti reducirani problem  $P(W, \hat{Q})$ .
2. Postaviti  $W = \hat{W}$  i riješiti reducirani problem  $P(\hat{W}, Q)$ .

Rješavanje ovih potproblema objašnjeno je u [18] pomoću dva teorema.

*Teorem 1:* Ako je  $\hat{Q} = \{\hat{q}_1, \hat{q}_2, \dots, \hat{q}_k\}$  postavljena kao fiksna vrijednost, tada je funkcija  $P(W, \hat{Q})$  minimizirana ako i samo ako vrijedi jednadžba (2.38).

$$w_{ij} = \begin{cases} 1 & \text{ako } d(o_i, \hat{q}_j) = \min_{1 \leq t \leq k} d(o_i, \hat{q}_t) \\ 0 & \text{inače} \end{cases} \quad (2.38)$$

*Teorem 2:* Ako je  $\hat{W}$  postavljeno kao fiksna vrijednost tada je funkcija  $P(\hat{W}, Q)$  minimizirana ako i samo ako vrijedi jednadžba (2.39),

$$q_{jd} = \frac{\sum_{i=1}^n \hat{w}_{ij} o_{id}}{\sum_{i=1}^n \hat{w}_{ij}} \quad (2.39)$$

gdje je  $d = 1, 2, \dots, m$  dimenzija prostora.

Vremenska složenost algoritma po iteraciji je  $O(nkm)$ , gdje je  $n$  broj objekata u skupu objekata,  $k$  je broj grozdova i  $m$  broj prostornih dimenzija. Budući da se niz  $P$  smanjuje on će nakon konačnog broja iteracija konvergirati u točku lokalnog minimuma i tako zaustaviti proces razvrstavanja [19].

Metoda predstavljena u ovom poglavlju naziva se konvencionalni ili standardni *k-means*. Postoje različite inačice *k-means* metode koje su razvijene kako bi se poboljšale performanse postupka, a predstavljene su u sljedećim poglavljima. Standardni *k-means* ima sljedeće nedostatke:

- nije učinkovit za razvrstavanje velikih skupova objekata jer je njegova složenost linearno proporcionalna veličini skupa objekata,
- često završava razvrstavanje kada se postigne lokalni optimum,
- grozdovi su konveksnog oblika,
- koristi se samo za brojčane podatke,
- performanse razvrstavanja su jako ovisne o početnom odabiru središta grozdova.

Ovo zadnje svojstvo se može riješiti korištenjem jedne od razvijenih inicijalizacijskih metoda koje predlažu različite načine izbora za početna središta grozdova [20, 21].

### 2.5.2. Kontinuirana *k-means* metoda

Prijedlog za kontinuiranu *k-means* metodu dan je u [22]. Ova metoda odlikuje se bržim procesom razvrstavanja od standardnog *k-means*-a iz dva razloga. Prvi je da se početna središta grozdova u ovoj metodi odabiru slučajnim odabirom iz skupa objekata, dok standardni *k-means* središta grozdova odabire proizvoljno. Drugi razlog poboljšane učinkovitosti leži u činjenici da neprekidni *k-means* u svakoj iteraciji ne uzima u obzir sve objekte u skupu objekata, nego samo dio skupa objekata. Slučajno uzorkovanje objekata predstavlja izvorni koncept koji je korišten za razvrstavanje u neprekidnom prostoru. Prema MacQueenu [17] funkcija pogreške  $E_i$  za svako područje  $R_i$  dana je jednadžbom (2.40),

$$E_i = \int_{R_i} \rho(x) \|x - c_i\|^2 dx \quad (2.40)$$

gdje je  $\rho(x)$  neprekidna funkcija gustoće vjerojatnosti nad područjem  $R_i$  sa središtem u  $c_i$ .

Zbrajanjem svih funkcija pogrešaka  $E_i$  dobije se ukupna funkcija pogreške. Tako se korištenjem uzorkovanja slučajnog skupa objekata može procijeniti funkcija gustoće vjerojatnosti  $\rho(x)$  i na taj način dobiti funkcija pogreške bez korištenja svih objekata iz skupa objekata. Zbog korištenja manjeg broja objekata kontinuirani *k-means* brži je od standardnog.

### 2.5.3. Usporedna *k-means* metoda

Kako bi se ubrzao standardni *k-means* postupak razvrstavanja u [23] predstavljena je nova *k-means* metoda koja na jednostavan način izbjegava računanje nepotrebnih usporedbi. Ako je  $o$  objekt iz skupa objekata  $O$ , a  $\mu_i$  i  $\mu_j$  su dva središta, tada se pomoću nejednakosti trokuta može izračunati koji od središta je nepotrebno uspoređivati prema jednadžbi (2.41).

$$d(o, \mu_i) + d(o, \mu_j) \geq d(\mu_i, \mu_j) \quad (2.41)$$

Ako je ispunjeno da je  $d(o, \mu_i) \geq d(\mu_i, \mu_j) - d(o, \mu_j)$  i ako je zadovoljen sljedeći uvjet  $d(\mu_i, \mu_j) \geq 2d(o, \mu_i)$  može se zaključiti da vrijedi  $d(o, \mu_j) \geq d(o, \mu_i)$ . U tome slučaju izbjegnuto je računanje udaljenosti  $d(o, \mu_j)$ . Budući da broj grozdova manji od broja objekata onda se udaljenosti između svih parova središta izračunava u kratkom vremenu prije svake iteracije. Prije usporedbe objekta  $o$  sa središtem  $\mu_j$  potrebno je provesti provjeru iz jednadžbe

(2.41) koristeći najbliže središe podatku  $o$ . Broj usporedbi koje napravi ova metoda teško je izračunati, no nije veći od vrijednosti  $k^2d + nkd$ , gdje je  $n$  je broj objekata,  $k$  je broj grozdova i  $d$  broj prostornih dimenzija.

#### 2.5.4. Razvrstana $k$ -means metoda

Razvrstana  $k$ -means metoda je poboljšanje usporedne  $k$ -means metode i također je objašnjen u [23]. Ovaj metoda razvrstava središta grozdova po udaljenosti od svakog središta posebno i na taj način postiže dodatno ubrzanje. Ako je  $D_{ij}$  jednak udaljenosti između dva središta grozdova  $d(\mu_i, \mu_j)$  za koje vrijedi  $i, j = 1, 2, \dots, k$ , tada se matrica  $M$  sastoji od  $k \times k$  elemenata. Redak  $i$  u matrici  $M$  predstavljaju permutacije  $(m_{i1}, m_{i2}, \dots, m_{ik})$  od 1 do  $k$  za koje vrijedi  $d(\mu_i, \mu_{m_{i1}}) \leq d(\mu_i, \mu_{m_{i2}}) \leq \dots \leq d(\mu_i, \mu_{m_{ik}})$  iz razloga što su udaljenosti sortirane.

Vremenska složenost jedne iteracije algoritma je  $O(nd\gamma + k^2d + k^2 \log k)$ , gdje je  $n$  je broj objekata,  $k$  je broj grozdova,  $d$  broj dimenzija i  $\gamma$  je aritmetička sredina objekata  $o$  u svim grozdovima koji nisu udaljeniji više od dvije udaljenosti objekta  $o$  od svog središta koje mu je dodijeljeno u prethodnoj iteraciji.

#### 2.5.5. $x$ -means metoda

Kod standardnog  $k$ -means-a broj grozdova  $k$  poznat je i određen na samom početku procesa i predaje se kao ulazni parametar. Kod  $x$ -means metode [24] koristi Bayesov informacijski kriterij (eng. Bayesian Information Iriterion - BIC) za određivanje optimalnog broja grozdova. Ako je  $D = \{o_1, o_2, \dots, o_n\}$  skup podataka s  $n$  objekata, u  $m$ -dimenzionalnom prostoru i  $M_j = \{C_1, C_2, \dots, C_k\}$  skup zamjenskih modele, gdje svaki model odgovara različitom broju grozdova  $k$ , onda se za svaki model računa vjerojatnost  $P(M_j|D)$  kako bi se ocijenili pojedini modeli. BIC to jest Schwartzov kriterij opisan je jednadžbom (2.42),

$$BIC(M_j) = \hat{l}_j(D) - \frac{p_j}{2} \log n \quad (2.42)$$

gdje je  $\hat{l}_j(D)$  logaritamska mogućnost skupa objekata  $D$  prema  $j$ -tom modelu u točki najveće mogućnosti,  $p_j = (d + 1)k$  je broj parametara u  $M_j$ . Kao najbolji model odabire se onaj s najvećom ocjenom.

Za Gaussovu razdiobu procjena vrijednosti varijance dana je jednadžbom (2.43),

$$\hat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^n (o_i - \mu_{(i)})^2 \quad (2.43)$$

gdje je  $\mu_{(i)}$  središte objekta  $o_i$ , a  $i$  označava indeks središta koje je najbliže objektu  $o_i$ . Vjerojatnost za pojedini objekt može se izračunati pomoću jednadžbe (2.44):

$$\hat{P}(o_i) = \frac{|C_{(i)}|}{n} \frac{1}{\sqrt{2\pi}\hat{\sigma}^d} \exp\left(-\frac{1}{2\hat{\sigma}^2} \|o_i - \mu_{(i)}\|^2\right) \quad (2.44)$$

I na kraju slijedi da se logaritamska mogućnost svih objekata može izračunati jednadžbom (2.45),

$$l(D) = \log \prod_{i=1}^n P(o_i) = \sum_{i=1}^n \left( \log \frac{1}{\sqrt{2\pi}\hat{\sigma}^d} - \frac{1}{2\hat{\sigma}^2} \|o_i - \mu_{(i)}\|^2 + \log \frac{|C_{(i)}|}{n} \right) \quad (2.45)$$

U *x-mean* metodi BIC kriterij koristi se globalno za određivanje najboljeg modela, a lokalno za diobu grozdova. Tijek procesa razvrstavanja može se opisati na sljedeći način. Na početku se određuju granice raspona broja grozdova tj.  $k_{min}$  i  $k_{max}$ . U prvoj iteraciji postavlja se  $k = k_{min}$  te se u svakoj novoj iteraciji dodaje novi grozd sve dok se ne dosegne najveći broj grozdova  $k_{max}$ . Novi grozd dodaje se tako da se jedan grozd razdvoji na dva grozda prema BIC kriteriju. Iteracija s najvećom ocjenom, na kraju procesa, uzima se kao optimalni broj grozdova i optimalno razvrstavanje.

### 2.5.6. *k*-harmonijska *means* metoda

Ova metoda razvrstavanja oko centrale točke nastala je iz *k-means* metode [25], a razlikuje se od standardnog *k-means*-a po svojstvu da na proces ove metode ne utječe početni izbor središta grozdova. Ime je metoda dobila po načinu izračunavanja srednje vrijednosti. Funkcija pogreške *k-means* metode  $E$  može se računati na način prikazan jednadžbom (2.46),

$$E = \sum_{i=1}^n \min(d(o, \mu_j)), \quad j = 1, 2, \dots, k \quad (2.46)$$

gdje je  $\mu_j$  središte  $j$ -tog grozda. Funkcija pogreške *k*-harmonijske *means* metode dobije se zamjenjivanjem funkcije minimuma s funkcijom harmonijske srednje vrijednosti i kvadratne euklidske udaljenosti. Harmonijsku sredinu moguće je definirati kao recipročnu aritmetičku sredinu iz obrnutih ili recipročnih elemenata. Funkcija pogreške *k*-harmonijske *means* metode računa se jednadžbom (2.47),

$$E = \sum_{i=1}^n HA(d_{keuklid}(o, \mu_j)) = \sum_{i=1}^n \frac{k}{\sum_{j=1}^k \frac{1}{(o - \mu_j)^T (o - \mu_j)}}, \quad j = 1, 2, \dots, k \quad (2.47)$$

gdje je  $d_{keuklid}$  kvadratna euklidska udaljenost, a HA harmonijska sredina opisana općom jednadžbom (2.48).

$$HA(x_i) = \frac{m}{\sum_{i=1}^m \frac{1}{x_i}}, \quad i = 1, 2, \dots, m \quad (2.48)$$

Rekurzivna formula za funkciju pogreške  $k$ -harmonijske *means* metode može se izračunati koristeći djelomično deriviranje jednadžbe (2.47) s obzirom na središta  $\mu_l$  gdje je  $l = 1, 2, \dots, k$  i izjednačujući ih s nulom. Kao rezultat dobije se jednadžba (2.49),

$$\frac{\partial E}{\partial \mu_l} = -k \sum_{i=1}^n \frac{2(o_i - \mu_l)}{d_{il}^4 (\sum_{j=1}^k d_{ij}^{-2})^2} = 0 \quad (2.49)$$

gdje je  $d_{ij} = d(o, \mu_j)$ , a  $d$  je euklidska funkcija udaljenosti.

Rješavanjem jednadžbe (2.49) dobiju se nova središta grozdova  $\mu_l^*$  pomoću jednadžbe (2.50).

$$\mu_l^* = \frac{\sum_{i=1}^n d_{il}^{-4} (\sum_{j=1}^k d_{ij}^{-2})^{-2} o_i}{\sum_{i=1}^n d_{il}^{-4} (\sum_{j=1}^k d_{ij}^{-2})^{-2}} \quad l = 1, 2, \dots, k \quad (2.50)$$

Jednadžba (2.50) se koristi u rekurzivnoj funkciji sve dok središta ne konvergiraju.

### 2.5.7. Ostale metode razvrstavanja oko središnje točke

Metoda pomicanja središta (eng. Mean Shift) jednostavan je iterativni proces koji pomiče svaki objekt na prosjek objekata u njegovoj okolini [26]. K-modes metoda je prilagođena  $k$ -*means* metoda za razvrstavanje kategoričkih podataka [27, 28]. Ideja je ove metode postaviti  $k$  početnih modova te smjestiti sve objekte u njima najbliži mod. Algoritam  $k$ -vjerojatnosti, opisan u [29], može se koristiti za razvrstavanje mješovitih tipova objekata. On koristi opći koeficijent udaljenosti između dva objekta. Algoritam  $k$ -prototipa opisan u [28], koristi se za razvrstavanje mješovitih tipova objekata. U ovom algoritmu prototip predstavlja središte grozda. MEC je metoda za razvrstavanje *fuzzy* objekata, odlučivanjem o članstvu u *fuzzy* grozdovima na temelju maksimiziranja entropije zadane srednjom varijancom.



## 2.6. Metode za razvrstavanje objekata prema gustoći

U ovom poglavlju predstavljene su različite metode za razvrstavanje prema gustoći (eng. Density-Based). Ovaj način razvrstavanja ima mogućnost pronalaska grozdova nepravilnih oblika. Pronalaze se grozdovi koji su područja s velikom gustoćom objekata. Između takvih područja nalaze se područja male gustoće. U ovom tipu razvrstavanja ne postoje iterativna ponavljanja jer se ovim metodama samo jednom pregledava čitavo područje skupa objekata i objekti se dodjeljuju grozdovima. Prednost ove metode nalazi se u činjenici da broj grozdova nije unaprijed poznat. On se određuje pregledom područja skupa objekata i određuje ovisno o gustoći objekata u promatranom području.

U [30] predstavljena je metoda koji može detektirati grozdove nepravilnog oblika nazvana akronimom DBSCAN (eng. Density-Based Spatial Clustering of Applications with Noise). Metoda u razvrstavanju koristi koncept zvan  $\epsilon$ -susjedstvo objekta. Ako postoji točka  $x$  tada se njezino  $\epsilon$ -susjedstvo označava s  $N_\epsilon(x)$  i definira na sljedeći način:

- $\epsilon$ -susjedstvo točke  $x$  opisuje se jednadžbom (2.51),

$$N_\epsilon(x) = \{y \in D: d(x, y) \leq \epsilon\} \quad (2.51)$$

gdje je  $D$  skup podataka i  $d(x, y)$  je jedna od funkcija udaljenosti.

- Za točku  $x$  može se reći da je izravno dostupna kroz gustoću od točke  $y$  s obzirom na  $\epsilon$  i  $N_{min}$  ako vrijedi jednadžba (2.52),

$$1. x \in N_\epsilon(y) \quad (2.52)$$

$$2. |N_\epsilon(y)| \geq N_{min}$$

gdje je  $|N_\epsilon(y)|$  jednak broju podataka u  $N_\epsilon(y)$ . Može se reći da je izravna dostupnost kroz gustoću simetrična za dvije točke koje se nalaze unutar iste grupe, ali to nije uvijek slučaj za točke unutar grupe i točke na granici grupe.

- Za razliku od izravne dostupnosti kroz gustoću, dostupnost kroz gustoću ne mora biti simetrična. Točka  $x$  dostupna je kroz gustoću od točke  $y$  ako postoji skup točaka  $x = x_1, x_2, \dots, x_i = y$  takav da je  $x_l$  izravno dostupan kroz gustoću objekta  $x_{l+1}$  za svaki  $l = 1, 2, \dots, i - 1$ .
- Za razliku od dostupnosti, pojam povezanost kroz gustoću je uvijek simetrična i za točke unutar grozda i točke na granici grozda. Dvije točke  $x$  i  $y$  kaže se da su povezane kroz gustoću ako postoji točka  $z$  od koje su oba objekta  $x$  i  $y$  dostupni kroz gustoću s obzirom na  $\epsilon$  i  $N_{min}$ .

- Grozd  $C$ , s obzirom na  $\epsilon$  i  $N_{min}$ , je neprazni podskup  $D$  skupa podataka, ako zadovoljava sljedeće uvjete:
  - za svaki  $x, y \in D$ , ako je  $x \in C$  i  $y$  je dostupnost kroz gustoću od  $x$ , s obzirom na  $\epsilon$  i  $N_{min}$ , onda je  $y \in C$  i
  - za svaki  $x, y \in C$ ,  $x$  i  $y$  su povezani kroz gustoću, s obzirom na  $\epsilon$  i  $N_{min}$ .

Tako se grozd može definirati kao skup točaka povezanih kroz gustoću koji je maksimalan u odnosu na dostupnost kroz gustoću.

DBSCAN metoda može se opisati na sljedeći način. Odabire se točka  $x$  i njoj se pridružuju sve točke koje su joj dostupne kroz gustoću. Ovaj način razvrstavanja primjeren je ukoliko se odabrana točka  $x$  nalazi unutar grupe točaka, no ako je izabrana rubna točka tada nema drugih točaka koje su dostupne kroz gustoću i DBSCAN metoda prelazi na sljedeću nerazvrstanu točku. DBSCAN metoda može spajati grozdove u jedan ako su dovoljno blizu jedan drugome.

Ulazni parametri koje zahtijeva ova metoda su  $\epsilon$  i  $N_{min}$ . Ovi parametri moraju biti isti za sve grozdove tijekom razvrstavanja, stoga nije lagano izabrati optimalne vrijednosti za početne parametre. U [30] se predlaže da se početne vrijednosti za  $\epsilon$  i  $N_{min}$  odrede prema „najtanjem“ grozdu u skupu objekata, to se naziva sortirani graf  $k$ -udaljenosti.

U [31] predstavljena je hibridna metoda BRIDGE koja je nastala kombinacijom  $k$ -means i DBSCAN metode. Korištenjem ove metode omogućava DBSCAN metodi razvrstavanje velikih skupina podataka, a istovremeno poboljšavaju se karakteristike  $k$ -means metode uklanjanjem šuma. U BRIDGE metodi prvo se  $k$ -means-om podijeli skup podataka (točaka, objekata) u  $k$  grozdova, a zatim se primjenom DBSCAN metode nad svakim grozdom pronalaze novi grozdovi po gustoći. Ti rezultati se uspoređuju s rezultatima DBSCAN metode nad cijelim skupom podataka i pronalazi se šum. Na kraju se ponovo provodi  $k$ -means metoda za pronalaženje centara grozdova, ali bez uklonjenog šuma.

Ostale metode razvrstavanja prema gustoći koje treba spomenuti su DBCLASD, DENCLUE i CUBN metoda koje su prikazane u [10].

## 2.7. Ostali modeli jednoznačnog raščlanjujućeg razvrstavanja objekata

Za ovu vrstu razvrstavanja postoje još dvije vrste modela. To su razvrstavanje objekata pomoću grafa i razvrstavanje objekata pomoću rešetke (eng. grid). Modeli razvrstavanja koji su temeljeni na grafovima najprije naprave graf ili hipergraf podataka, a zatim nad tim grafom

primjenjuju neku od metoda za razvrstavanje na pojedine dijelove grafa. Povezanost između teorije grafova i razvrstavanja skupa objekata prikazana je u [32]. Najpoznatije metode ovog tipa su kameleon, CACTUS i ROCK metoda.

Modeli razvrstavanja pomoću rešetke znatno smanjuju vremensku složenost, pogotovo kod velikih skupova objekata. Razlikuju se od ostalih modela u tome što ne promatraju vrijednosti pojedinih objekata, nego prostor koji ih okružuje. Prema [33] koraci od kojih se spomenuto razvrstavanje sastoji su:

1. Stvaranje rešetkaste strukture, to jest podjela skupa podataka u konačan broj ćelija.
2. Računanje gustoće ćelije, za sve ćelije.
3. Sortiranje ćelija prema njihovoj gustoći.
4. Pronalazak središta grozdova.
5. Obuhvaćanje susjednih ćelija.

Najpoznatija metoda bazirana na ovoj vrsti razvrstavanja je STING (eng. STatistical INformation Grid-based clustering method). Skup podataka dijeli se u pravokutne ćelije koje su predstavljene hijerarhijski. Hijerarhija je predstavljena roditeljsko – potomskim ćelijama. Ako je roditelj na prvoj razini, onda su njegovi potomci na drugoj razini i tako dalje. Ćelija na  $i$  razini nastaje unijom svojih ćelija potomaka na  $i - 1$  razini. U STING metodi svaka ćelija ima četiri potomka i svaki potomak predstavlja jedan od četiri kvadranta roditeljske ćelije, ovo vrijedi samo za prostore s dvije dimenzije. Ova metoda nije pogodana za razvrstavanje podataka s velikim brojem dimenzija. Ostale metode ovog modela razvrstavanja koje treba spomenuti su OPTIGrid, GRIDCLUS i GDILC metoda, koje su prezentirane u [10].

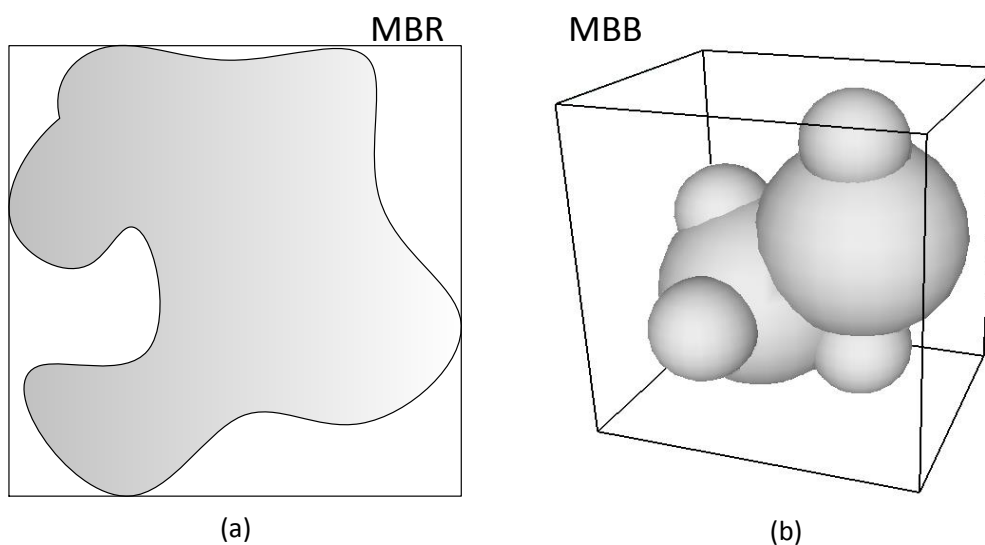
### 3. METODE RAZVRSTAVANJA I ODBACIVANJA SREDIŠTA GROZDOVA ZA OBJEKTE S PROSTORNOM NEODREĐENOSTI

Objekti s prostornom neodređenosti su objekti koji sadrže neodređenost u svojim prostornim podacima. Pozicija takvog objekata  $o_i$  ne može se predstaviti točkom u prostoru  $T(x, y, z)$ , već se predstavlja funkcijom gustoće vjerojatnosti (eng. Probability Density Function - PDF) [34]. Neodređenost objekta prikazuje se pomoću PDF-a u svrhu povećanja sličnost dobivenih rezultata sa stvarnim vrijednostima podataka o objektu [35]. Neodređenost, sama po sebi, proizlazi iz ograničenja ljudske percepcije i razumijevanja stvarnosti. Na neodređenost utječu još i ograničenja resursa, uređaja, baza podataka, analize i razumijevanje podataka. Objekt može imati prostornu neodređenost iz različitih razloga. Neodređenost može proizlaziti iz nepreciznosti i šuma prilikom mjerenja, pogreške u uzorkovanju podataka ili zbog neredovitog osvježavanja prostornih podataka s novim vrijednostima. U tim se slučajevima podaci o poziciji objekta moraju aproksimirati. Kao primjer mogu se uzeti pokretni objekti koji periodički izvještavaju svoj položaj. Određivanje položaja između dva izvještaja procjenjuje se s određenom neodređenosti koristeći pretpostavke o smjeru i brzini kretanja pokretnog objekta.

U ovoj disertaciji detaljno je opisano razvrstavanje objekata koji sadrže neodređenost u svojim prostornim podacima oko jednog središta. Promatrani su objekti u 2D i 3D prostoru koji imaju neodređenost u dvije ili tri dimenzije, ovisno o promatranom prostoru. U nekim primjenama broj objekata koji su zadani s neodređenosti može biti jako velik. Podatke o njihovoj poziciji, tj. njihovu neodređenost, potrebno je spremiti u bazu podataka. Baze podataka u kojima spremamo informacije o objektima koji sadrže neodređenost nazivaju se neodređene baze podataka. Zbog velikog broja objekata u bazi podataka i načina na koji su opisani razvrstavanje u grozdove je vremenski vrlo zahtjevno. Kako bi se minimiziralo ukupno vrijeme za razvrstavanje, potrebni su specifični algoritmi razvrstavanja i velika računalna moć. Kao i kod algoritama za razvrstavanje objekata, čija se pozicija može opisti samo jednom točkom u prostoru, objekti se dijele u grozdove prema sličnosti u svojim prostornim podacima i grupiraju oko jednog središta. Na kraju razvrstavanja dobiju se podaci koji objekt pridružuju pojedinom grozdu i određuju gdje se nalazi središte svakog grozda. Ti se podaci mogu iskoristi na različite načine; kako bi se ubrzala komunikacija i vrijeme reakcije, smanjila potrošnja energije i slično. Kako su objekti u pojedinom grozdu prostorno bliski, tako su i sva djelovanja između objekata lokalna.

Prostorna neodređenost objekta opisuje se PDF-om i to predstavlja problem klasičnim algoritma, opisanim u drugom poglavlju, za razvrstavanje objekata koji su opisani jednom točkom u prostoru. Te algoritme gotovo je nemoguće upotrijebiti za razvrstavanje objekata s neodređenosti. Ako se želi upotrijebiti klasične algoritme, opisane u drugom poglavlju ove disertacije, bilo bi potrebno neodređene objekte sažeti i predstaviti ih pomoću središnje točke. Taj način razvrstavanje je samo približno točan i može rezultirati krivim rezultatima. U [36] opisan je drugi način predstavljanja prostorne neodređenosti objekata. Objekti su opisani pomoću minimalnog graničnog područja u kojemu se mora nalaziti objekt. U 2D prostoru radi se o minimalnom graničnom pravokutniku (eng. Minimum Bounding Rectangle - MBR), a u 3D prostoru radi se o minimalnoj graničnoj kutiji (eng. Minimum Bounding Box - MBB). Korištenjem ovakvih opisa prostorne neodređenosti objekata dobiju se bolji rezultati, nego predstavljanjem pomoću središnje točke i razvrstavanjem klasičnim algoritmima. Na slici 3.1 prikazani su objekti s prostornom neodređenosti koji su opisani minimalnim graničnim područjem u 2D i 3D.

Prostorna neodređenost objekta može se promatrati na dva različita primjera. Prvi primjer se naziva egzistencijalna neodređenosti, u kojoj nije poznato postoji li uopće objekt na mjestu na kojem bi trebao biti [37]. Pozicija objekta opisana je jednom točkom u prostoru i vjerojatnošću postojanja samog objekta [38] te je na takav način spremljena u bazu podataka. Nasuprot tome postoji primjer kada je postojanje objekta u promatranom prostoru sigurno, ali je njegova pozicija neodređena. U ovom je slučaju objekt predstavljen pomoću minimalnog graničnog područja neodređenosti koje ga okružuje (MBR ili MBB). Vjerojatnost da se objekt



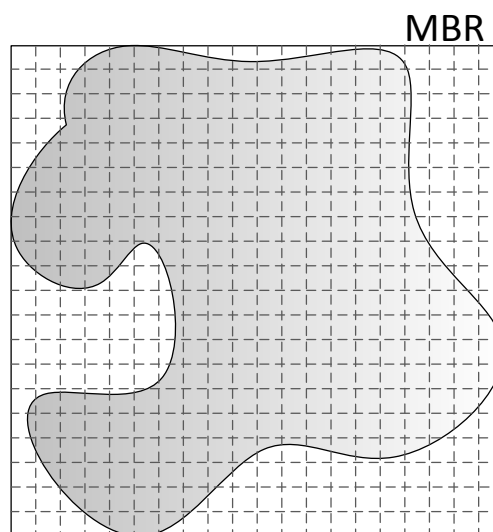
Slika 3.1: Objekti s prostornom neodređenosti koji su opisani minimalnim graničnim područjem: (a) u 2D prostoru (b) u 3D prostoru.

nalazi izvan tog područja je jednaka nuli [36, 39]. U ovoj disertaciji proučava se drugi slučaj u kojem je sigurno da objekt postoji, ali njegovi prostorni podaci sadrže neodređenost.

Ovako predstavljanje neodređenosti pruža mogućnost da se PDF zamjeni s minimalnim područjem neodređenosti koje je podijeljeno na uzorke. Svakome uzorku potrebno je dodijeliti vjerojatnost da se objekt nalazi na tome uzorku. Zbroj svih vjerojatnosti uzoraka mora biti jednak jedan, kako je prikazano u jednadžbi (3.1).

$$\int_{x \in R^m} f_i(x) dx = 1 \quad (3.1)$$

Cilj razvrstavanja može biti pronalaženje područja najveće gustoće objekata s neodređenosti [40, 41, 30] ili minimiziranje ciljne funkcije. Kao primjer ciljne funkcije može se uzeti minimiziranje euklidske udaljenosti od objekata do središta grozda [34], ukupne kvadratne udaljenosti [17], Minkowski udaljenosti ili Manhattan udaljenosti. Većina algoritama različitih metoda za razvrstavanje objekata s neodređenosti koristi skup uzoraka za opisivanje neodređenosti. Kako je ranije objašnjeno, svaki taj uzorak sadrži vjerojatnost da se objekt nalazi baš tamo. Kako bi se zadovoljila preciznost, potreban je veliki broj uzoraka za opisivanje PDF-a. Na slici 3.2 prikazan je MBR podijeljen na 400 uzoraka za primjer objekta s dvodimenzionalnom prostornom neodređenosti sa slike 3.1(a). Prilikom razvrstavanja potrebno je izračunati očekivanu udaljenost (eng. Expected Distance - ED) između svakog objekata i svih grozdova. Računanje ED za jedan objekt s prostornom neodređenosti zahtijeva računanje udaljenosti od svih uzoraka MBR-a kojim je opisan objekt do središta grozda. U [34]



Slika 3.2: Objekt s dvodimenzionalnom prostornom neodređenosti opisan MBR-om i podijeljen na 400 uzoraka.

objašnjeno je da je zbog toga potrebno vrijeme računanja ED-a i po nekoliko stotina puta duže od jednostavnog računanja euklidske udaljenosti. Za navedni primjer sa slike 3.2, gdje je za predstavljanje PDF-a korišteno 400 uzoraka, potrebno je računati 400 udaljenosti između tog objekta i grozda. Iz tog proizlazi zaključak da je računanje ED-a 400 puta vremenski zahtjevnije nego računanje euklidske udaljenosti.

Kako je već navedeno, problem klasičnih algoritma za razvrstavanje objekata koji su opisani jednom točkom u prostoru je taj da su gotovo neupotrebljivi kada je pozicija objekta zadana s neodređenosti. Kako ti algoritmi nemaju mogućnost razvrstavati neodređene objekte bilo je potrebno razviti nove ili izmijeniti postojeće algoritme. Jedan od najjednostavniji algoritama za razvrstavanje objekata s neodređenosti opisan je u [36]. Radi se o *uk-means* algoritmu koji je nastao izmjenom *k-means* algoritma. Koristi se za razvrstavanje objekata s neodređenosti oko središnje točke grozda na takav način da ukupni ED od objekata do grozdova bude minimalan. Tu se opet vidi nedostatak zbog potrebe računanja ED od svakog objekata do svih središta grozdova. Iako *uk-means* algoritam daje točne rezultate razvrstavanja objekata s neodređenosti kod velikog broja objekata on je neučinkovit. Detaljan opis *uk-means* metode bit će opisan nešto kasnije. Neučinkovitost *uk-means* algoritma proizlazi iz računanja ED-a koje je vremenski i strojno najzahtjevniji dio postupka razvrstavanja. Iz tog razloga nastala je potreba za novim algoritmima u kojima se nastoji smanjiti vrijeme potrebno za razvrstavanje objekata na način da se izbjegne računanje ED-a. Taj se postupak naziva odbacivanje (eng. Pruning) središta grozdova kao mogućih kandidata za pojedine objekte. Ovim postupkom pronalaze se ona središta grozdova kojima objekt sigurno ne može biti dodijeljen te se oni odbacuju bez računanja očekivane udaljenosti. Takvi algoritmi pridonose većoj učinkoviti i smanjenju trajanja postupka razvrstavanja.

Najpoznatije i najčešće korištene metode odbacivanja središta grozdova kao kandidata, bez računanja ED, su MinMax [42], Voronojevi dijagrami [43, 44] i SPP metoda [45, 46]. MinMax metoda pronalazi minimalnu i maksimalnu udaljenost od svakog objekta do središta grozda. Zatim se iz dobivenih rezultata za maksimalnu udaljenost pronalazi najmanja vrijednost. Ona se zatim uspoređuje sa svim minimalnim vrijednostima udaljenosti. Ako za neko središte grozda vrijedi da je minimalna udaljenost veća od najmanje maksimalne udaljenosti, tada se taj grozd odbacuje kao kandidat za promatrani objekt. Na ovaj način nema potrebe računati ED. Ova metoda detaljno je objašnjena u drugom potpoglavlju. Metoda odbacivanja središta grozdova kao kandidata bez računanja ED pomoću Voronojevih dijagrama dijeli ukupni prostor skupa objekata na  $k$  Voronojevih ćelija, gdje je  $k$  jednak broju grozdova. Za svaki objekt, tj. njegovo minimalno granično područje neodređenosti,

provjerava se nalazi li se cijelom svojom površinom (2D) ili volumenom (3D) unutar jedne Voronojeve ćelije. Ukoliko se nalazi, onda se objekat pridružuje tom grozdu, a ostali grozdovi se odbacuju kao kandidati bez potrebe računanja ED-a. Ova metoda detaljno je objašnjena u trećem potpoglavlju. Odbacivanje grozdova pomoću SPP metode odbacuje jedan od dva grozda kao kandidata za promatrani objekat. Postupak se nastavlja za preostale grozdove sve dok na kraju ne budu odbačeni svi grozdovi osim jednoga. O ovoj metodi više se govori u četvrtom potpoglavlju.

### 3.1. Razvrstavanje objekata s prostornom neodređenosti pomoću *uk-means* metode

Objekti koji su prostorno neodređeni ne mogu se razvrstavati klasičnim algoritmima koji su opisanim u drugom poglavlju. Iz tog razloga bilo je potrebno izmijeniti već postojeće algoritme za razvrstavanje objekata koji su opisani jednom točkom u prostoru na način da se koriste za razvrstavanje objekata koji su opisani PDF-om. Tako je izmjenom *k-means* algoritma nastao *uk-means*. U [17] je predložen *k-means* u svrhu razvrstavanja  $n$  objekata s broječanim podacima u  $k$  grozdova predstavljenih pomoću središnje točke. Broj grozdova mora biti unaprijed poznat i ne mijenjati se tijekom iteracija. Objekat se dodjeljuje onom grozdu čije je središte najbliže. Nakon prve iteracije, kada se svi objekti pridodjele jednom od središta grozdova, postupak se nastavlja tako što se računaju nove pozicije središta grozdova kao srednja vrijednost svih objekata dodijeljenih tom grozdu. Postupak se nastavlja sve dok se ne zadovolji ciljna funkcija. Izmjenom *k-means* algoritma za razvrstavanje objekata s neodređenosti razvijen je *uncertain k-means* ili skraćeno *uk-means* algoritam [36]. Prva aplikacija algoritma bila je za potrebe praćenje pokretnih objekata koji imaju neodređenost u prostornim podacima [39], tzv MOD aplikacije (eng. Moving Objects Detection). To je dovelo do razvoja sustava za upravljanje bazama podataka (eng. Database Management System - DBMS), koji omogućavaju spremanje velikih količina informacija o objektima. Upit za lokaciju jednog od objekata nad takvom bazom podataka rezultira odgovorom njegove pretpostavljene lokacije u obliku  $(x, y, z)$  koordinate i područjem neodređenosti u kojem se objekat sigurno mora nalaziti. Neodređenost pozicije objekata smanjuje se učestalijim nadopunjavanjem baze podataka, što rezultira većom opterećenošću komunikacijskog kanala. Stoga se radi kompromis između neodređenosti i komunikacijskog prometa.



Bilo koja dva objekta s prostornom neodređenosti iz skup objekat  $O = \{o_1, o_2, \dots, o_n\}$  u prostoru  $\mathbb{R}^m$  s  $m$  dimenzija imaju udaljenost veću od nula, kako je prikazano u (3.2).

$$d(o_i, o_j) > 0 \quad (3.2)$$

Jednadžba (3.3) prikazuje da funkcija gustoće vjerojatnosti  $f_i(x)$ , tj. PDF neodređenog objekta u svakoj točki  $x$  mora biti veća od nule pa za sve točke unutar minimalnog područja neodređenosti vrijedi jednadžba (3.1).

$$f_i(x) > 0 \quad \forall x \in \mathbb{R}^m \quad (3.3)$$

PDF se modelira tako da se minimalno područje neodređenosti  $S_i$  podijeli na odgovarajući broj uzoraka. Svakom uzorku dodijeli se odgovarajuća vjerojatnost da se objekt nalazi u tom uzorku. Te se vjerojatnosti spremaju u bazu podataka. Za računanje ED objekta do središta grozda mora se izračunati udaljenost od svakog uzorka tog objekta do središta grozda. Udaljenosti se množe sa PDF-om i zbrajaju u integralu. Što je veći broj uzoraka za predstavljanje PDF-a to je veća preciznost razvrstavanja. Način računanja ED-a između objekta  $o_i$  i središta grozda  $c_j$  prikazanje jednadžbom (3.4).

$$ED(o_i, c_j) = \int_{x \in S_i} d(x, c_j) f_i(x) dx \quad (3.4)$$

Cilj razvrstavanja je pronaći sve relacije između objekata  $O = \{o_1, o_2, \dots, o_n\}$  i skupa grozdova  $C = \{c_1, c_2, \dots, c_k\}$ , da ukupna očekivana udaljenost (eng. Total Expected Distance - TED) između objekata i središta grozdova bude minimizirana.

$$TED = \sum_{i=1}^n ED(o_i, c_{j(i)}) \quad (3.5)$$

Pseudo kôd za *uk-means* razvrstavanje dan je algoritmom 3.1. Kako se vidi iz samog algoritma potrebno je za svaki objekt i svako središte grozda izračunati ED po formuli (3.4).

Algoritam 3.1: Pseudo kôd *uk-means* algoritma za razvrstavanje.

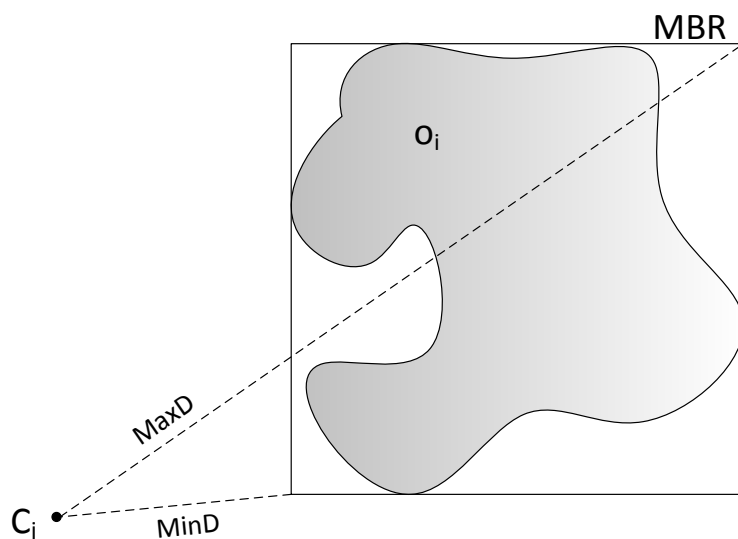
- 
1. **Ulaz:** početne pozicije  $k$  grozdova i skup objekata  $O$
  2. **do**
  3.     **for all**  $o_i \in O$
  4.         izračunati  $ED(o_i, c_j)$  prema (3.4)
  5.         pridružiti objekat grozdu s najmanjim  $ED$
  6.     **end for all**
  7.     **for all**  $c_j \in C$
  8.         izračunati nove pozicije središta grozdova
  9.     **end for all**
  10. **while** središta grozdova ne konvergiraju
-

Može se zaključiti da je razvrstavanje *uk-means* metodom vremenski neučinkovito zbog velikog broja uzoraka koji su potrebni za predstavljanje PDF-a.

### 3.2. MinMax metoda za odbacivanje središta grozdova

Zbog neučinkovitosti *uk-means* metode za razvrstavanje objekata s prostornom neodređenošću razvijene su različite metode za odbacivanje središta grozdova kao kandidata bez računanja ED. Zaključeno je da računanje ED-a vremenski vrlo zahtjevan proces zbog velikog broja uzoraka. Veliki broj uzoraka potreban je kako bi se što bolje opisao PDF. Kako bi se objekti s neodređenošću podjelili u grozdove upotrebom *uk-means* metode mora se izračunati ED od svakog objekata do svih središta grozdova. Ako je broj objekata jednak  $n$ , broj grozdova jednak  $k$  i broj iteracija po objektu jednak  $t$ , tada je broj računanja očekivanih udaljenosti u čitavom procesu razvrstavanja jednak  $n \cdot k \cdot t$ . Ideja je ako bi se smanjio ukupni broj računanja ED-a smanjilo bi se i vrijeme potrebno za razvrstavanje objekata s neodređenošću.

MinMax metoda predstavlja jedno od rješenja kako smanjiti ukupan broj računanja ED. Ovaj algoritam koristi činjenicu da se unutar minimalnog područja neodređenošću, predstavljenog s MBR-om u 2D i MBB-om u 3D, mora nalaziti lokacija objekta. Ako se uzmu u obzir samo krajnje točke MBR-a, kao što je prikazano na slici 3.3, moguće je umjesto očekivane udaljenosti ED koristiti euklidsku udaljenost koja je jednostavnija za računanje. Korištenjem ovakvog postupka razvrstavanja objekata mogu se neki grozdovi odbaciti kao mogući kandidati za pojedini objekt i to bez računanja ED [42].



Slika 3.3: Minimalna i maksimalna udaljenost između središta grozda  $c_j$  i objekt s prostornom neodređenošću  $o_i$  prikazanog pomoću minimalnog graničnog pravokutnika.

Postupak razvrstavanja počinje pronalaskom minimalne udaljenosti između objekta  $o_i$  i svih središta grozdova, kao što je prikazano formulom (3.6).

$$\text{MinD}(o_i, c_j) = \min_{x \in \text{MBR}_i} d(x, c_j) \quad (3.6)$$

Nakon toga potrebno je pronaći i maksimalne udaljenosti između tog objekta i svih središta grozdova, kao što je prikazano formulom (3.7).

$$\text{MaxD}(o_i, c_j) = \max_{x \in \text{MBR}_i} d(x, c_j) \quad (3.7)$$

Iz (3.6) i (3.7) jasno je da je minimalna udaljenost od objekta do središta grozda manja od ED te da je maksimalna udaljenost od objekta do središta grozda veća od ED.

$$\text{MinD}(o_i, c_j) \leq \text{ED}(o_i, c_j) \leq \text{MaxD}(o_i, c_j) \quad (3.8)$$

Ako je minimalna udaljenost od središta grozda  $c_p$  do objekta  $o_i$  veća ili jednaka od maksimalne udaljenost od središta grozda  $c_q$  do objekta  $o_i$ , tada se grozd  $c_p$  može odbaciti kao mogući kandidati za objekt  $o_i$ , kako je prikazano jednadžbom (3.9).

$$\text{MinD}(o_i, c_p) \geq \text{MaxD}(o_i, c_q) \xrightarrow{\text{proizlazi}} o_i \notin c_p \quad (3.9)$$

Jednostavniji način za odbacivanje grozdova je iz skupa svih maksimalnih udaljenosti između objekta  $o_i$  i svih središta grozdova pronaći najmanju maksimalnu udaljenost, kako je prikazano jednadžbom (3.10).

$$\text{MinMaxD}(o_i) = \min_{c_j \in C} \text{MaxD}(o_i, c_j) \quad (3.10)$$

Sada je potrebno usporediti sve izračunate minimalne udaljenosti  $\text{MinD}(o_i, C)$  s najmanjom maksimalnom udaljenošću  $\text{MinMaxD}(o_i)$ . Svi grozdovi koji zadovoljavaju jednadžbu (3.11) mogu se odbaciti kao mogući kandidati za objekt  $o_i$ .

$$\text{MinD}(o_i, c_p) \geq \text{MinMaxD}(o_i) \quad (3.11)$$

Grozdovi koji nisu odbačeni ovim postupkom trebaju se razvrstati korištenjem *uk-means* algoritma, tj. potrebno je izračunati njihovu ED do objekta i objekt dodijeliti onom grozdu koji ima najmanju ED. Postupak treba ponoviti za sve ostale objekte iz skupa objekata s neodređenošću  $O = \{o_1, o_2, \dots, o_n\}$ . Pseudo kôd za odbacivanje središta grozdova kao kandidata pomoću MinMax metode prikazan je na algoritmu 3.2.

MinMax metoda, osim navedenog poboljšanja, nudi još neka rješenja u odnosu na *uk-means* algoritam. Nakon što su neka središta grozdova odbačena po jednadžbi (3.11), ostala se

Algoritam 3.2: Pseudo kôd MinMax algoritma za razvrstavanje i odbacivanje središta grozdova kao kandidata.

---

```

1. Ulaz: skup objekata  $O$  i početne pozicije grozdova  $C$ 
2. for all  $c_j \in C$  i objekt  $o_i$ 
3.   izračunati  $MinD(o_i, c_j)$  i  $MaxD(o_i, c_j)$  prema (3.6) i (3.7)
4.   izračunati  $MinMaxD(o_i, C)$  prema (3.10)
5.   for all  $c_j \in C$ 
6.     if  $MinD(o_i, c_j) \geq MinMaxD(o_i, C)$ 
7.       ukloni  $c_j$  iz skupa mogućih kandidata za objekt  $o_i$ 
8.     end if
9.   end for all
10. end for all
11. for all ostale grozdove
12.   izračunati  $ED$  prema (3.4)
13. end for all

```

---

središta mogu odbaciti po jednadžbi (3.12). Svaki put kad se izračuna neka  $ED(o_i, c_p)$  ona se uspoređuje s do tada najmanjom vrijednošću očekivane udaljenosti  $MinED(o_i, c_j)$  i ako je minimalna udaljenost nekog grozda  $c_p$  veća od najmanje očekivane udaljenosti taj se grozd odbacuje. U slučaju da je izračunata  $ED(o_i, c_p)$  manja od  $MinED(o_i, c_j)$ ,  $ED(o_i, c_p)$  se proglašava minimalnom očekivanom udaljenošću  $MinED(o_i, c_p)$  dok se dotadašnji grozd  $c_p$  odbacuje kao kandidat za objekt  $o_i$ .

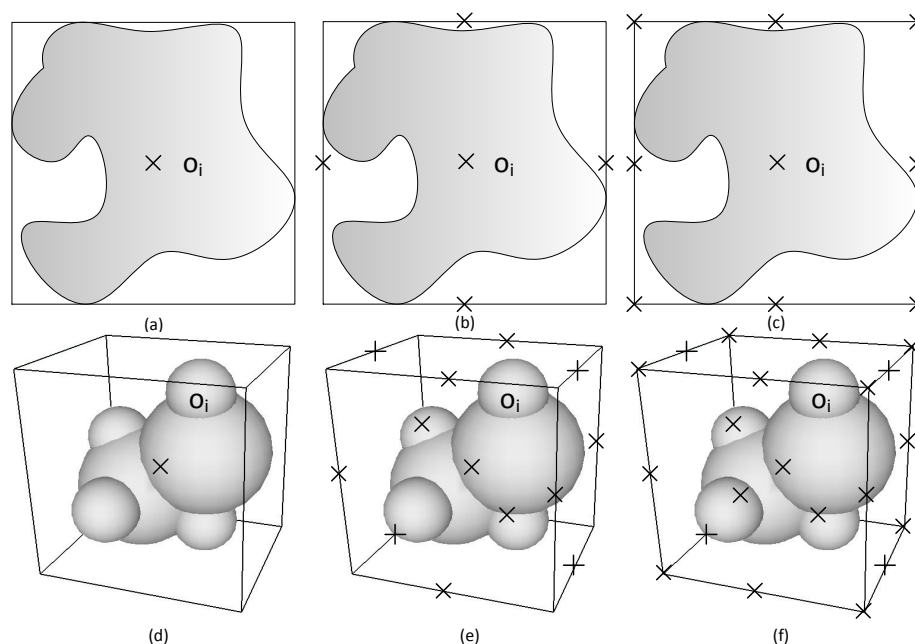
$$ED(o_i, c_p) > MinED(o_i, c_j) \quad (3.12)$$

Na taj način dodatno je smanjen broj računanja ED.

Učinkovitost MinMax algoritma vidljiva je u jednostavnom odbacivanju grozdova koji se nalaze daleko od najbližih grozdova objekta  $o_i$ , bez potrebe računanja ED-a. Učinkovitost algoritma, tj. broj odbačenih grozdova, ovisi o veličini minimalnog područja neodređenosti. Ako je područje neodređenosti relativno malo, u omjeru s udaljenosti između središta grozdova, većina će grozdova biti odbačena bez računanja ED-a, no ako je područje neodređenosti veliko (u bilo kojem smjeru) tada će najmanja maksimalna udaljenost biti velika i rezultat odbacivanja grozdova neće biti toliko dobar. Može se zaključiti da je uspješnost MinMax algoritma ovisna o veličini područja neodređenosti i ako je područje neodređenosti malo algoritam će imati više odbačenih grozdova te time i kraće vrijeme razvrstavanja objekata s neodređenosti u grozdove.

Kako bi se poboljšala učinkovitost MinMax algoritma kod objekata s velikim područjima neodređenosti razvijene su dodatne metode. U [47] predložena je metoda nejednakosti trokuta, koja se koristi za određivanje najbolje gornje i donje granice udaljenosti između dviju točaka. Kako bi se ova metoda mogla primijeniti potrebno je pronaći sidrišne

točke. U praktičnoj primjeni treba dobro procijeniti koliko sidrišnih točaka se može uzeti, a da se i dalje dobije ušteda na vremenu izvođenja. Sam proces procjene najbolje gornje i donje granice udaljenosti za rezultat ima povećanje vremena izvođenja. Ako se koristi  $r$  sidrišnih točaka u skupu objekata s neodređenosti koji ima  $n$  članova tada je potrebno izračunati  $n \cdot r$  predračunskih ED-a. Iz toga se lako zaključuje da će metoda procjene biti korisna samo u onome slučaju kad je broj grozdova koji će biti odbačeni u svim iteracijama pomoću ove metode biti veći od  $n \cdot r$ . Zato je vrlo bitno odabrati ispravan broj sidrišnih točaka kako bi ovaj postupak bio isplativ i rezultirao skraćivanjem vremena izvođenja. Ne postoji jednostavan način pomoću kojeg se može izračunati koja točka minimizira ED. U praksi se uzimaju različite točke unutar područja neodređenosti objekta  $o_i$ . Najboljim se pokazalo odabrati točke na različitim stranama MBR-a i MBB-a. Ovisno o tome koliko se sidrišnih točaka koristi postoje različiti modeli odabira točaka. Ako se uzima samo jedna sidrišna točka tada je najbolje odabrati središte MBR/MBB-a. U modelu s više točaka uzima se središte i mogu se uzeti središta svih bridova i vrhova koji određuje MBR/MBB, što osigurava da je barem jedna od točaka bliže reprezentativnoj točki objekta, nego neka točka izvan MBR/MBB-a, kao što je prikazano na slici 3.4. Primjena navedene metode za procjenu granica najpogodnija je u slučajevima kada dio objekata ne mijenja svoju poziciju često. Za takve objekte nije potreban stalan predračun ED-a što dodatno smanjuje vrijeme izvođenja. U [42] predstavljeni su rezultati pokusa koji pokazuju učinkovitost pojedinih granica i broj sidrišnih točaka.



Slika 3.4: Objekt s prostornom neodređenosti  $o_i$  prikazan pomoću modela sa: (a) jednom, (b) pet, (c) devet (d) jednom, (e) trinaest i (f) dvadestet jednom sidrišnom točkom.

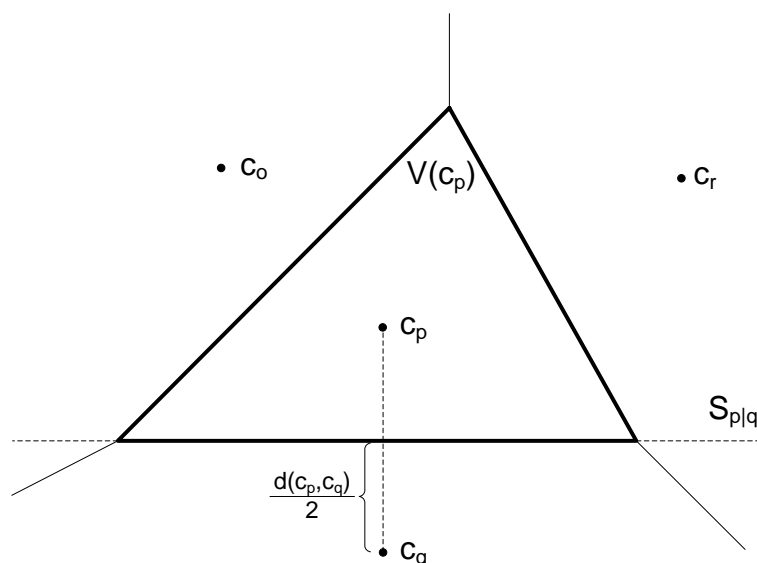
### 3.3. Metoda za odbacivanje središta grozdova pomoću Voronojevih dijagrama

U dosadašnjim poglavljima objašnjene su *uk-means* i MinMax metoda za razvrstavanje objekata s prostornom nesigurnosti. Pokazano je također da MinMax metoda značajno povećava učinkovitost razvrstavanja u odnosu na *uk-means* metodu, tako što koristi minimalno područje neodređenosti za određivanje granica ED. Na taj način može se izbjeći računanja većine ED-ova koji inače produžavaju proces razvrstavanja. U [43] dan je opis metode za razvrstavanje objekata s prostornom neodređenosti korištenjem Voronojevih dijagrama. Voronojevi dijagrami koriste prednosti geometrijske strukture prostora  $\mathbb{R}^m$  te na taj način odbacuju veći broj grozdova od MinMax metode.

Ideja ovog načina razvrstavanje je podijeliti geometrijski prostor  $\mathbb{R}^m$ , koristeći skup grozdova  $C = \{c_1, c_2, \dots, c_k\}$ , na  $k$  Voronojevih dijagrama. Ti Voronojevi dijagrami ili drugim nazivom Voronojeve ćelije označavaju se s  $V(c_j)$  te imaju svojstvo da bilo koja točka  $x$  koja se nalazi u ćeliji  $V(c_p)$  ima manju udaljenost do središta ćelije  $c_p$ , nego do središta bilo koje druge ćelije u tom prostoru  $\mathbb{R}^m$ , kako je opisano jednadžbom (3.13)

$$d(x, c_p) < d(x, c_q) \quad \forall x \in V(c_p), c_p \neq c_q \quad (3.13)$$

U geometrijskom prostoru s  $\mathbb{R}^2$  granica između dvije susjedne ćelije  $V(c_p)$  i  $V(c_q)$  je dio pravca simetrale koja prolazi polovinom dužine, koja spaja središta grozdova  $c_p$  i  $c_q$ , te je i okomit na tu dužinu, označava se sa  $S_{p|q}$ , kao što je prikazano na slici 3.5.



Slika 3.5: Konstrukcija  $V(c_p)$  ćelije u  $\mathbb{R}^2$  geometrijskom prostoru.

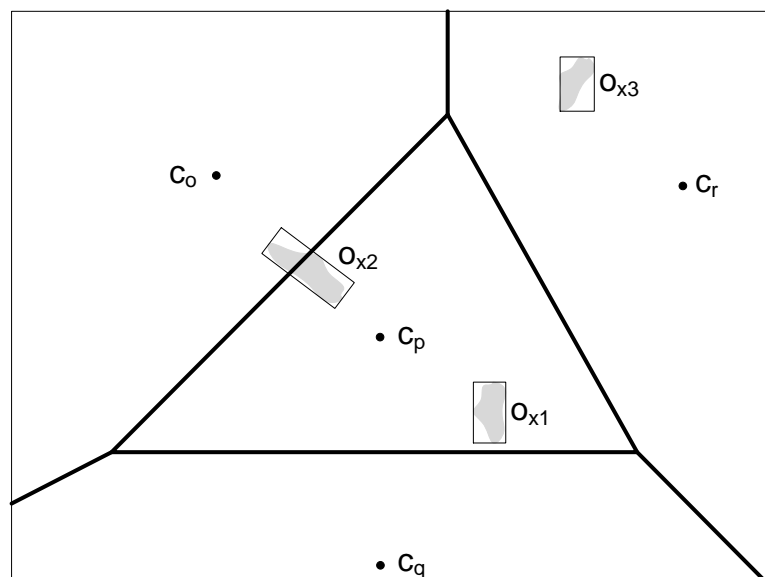
Prilikom konstrukcije simetrala  $S_{p|q}$  u geometrijskom prostoru s dvije dimenzije  $\mathbb{R}^2$  nastaju dvije poluravnine. Na strani gdje se nalazi središte grozda  $c_p$  nastaje poluravnina  $HP_{p|q}$ , a na strani gdje se nalazi središte grozda  $c_q$  nastaje poluravnina  $HP_{q|p}$ .

Svojstvo prikazano jednadžbom (3.13) može se primjeniti za pojedini objekt  $o_i$  tako da se provjeri leži li u cjelini njegovo područje neodređenosti unutar neke od Voronojevih ćelija. Ako je zadovoljen uvjet prikazan formulom (3.14) onda se objekt  $o_i$  dodijeljuje grozdu  $c_p$ .

$$ED(o_i, c_p) < ED(o_i, c_q) \quad \forall c_q \in C \setminus \{c_p\} \quad (3.14)$$

U ovom slučaju nema potrebe provesti niti jedan izračun ED-a. Sva središta grozdova osim  $c_p$  su odbačena kao mogući kandidati za objekt  $o_i$ . Na slici 3.6 prikazana su tri objekta s prostornom neodređenošću u 2D prostoru. Svaki je prikazan sa svojim MBR-om. Vidljivo je da se objekt  $o_{x1}$ , tj. njegov MBR u potpunosti nalazi unutar ćelije  $V(c_p)$ , pa je zato objekt  $o_{x1}$  dodijeljen grozdu  $c_p$ . Na slici 3.6 nalaze se još objekti  $o_{x2}$  i  $o_{x3}$ . Objekt  $o_{x2}$  nalazi se na samoj granici između dvije Voronojeve ćelije ( $V(c_p)$  i  $V(c_o)$ ), dok se objekat  $o_{x3}$  nalazi posve izvan  $V(c_p)$  ćelije. Iz tog razloga oba se objekta ne mogu dodijeliti grozdu  $c_p$ . Objekt  $o_{x3}$  će biti dodijeljen u nekoj od sljedećih iteracija grozdu  $c_r$ , no objekt  $o_{x2}$  će ostati nedodijeljen i za njegovo razvrstavanje bit će potrebno izvršiti izračun ED za svaki od njegovih uzoraka.

Pseudo kôd za odbacivanje središta grozdova kao kandidata pomoću metode Voronojevih dijagrama prikazan je algoritmom 3.3. Linijama 2 do 7 prikazan je dio kôda kojim se objekti  $o_{x1}$  i  $o_{x3}$  dodijeljuju grozdovima  $c_p$  i  $c_r$ . Dok je linijama 8 do 15 prikazan način razvrstavanje objekta  $o_{x2}$ .



Slika 3.6: Prikaz odbacivanja središta grozdova pomoću Voronojevih dijagrama 2D prostoru

Algoritam 3.3: Pseudo kôd metode Voronojevih dijagrama za razvrstavanje i odbacivanje središta grozdova kao kandidata.

---

```

1. Ulaz: skup objekata  $O$  i početne pozicije grozdova  $C$ 
2. izračunaj Voronojeve dijagrame za  $C = \{c_1, c_2, \dots, c_k\}$ 
3. for all  $c_j \in C$  i objekt  $o_i$ 
4.     if  $MBR_i \subseteq V(c_j)$ 
5.         objekt  $o_i$  dodjeli grozdu  $c_j$ 
6.     end if
7. end for all
8. for all nedodjeljene  $o_i$  i parove grozdova  $c_p$  i  $c_q$ 
9.     if  $MBR_i \subseteq H_{p|q}$ 
10.        ukloni  $c_q$  iz skupa mogući kandidati za objekt  $o_i$ 
11.    end if
12. end for all
13. for all ostale nedodjeljene objekte i grozdove
14.    izračunati  $ED$  prema (3.4)
15. end for all

```

---

U algoritmu 3.3 od 8. do 11. linije dan je pseudo kôd za odbacivanje grozdova pomoću simetrala, koje nastaju prilikom konstrukcije Voronojevih ćelija. Računanje tih simetrala ne zahtijeva puno dodatnog vremena u razvrstavanju, a korisne su jer dijele prostor na dvije poluravnine. Tako se za svaki objekt  $o_i$  provjerava nalazi li se njegov MBR u cijelosti unutar poluravnine  $HP_{p|q}$ . Ako je taj uvjet zadovoljen, grozd  $c_q$  se odbacuje kao kandidat za objekt  $o_i$ . Ovim postupkom može se odbaciti određeni dio grozdova kao potencijalni kandidati za objekt  $o_i$  i time dodatno smanjiti broj računanja ED-a.

U [43] dan je i teorem koji dokazuje da se korištenjem Voronojevih dijagrama može odbaciti više grozdova kao potencijalnih kandidata, nego korištenjem MinMax metode.

*Teorem 1:* Za neki objekt  $o_i \in O$  i grozd  $c_j \in C$ , vrijedi da ako algoritam Voronojevih dijagrama ne odbaci grozd  $c_j$  kao kandidata za objekt  $o_i$  tada ga neće odbaciti niti MinMax algoritam.

Kako bi se dokazao ovaj teorem koristi se grozd  $c_p$  koji ima najmanju maksimalnu udaljenost do objekta  $o_i$ , tj.  $MaxD(o_i, c_p) = MinMaxD(o_i)$ . Sada se mogu razmotriti dva slučaja. Prvi slučaj je kada je  $c_p = c_j$ . Za ovaj slučaj vrijede jednakosti u jednadžbi (3.15).

$$\begin{aligned}
 MinD(o_i, c_j) &\leq MaxD(o_i, c_j) \\
 MaxD(o_i, c_j) &= MaxD(o_i, c_p) \\
 MaxD(o_i, c_p) &= MinMaxD(o_i)
 \end{aligned} \tag{3.15}$$

MinMax algoritam neće odbaciti grozd  $c_j$  kao kandidata za objek  $o_i$  jer nije zadovoljen uvjet  $MinD(o_i, c_j) \geq MinMax(o_i)$  iz čega slijedi da je teorem zadovoljen.

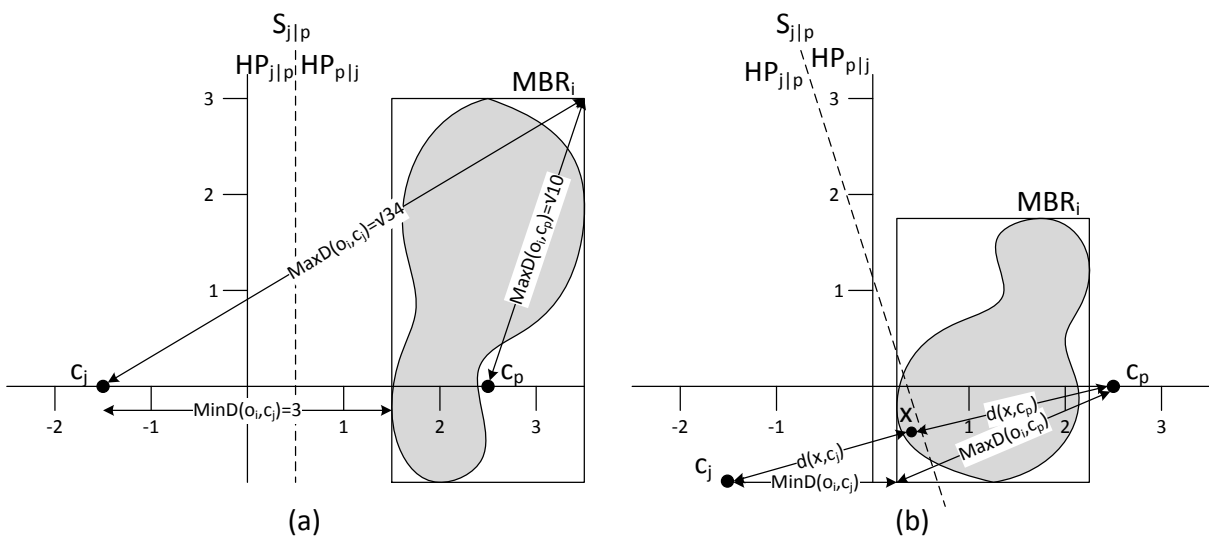


U drugom slučaju kada vrijedi  $c_p \neq c_j$  i simetrala  $S_{p|j}$  dijeli prostor  $\mathbb{R}^2$  na dva dijela  $HP_{p|j}$  i  $HP_{j|p}$  mogu postojati dva nova podslučaja. U prvom podslučaju MBR objekta  $o_i$  nalazi se u cjelini  $HP_{p|j}$  ravnini. U ovom podslučaju grozd  $c_j$  će biti odbačen linijom 10 u algoritmu 3.3, što nije slučaj za algoritam MinMax pa je teorem i dalje zadovoljen. Sa slike 3.7(a) vidi se da je minimalna udaljenost objekta  $o_i$  i središta grozda  $c_j$  manja od najmanje maksimalne udaljenosti koja je jednaka maksimalnoj udaljenosti objekta  $o_i$  i središta grozda  $c_p$  prikazano ovom nejednakosti  $MinD(o_i, c_j) = 3 < MinMax(o_i) = MaxD(o_i, c_p) = \sqrt{10}$ . U drugom podslučaju MBR objekta  $o_i$  nalazi se dijelom u  $HP_{p|j}$  ravnini, a dijelom u  $HP_{j|p}$  ravnini. Ako se promatra točka  $x$ , koja se nalazi u dijelu  $MBR_i$  koji leži u  $HP_{j|p}$  ravnini i prikazana je na slici 3.7(b), onda za nju vrijede jednakosti u jednadžbi (3.16).

$$\begin{aligned}
 MinD(o_i, c_j) &\leq d(x, c_j) \\
 d(x, c_j) &\leq d(x, c_p) \\
 d(x, c_p) &\leq MaxD(o_i, c_p) \\
 MaxD(o_i, c_p) &= MinMaxD(o_i)
 \end{aligned} \tag{3.16}$$

Iz jednadžbe (3.16) vidljivo je da uvjeti za MinMax algoritam nisu zadovoljeni te je teorem i u ovom slučaju valjan.

S navedenim primjerima dokazan je Teorem 1 i uz to je pokazano da se algoritmom Voronojevih dijagrama ostvaruje veći broj odbacenih grozdova, nego što je to u slučaju korištenja MinMax algoritma.



Slika 3.7: Prikaz uspješnog odbacivanja središta grozda  $c_j$  pomoću algoritma Voronojevih dijagrama za (a) prvi podslučaj i (b) drugi podslučaj kada su  $c_p \neq c_j$ .

Za sve ostale grozdove koji nisu odbačeni do 13 reda u algoritmu 3.3 mora se računati ED. Kako bi se smanjilo potrebno vrijeme za računanje ED-a može se koristiti djelomično računanje očekivane udaljenosti. Na taj način nije potrebno računati čitav integral iz jednadžbe (3.4), već samo njegov dio. Ideja je, ako se promatraju dva grozda  $c_p$  i  $c_j$  te objekt  $o_i$  čiji MBR siječe simetralu  $S_{p|j}$ , tada se MBR tog objekta dijeli na dva dijela. Dio X i dio Y za koje vrijede svojstva opisana jednadžbom (3.17) i prikazano slikom 3.8.

$$X \cup Y = MBR_i; X \cap Y = 0; X \in V(c_p) \quad (3.17)$$

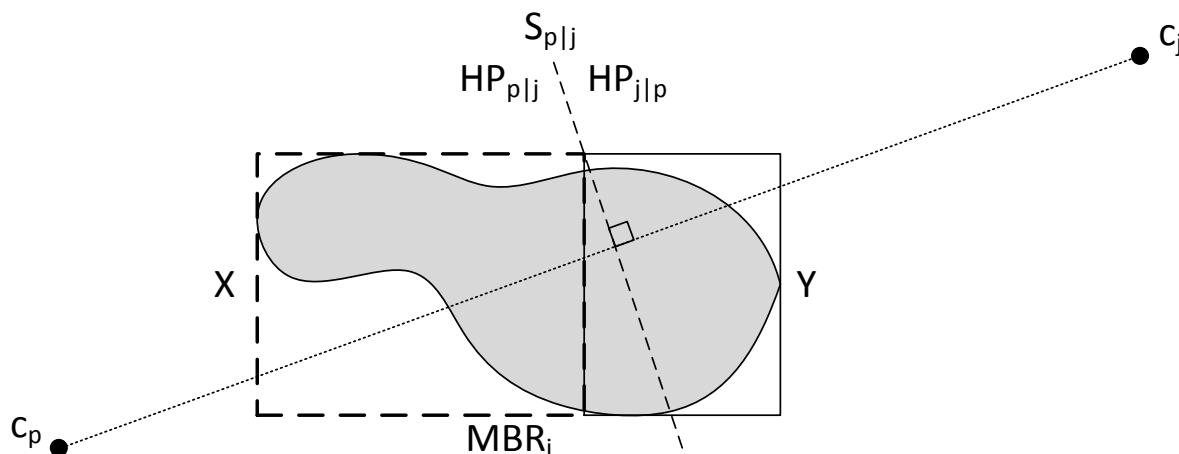
Na ovaj način računanje ukupne ED može se predstaviti sumom dva manja integrala, kako je prikazano formulom (3.18).

$$\begin{aligned} ED(o_i, c_p) &= \int_{x \in MBR_i} d(x, c_p) f_i(x) dx \\ &= \int_{x \in X} d(x, c_p) f_i(x) dx + \int_{x \in Y} d(x, c_p) f_i(x) dx \\ &= ED_X(o_i, c_p) + ED_Y(o_i, c_p) \end{aligned} \quad (3.18)$$

Jednadžba (3.18) može se primijeniti i za računanje ED za središte grozda  $c_j$ .

$$ED(o_i, c_j) = ED_X(o_i, c_j) + ED_Y(o_i, c_j) \quad (3.19)$$

Koristeći činjenicu iz jednadžbe (3.17) da je  $X \in V(c_p)$  slijedi da je ED za dio X  $ED_X(o_i, c_p) < ED_X(o_i, c_j)$ . Nakon toga potrebno je samo izračunati vrijednosti integrala  $ED_Y(o_i, c_p)$  i  $ED_Y(o_i, c_j)$  te ako vrijedi da je  $ED_Y(o_i, c_p) < ED_Y(o_i, c_j)$  onda se može zaključiti da je i  $ED(o_i, c_p) < ED(o_i, c_j)$ . Tada se grozd  $c_j$  može odbaciti kao kandidat za



Slika 3.8: Podjela područja neodređenosti na dva dijela.

objekt  $o_i$  bez računanja  $ED_X(o_i, c_j)$ . U suprotnom potrebno je izračunati ED za grozd  $c_j$ . Prednost je u tome što se samo mora izračunati  $ED_X(o_i, c_j)$  i zbrojiti s već ranije izračunatom  $ED_Y(o_i, c_j)$  kako bi se dobila ukupna očekivana udaljenost  $ED(o_i, c_j)$ .

### 3.4. SPP metoda za odbacivanje središta grozdova

U prethodnom poglavlju objašnjena je metoda Voronojevih dijagrama za razvrstavanje objekata s prostornom nesigurnosti. Pokazane su prednosti te metode na MinMax i *uk-means* metodama. Voronojevi dijagrami koriste prednosti geometrijske strukture prostora  $\mathbb{R}^m$  te na taj način odbacuju veći broj grozdova od MinMax metode. Također, objašnjen je algoritam Voronojevih dijagrama koji za svaki objekt provjerava leži li njegovo područje neodređenosti u potpunosti unutar neke od Voronojevih ćelija. To je vremenski vrlo zahtjevna operacija jer se prvo moraju konstruirati Voronojeve ćelije i provjeravati nalazi li se MBR unutar te ćelije.

U [46] predstavljena je ideja odbacivanja središta grozdova pomoću simetrale podjele prostora. Metoda je nazvana SPP i zasniva se na podjeli prostora između dvaju grozdova te odbacivanju jednog od grozdova kao kandidat za promatrani objekt. Za svaka dva središta grozdova konstruira se pravac koja prolazi polovicom dužine između dva središta grozda i okomita je na tu dužinu. Taj se pravac naziva simetrala i označava s oznakama oba grozda. Tako između grozdova  $c_p$  i  $c_q$  postoji simetrala  $S_{p|q}$  koja dijeli prostor  $\mathbb{R}^2$  na dvije poluravnine  $HP_{p|q}$  (na stani groza  $c_p$ ) i  $HP_{q|p}$  (na stani groza  $c_q$ ). Na temelju ovih činjenica mogu se definirati sljedeća pravila prikazana u jednadžbama (3.20) i (3.21).

$$d(x, c_p) < d(x, c_q) \quad \forall x \in HP_{p|q} \quad (3.20)$$

$$d(x, c_p) = d(x, c_q) \quad \forall x \in S_{p|q} \quad (3.21)$$

Jednadžba (3.20) opisuje svojstvo točke  $x$  koja se nalazi u poluravnini  $HP_{p|q}$ . Udaljenost takve točke  $x$  do središta grozda  $c_p$  manja je nego do središta grozda  $c_q$ . Vrijedi i obratno. Jednadžba (3.21) opisuje slučaj kada se točka  $x$  nalazi na samoj simetrali  $S_{p|q}$ . Za ovaj primjer vrijedi da je udaljenost od točke  $x$  do središta grozdova  $c_p$  i  $c_q$  jednaka.

SPP metoda provjerava za svaki objekt  $o_i$  leži li njegov MBR u cijelosti u poluravnini  $HP_{p|q}$ . Ako je taj uvjet ispunjen, grozd  $c_q$  odbacuje se kao kandidat za objekt  $o_i$ , a grozd  $c_p$  uspoređuje se s preostalim grozdovima. Prednost SPP metode nad metodom Voronojevih dijagrama nalazi se u činjenici da nije potrebno konstruirati Voronojeve ćelije i što se

odbačeni grozd  $c_q$  ne uspoređuje s preostalim grozdovima. Iz tog razloga broj usporedbi između grozdova znatno se smanjuje, a time se smanjuje i vrijeme potrebno za razvrstavanje. Nakon što se obave sve usporedbe, u najboljem slučaju, ostati će samo jedan grozd ili će većina grozdova biti odbačena kao potencijalni kandidati za objekt  $o_i$ .

U prethodnom potpoglavlju dokazano je da se upotrebom metode Voronojevih dijagrama odbacuje veći broj grozdova nego MinMax metodom. Broj odbačenih grozdova upotrebom metode Voronojevih dijagrama jednak je broju odbačenih grozdova upotrebom SPP metode. To je iz razloga što obje metode odbacuju grozdove po sličnom principu. Voronojeve ćelije konstruiraju se pomoću simetrala, a SPP metoda na osnovu simetrala dijeli prostor u poluravnine. Iz toga se može zaključiti da će SPP metoda odbacivati više grozdova od MinMax metode, što je i dokazano pokusima u [48]. Uz sve navedeno SPP metoda predstavlja način brze konstrukcije simetrala te tako ubrzava proces odbacivanja grozdova dobiven korištenjem metode Voronojevih dijagrama.

Jednadžba pravca simetrale računa se po jednadžbi (3.22). Za svaki par središta grozdova  $c_p$  i  $c_q$  iz skupa grozdova  $C = \{c_1, c_2, \dots, c_k\}$  koji su zadani koordinatnima točkama  $c_p = (x_{cp}, y_{cp})$  i  $c_q = (x_{cq}, y_{cq})$  u geometrijskom prostoru  $\mathbb{R}^2$  računa se simetrala  $S_{p|q}$ .

$$S_{p|q} = ax + b \tag{3.22}$$

$$a = -\frac{x_{cp} - x_{cq}}{y_{cp} - y_{cq}} \quad b = \frac{x_{cp}^2 - x_{cq}^2 + y_{cp}^2 - y_{cq}^2}{2(y_{cp} - y_{cq})}$$

Računanje simetrale na ovaj način brže je nego kod metode Voronojevih dijagrama. U metodi Voronojevih dijagrama prvo se računaju ćelije pomoću *qhull* algoritma, a onda se iz stranica ćelija mogu izračunati simetrale. Osim bržeg računanja simetrala SPP metoda ima brži proces odbacivanja grozdova jer provjeravanje s koje strane simetrale se nalazi MBR objekta  $o_i$  brže je nego provjeravanje nalazi li se MBR objekta  $o_i$  unutar Voronojeve ćelije, što je također dokazano u [48].

Nakon što je izračunata jednadžba simetrale potrebno je utvrditi s koje strane simetrale se nalazi MBR objekta  $o_i$ . Ako se MBR objekta  $o_i$  nalazi s iste strane simetrale kao središte grozda  $c_p$ , onda se grozd  $c_q$  odbacuje kao kandidat za objekt  $o_i$ . Također vrijedi i obrnuto. U slučaju kada MBR objekta  $o_i$  siječe simetralu ne može se odbaciti niti jedan od grozdova, nego se nastavljaju uspoređivati s preostalim grozdovima kandidatima za objekt  $o_i$ . Na kraju svih usporedbi, ako su odbačeni svi grozdovi osim jedanog grozda onda je proces

razvrstavanja završen, a ako je ostalo više kandidata za njih je potrebno izračunati ED do objekta  $o_i$  i grozdu s najmanjom ED dodijeliti objekt.

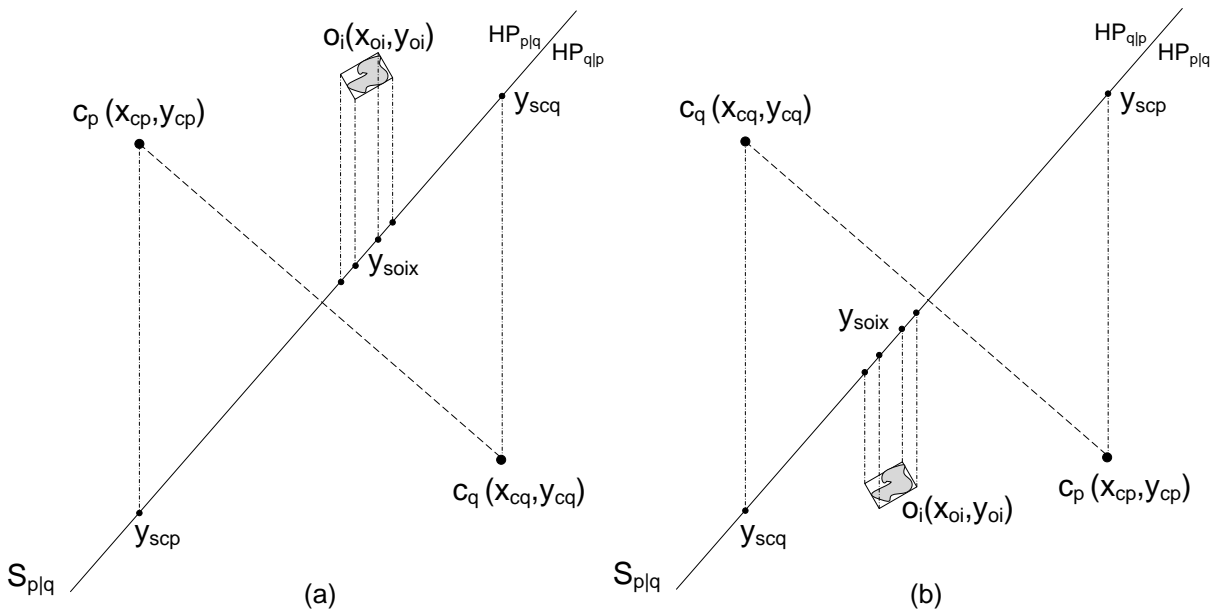
Postupak odbacivanje grozda kao kandidata za objekt  $o_i$  započinje proiciranjem koordinata središta grozdova  $c_p$  i  $c_q$  te proiciranjem  $MBR_i$  na simetralu  $S_{p|q}$ . Za svaki MBR potrebno je projicirati četiri koordinate vrhova. Na taj način osigurava se provjera siječe li simetrala MBR objekta  $o_i$ . Na slici 3.9 prikazana je konstrukcija simetrale te projiciranje koordinata središta grozdova i MBR-a.

Uvjeti za uspješno odbacivanje grozda  $c_q$  prikazani su jednadžbom (3.23).

$$(y_{scp} < y_{cp} \text{ and } y_{soix} < y_{oix}) \vee (y_{scp} > y_{cp} \text{ and } y_{soix} > y_{oix}) \quad (3.23)$$

$$\forall x \in MBR_i = \{1,2,3,4\}$$

Iz gornje jednadžbe vidljivo je da se uspoređuju samo y koordinate objekta i središta grozdova s njihovim projekcijama. Tako je s  $y_{cp}$  označena y koordinata središta grozda  $c_p$ , a s  $y_{scp}$  njezina projekcija na simetrali  $S_{p|q}$ . Sve y koordinate četiri vrha MBR-a objekta  $o_i$  označene su s  $y_{oix}$  gdje x može biti 1 do 4 pa su tako i njihove projekcije na simetrali  $S_{p|q}$  označene sa  $y_{soix}$  gdje x odgovara broju vrha. Ako je ispunjen uvjet, koji je dan jednadžbom (3.23), onda je grozd  $c_q$  uspješno uklonjen kao kandidat za objekt  $o_i$ . Postoje dva slučaja kada će uvjet biti ispunjen. U prvom slučaju MBR objekta  $o_i$  se u cijelost nalazi u poluravnini  $HP_{p|q}$  koja se nalazi iznad simetrale  $S_{p|q}$ , kao što je prikazano na slici 3.9(a). Kako bi se izračunala koordinata projekcije  $y_{scp}$  u jednadžbu pravaca simetrale uvrštava se x koordinata



Slika 3.9: Podjela  $\mathbb{R}^2$  prostora simetralama između središta grozdova  $c_p$  i  $c_q$  gdje se objekt  $o_i$  nalazi (a) iznad i (b) ispod simetrale.

grozda  $c_p$ . Isto tako se za sva četiri vrha MBR-a uvrštava x koordinata objekta  $o_i$  u jednadžbu za pravac simetrale i kao rezultat dobiju se koordinate  $y_{soi1}$ ,  $y_{soi2}$ ,  $y_{soi3}$  i  $y_{soi4}$ . Nakon izračuna projekcije grozdova i objekta na simetralu provjerava se je li zadovoljen uvjet za odbacivanje grozdova prema jednadžbi (3.23). U ovom slučaju prvi je uvjet zadovoljen jer se vidi da je  $y_{cp}$  veće od  $y_{scp}$ , kao što su i y koordinate MBR-a objekta  $o_i$  ( $y_{oi1}$ ,  $y_{oi2}$ ,  $y_{oi3}$  i  $y_{oi4}$ .) veće od njihovih projekcija na simetrali ( $y_{soi1}$ ,  $y_{soi2}$ ,  $y_{soi3}$  i  $y_{soi4}$ ). Zato se može u ovom slučaju sa sigurnošću reći da objekt  $o_i$  ne pripada grozdu  $c_q$  te se on može odbaciti kao kandidat. U drugom slučaju MBR objekta  $o_i$  se u cijelost nalazi u poluravnini  $HP_{p|q}$  koja se nalazi ispod simetrale  $S_{p|q}$ , kao što je prikazano na slici 3.9(b). Koordinate projekcija računaju se na isti način kao i u prvom slučaju i uspoređuju s koordinatama objekta  $o_i$  i središta grozda  $c_p$  prema jednadžbi (3.23). U ovom slučaju zadovoljen je drugi uvjet iz jednadžbe i grozd  $c_q$  može se odbaciti kao kandidat.

Nakon što je grozd  $c_q$  odbačen kao mogući grozd za objekt  $o_i$ , potrebno je sve ostale grozdove usporediti na isti način s grozdom  $c_p$ . Cilj je da na kraju procesa razvrstavanja objekta  $o_i$  SPP metodom budu odbačeni svi grozdovi osim jednoga. Tada će se objekt  $o_i$  dodijeliti tom grozdu i proces razvrstavanja je završen. No postoje situacije kada simetrala između dva središta grozda siječe MBR objekta koji se razvrstava. U tom slučaju nije moguće odbaciti niti jedan od promatranih grozdova te je za ispravno razvrstavanje potrebno izračunati ED za promatrane grozdove i objekt dodijeliti grozdu s najmanjim ED-om.

Pseudo kôd za razvrstavanje i odbacivanje središta grozdova kao kandidata SPP metodom prikazan je algoritmom 3.4.

Dodatna prednost SPP metode vidi se ako se promotre linije 6 i 9 u algoritmu 3.4. U

Algoritam 3.4: Pseudo kôd SPP metode za razvrstavanje i odbacivanje središta grozdova kao kandidata.

- 
1. **Ulaz:** skup objekata  $O$  i početne pozicije grozdova  $C$
  2. izračunaj simetrale za  $C = \{c_1, c_2, \dots, c_k\}$
  3. **for all**  $c_p \in C$  i  $c_q \in C$ ;  $c_p \neq c_q$
  4.     **if**  $MBR_i \subseteq HP_{p|q}$  prema (3.23)
  5.         ukloni  $c_q$  kao kandidata za  $o_i$
  6.         ukloni  $c_q$  iz for petlje
  7.     **else**  $MBR_i \subseteq HP_{q|p}$  prema (3.23)
  8.         ukloni  $c_p$  kao kandidata za  $o_i$
  9.         ukloni  $c_p$  iz for petlje
  10.     **end if**
  11. **end for all**
  12. **for all** ostale nedodjeljene objekte i grozdove
  13.     izračunati ED prema (3.4)
  14. **end for all**
-

slučaju kada se ukloni kao kandidat jedan od grozdova za objekt  $o_i$ , on se izbacuje iz programske petlje i više ne uspoređuje s preostalim grozdovima iz skupa  $C = \{c_1, c_2, \dots, c_k\}$ . U promatranom primjeru grozd  $c_q$  odbaćen je kao kandidat u usporedbi s grozdom  $c_p$ , stoga nema potrebe za usporedbom između grozda  $c_q$  i preostalih, do tada neodbačenih, grozdova. Budući da je grozd  $c_p$  dokazano bliži objektu  $o_i$  jednadžbom (3.23), on će odbaciti sve one grozdove koje bi odbacio i grozd  $c_q$  u kasnijim usporedbama. Iz tog razloga rade se samo usporedbe između grozda  $c_p$  i preostalih grozdova, sve do trenutka kad, eventualno, i on bude odbaćen. Broj izbjegnutih usporedbi  $N_i$  ovisan je o broju uspješno odbaćenih grozdova  $N_o$ , i računa se po jedanažbi (3.24), gdje je  $n$  ukupni broj objekata, a  $k$  ukupan broj grozdova.

$$N_i = \sum_{i=0}^n N_o(k-1) \quad (3.24)$$

Ovom akcijom uklanjanja grozdova iz petlje značajno je smanjen broj usporedbi koje se trebaju napraviti te je vremenski skraćen proces odbacivanja grozdova.

### 3.5. Usporedba postojećih metoda

U ovom poglavlju raspravljalo se o metodama razvrstavanja objekata s prostornom neodređenosti. Nakon objašnjenja kako dolazi do prostorne neodređenosti i s kojim se sve problemima susreće pri razvrstavanju ovakvih podataka, dan je model prikaza PDF-a objekta pomoću uzoraka minimalnog područja neodređenosti. Također, detaljno su opisani algoritmi *uk-means*, MinMax, Voronojevi dijagrami i SPP metode. Navedeni su postupci kojima se metode služe pri razvrstavanju objekata s prostornom neodređenosti te istaknute prednosti i nedostaci pojedine metode.

Upotreba *uk-means* metode, iako dokazano neučinkovita, vrlo je važna u slučajevima kada nema drugog načina za razvrstavanje objekata s neodređenosti. Kod ove metode računanje ED-a vremenski je vrlo zahtjevan proces zbog velikog broja uzoraka koji je potreban kako bi se što bolje opisao PDF. Kako bi se objekti s neodređenosti podjelili u grozdove upotrebom *uk-means* metode mora se izračunati ED od svakog objekata do svih središta grozdova. Ako je broj objekata jednak  $n$ , broj grozdova jednak  $k$  i broj iteracija po objektu jednak  $t$ , tada je broj računanja ED-a u čitavom procesu razvrstavanja jednak  $n \cdot k \cdot t$ .

MinMax metoda koristi činjenicu da se unutar minimalnog područja neodređenosti, predstavljenog s MBR-om u 2D i MBB-om u 3D, mora nalaziti lokacija objekta. Ako se

uzmu u obzir samo krajnje točke područja neodređenosti moguće je umjesto ED koristiti euklidsku udaljenost koja je jednostavnija za računanje. Također, korištenjem ovakvog postupka razvrstavanja neki se grozdovi mogu odbaciti kao mogući kandidati i bez računanja ED-a. Nedostatak MinMax metode prepoznata je u tome što ona uzima u obzir cijeli geometrijski prostora  $\mathbb{R}^m$ . Korištenjem metode Voronojevih dijagrama taj se problem rješava podjelom geometrijski prostora  $\mathbb{R}^m$  na manje područja.

Predstavljenim teoremom i njegovim dokazima dokazano je da se korištenjem metode Voronojevih dijagrama odbacuje veći broj grozdova kao mogućih kandidata za neki objekt nego korištenjem MinMax metode. Nasuprot tome, glavni nedostatak metode Voronojevih dijagrama sporiji je proces odbacivanja grozdova, nego kod MinMax-a, što umanjuje korist dobivenu od većeg broja odbačenih grozdova.

U prethodnom potpoglavlju predstavljena je i SPP metoda koja koristi ideju o podjeli geometrijskih prostora  $\mathbb{R}^m$ , no to čini konstrukcijom simetrala. Postupak izračunavanja simetrala i projekcija na njih vrlo je jednostavan i računalno nije vremenski zahtjevan proces. Ovom metodom dobije se jednak broj odbačenih grozdova kao i kod metode koja koristi Voronojeve dijagrame, no uz kraće vrijeme razvrstavanja. Moglo bi se zaključiti da je SPP metoda i najbolje rješenje za razvrstavanja objekata s prostornom neodređenosti što bi bilo pogrešno. Pri odabiru metode za razvrstavanje objekata s neodređenosti treba uzeti u obzir razne parametre, kao što su broj objekata, broj uzoraka kojima je modeliran PDF, broj grozdova i veličina minimalnog područja neodređenosti.



## 4. RAZVRSTAVANJE OBJEKATA KOMBINIRANOM METODOM

U prethodnom poglavlju detaljno su objašnjene metode za razvrstavanje objekata s prostornom neodređenosti. Opisani su algoritmi razvrstavanja za *uk-means*, MinMax, Voronojevi dijagrami i SPP metodu. Navedeni su postupci koji se koriste u metodama pri razvrstavanju objekata s prostornom neodređenosti te istaknute prednosti i mane pojedine metode za objekte s prostornom neodređenosti u dvije dimenzije.

U ovom poglavlju predstavljen je pregled metoda za razvrstavanje objekata koji sadrže neodređenost u svojim prostornim podacima, a promatraju se u geometrijskom prostoru  $\mathbb{R}^3$ , koje su nastale prilagodbom do sada objašnjenih metoda. Također, objašnjena je nova metoda za razvrstavanje objekata s neodređenosti koja je razvijena u ovoj disertaciji. Provedeni su pokusi sa različitim parametrima te su napravljene usporedba i analiza rezultata spomenutih metoda razvrstavanja.

Razvrstavanje objekata koji imaju neodređenu prostornu lokaciju u sve tri dimenzije geometrijskog prostora  $\mathbb{R}^3$ , zahtjevnije je od razvrstavanja objekata koji se nalaze u  $\mathbb{R}^2$  prostoru, ne samo zbog jedne dimenzije više, nego i zbog velikog povećanja broja uzoraka koji su potrebni za predstavljanje PDF-a. Razvrstavanje takvih objekata vremenski je vrlo zahtjevno te su istraživanje u ovom području usmjerena razvitaku novih metoda koji će skratiti vrijeme izvođenja procesa razvrstavanja. U praktičnim primjenama potrebno je što prije dobiti rezultat razvrstavanja kako bi se na njegovu osnovu moglo saznati trenutno i predvidjeti buduće stanje te odrediti željene akcije u promatranom sustavu. U ovoj disertaciji proučene su prednosti i nedostaci postojećih metoda razvrstavanja u cilju stvaranja nove metode koja će skratiti vrijeme izvođenja procesa razvrstavanja.

MinMax metodu karakterizira brzo odbacivanje grozdova kao kandidata za neki udaljeni objekt, no nedostatak joj je nedovoljan broj odbačenih grozdova pa je potrebno računanje većeg broja očekivanih udaljenosti (ED-ova). Metoda Voronojevih dijagrama ima veći broj odbačenih grozdova od MinMax metode što rezultira manjim brojem računanja ED-a, ali je sam proces odbacivanja neučinkovit i ukupno trajanje razvrstavanja može trajati duže od onog s MinMax metodom. SPP metoda je učinkovitija od metode Voronojevih dijagrama, no ovisno o parametrima kao što su broj objekata, uzoraka, grozdova i veličina područja neodređenosti i ova metoda može vremenski trajati duže od MinMax metode.

## 4.1. Prilagođavanje postojećih metoda trodimenzionalnom prostoru

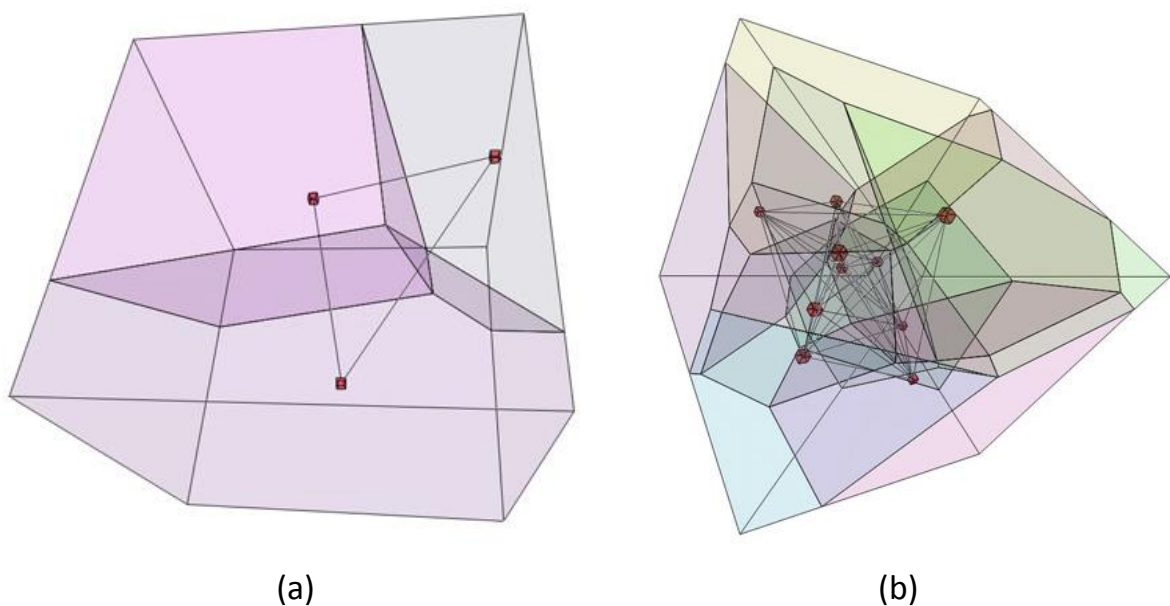
U literaturi [36] [42] [43] [46] dani su postupci i algoritmi za razvrstavanje objekata u geometrijskom prostoru  $\mathbb{R}^2$ . Kako bi se iste mogle primijeniti za razvrstavanje objekata s trodimenzionalnom prostornom neodređenosti potrebno ih je prilagoditi. Kako je već objašnjeno u ranijim poglavljima dvije metode koje imaju, u većini slučajeva, najkraće vrijeme razvrstavanja su MinMax metoda i SPP metodada. Metoda Voronojevih dijagrama je vremenski zahtjevana u postupcima izračunavanja ćelija i provjeravanju leži objektova MBR unutar neke ćelija. Ovi zahtjevi višestruko su veći kada se radi o  $\mathbb{R}^3$  prostoru. Na slici 4.1 [49] prikazan je  $\mathbb{R}^3$  geometrijski prostor s tri (a) i deset (b) središta grozdova, pomoću kojih su formirane Voronojeve ćelije.

### 4.1.1. MinMax metoda u $\mathbb{R}^3$ geometrijskom prostoru

MinMax metoda koristi euklidsku udaljenost, koja je jednostavnija za računanje od ED-a. Euklidska udaljenost između središta grozda  $c_j$  i jednog uzorka  $x$  objekta  $o_i$  u trodimenzionalnom prostoru računa se po jednadžbi (4.1).

$$d(x, c_j) = \sqrt{(x_x - c_{jx})^2 + (x_y - c_{jy})^2 + (x_z - c_{jz})^2} \quad (4.1)$$

Postupak razvrstavanja MinMax metodom u geometrijskom prostoru  $\mathbb{R}^3$  započinje pronalaskom minimalne i maksimalne udaljenosti između objekta  $o_i$  i svih središta



Slika 4.1: Voronojeve ćelije u 3D prostoru s (a) tri i (b) deset središta grozdova. [49]

grozdaova, kao što je prikazano jednadžbama (4.2) i (4.3).

$$\text{MinD}(o_i, c_j) = \min_{x \in \text{MBB}_i} d(x, c_j) \quad (4.2)$$

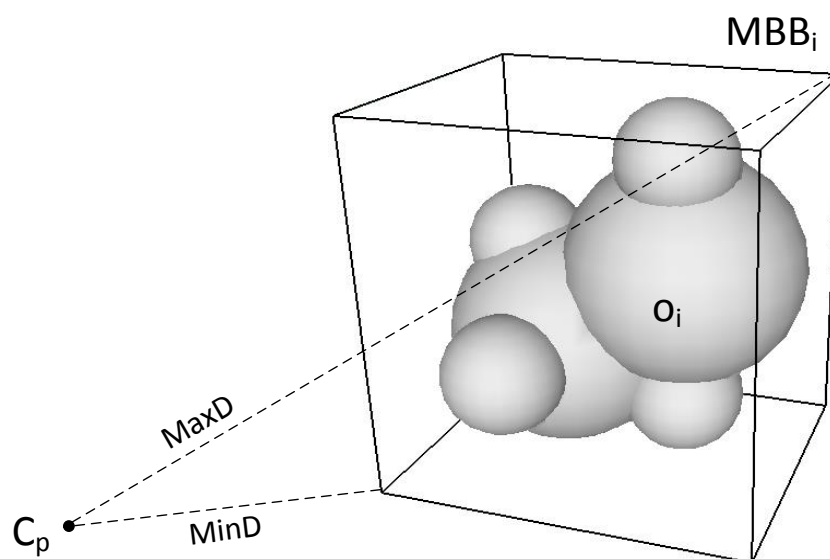
$$\text{MaxD}(o_i, c_j) = \max_{x \in \text{MBB}_i} d(x, c_j) \quad (4.3)$$

U ovom slučaju također vrijedi da je minimalna udaljenost od objekta do središta grozda manja od ED te da je maksimalna udaljenost od objekta do središta grozda veća od ED. Tako se i u ovom slučaju grozdovi odbacuju koristeći pravila dana jednadžbama od (3.9) do (3.11). Na slici 4.2 prikazan je objekt  $o_i$  i središte grozda  $c_p$ . Također su istaknute minimalna i maksimalna udaljenost od središta grozda do najbližih točaka na MBB-u.

Grozdovi koji se ne odbace ovim postupkom trebaju se razvrstati tako da se izračuna njihov ED do objekta i objekt dodijeliti onom grozdu koji ima najmanju ED. Postupak treba ponoviti za sve ostale objekte iz skupa objekata s neodređenosti  $O = \{o_1, o_2, \dots, o_n\}$ .

#### 4.1.2. SPP metoda kao temelj za metodu simetralnih ravnina

SPP metoda ne može se prilagoditi tako jednostavno  $\mathbb{R}^3$  prostoru kao što je to slučaj kod MinMax metode, no može se od nje krenuti kao početne ideje podjele prostora na dva dijela. SPP metoda dijeli prostor između dva grozda i odbacuje jednog od njih kao kandidata za promatrani objekt. Za svaka dva središta grozda  $c_p$  i  $c_q$  konstruira se simetrala  $S_{p|q}$  koja je okomita na dužinu između dva središta i prolazi njezinim polovištem. Nakon konstrukcije simetrale za svaki se objekt  $o_i$  provjerava s koje strane simetrale se nalazi njegov MBR. Ako



Slika 4.2: Minimalna i maksimalna udaljenost između središta grozda  $c_j$  i objekt s prostornom neodređenosti  $o_i$  prikazanog pomoću minimalne granične kutije.

se MBR objekta  $o_i$  nalazi u cijelosti na strani simetrale gdje se nalazi grozd  $c_p$  onda se grozd  $c_q$  odbacuje kao kandidat za objekt  $o_i$ . Grozd  $c_p$  uspoređuje se s preostalim grozdovima dok ne ostane samo jedan kandidat.

U trodimenzionalnom prostoru potrebno je između dva središta grozda konstruirati ravninu koja će podijeliti prostor u dva poluprostora. Ta ravnina mora prolaziti polovištem dužine između dva središta grozda i biti okomita na nju. Takva ravnina naziva se simetralna ravnina jer je okomita na dužinu i sadrži njezino polovište [50]. Tako se između grozdova  $c_p$  i  $c_q$  konstruira simetralna ravnina  $\Pi_{p|q}$  koja dijeli prostor  $\mathbb{R}^3$  na dva poluprostora  $HS_{p|q}$  (na strani groza  $c_p$ ) i  $HS_{q|p}$  (na strani groza  $c_q$ ). Pravila za udaljenost točke  $x$  koja se nalazi u promatranom prostoru dana su u jednadžbama (4.4) i (4.5).

$$d(x, c_p) < d(x, c_q) \quad \forall x \in HS_{p|q} \quad (4.4)$$

$$d(x, c_p) = d(x, c_q) \quad \forall x \in \Pi_{p|q} \quad (4.5)$$

Jednadžbama (4.4) i (4.5) opisna su svojstva točke  $x$  koja se nalazi u poluprostoru  $HS_{p|q}$  u (4.4) i na samoj simetralnoj ravnini  $\Pi_{p|q}$  u (4.5) jednadžbi. Udaljenost takve točke  $x$  do središta grozda  $c_p$  manja je nego do središta grozda  $c_q$  ako se ona nalazi u poluprostoru  $HS_{p|q}$ . No ako se točka  $x$  nalazi na samoj simetralnoj ravnini  $\Pi_{p|q}$  onda vrijedi da je udaljenost od točke  $x$  do središta grozdova  $c_p$  i  $c_q$  jednaka. Ova metoda koristi navedena svojstva kako bi provjerila leži li MBR objekta  $o_i$  u cijelosti u poluprostoru  $HS_{p|q}$ . Ako je taj uvjet ispunjen grozd  $c_q$  odbacuje se kao kandidat za objekt  $o_i$ , a grozd  $c_p$  uspoređuje se s preostalim grozdovima.

Jednadžba simetralne ravnine računa sljedećim postupkom. Za svaki par središta grozdova  $c_p$  i  $c_q$  iz skupa grozdova  $C = \{c_1, c_2, \dots, c_k\}$  koji su zadani s koordinatnima točkama  $c_p = (x_{cp}, y_{cp}, z_{cp})$  i  $c_q = (x_{cq}, y_{cq}, z_{cq})$  u geometrijskom prostoru  $\mathbb{R}^3$  prvo se računa polovište tih točaka po jednadžbi (4.6).

$$P\left(\frac{x_{cp} + x_{cq}}{2}, \frac{y_{cp} + y_{cq}}{2}, \frac{z_{cp} + z_{cq}}{2}\right) \quad (4.6)$$

U polovištu se računa normala  $n_{\Pi}$  simetralne ravnine  $\Pi_{p|q}$  po jednadžbi (4.7).

$$\vec{n}_{\Pi} = A\vec{i} + B\vec{j} + C\vec{k} \quad (4.7)$$

$$A = x_{cp} - x_{cq} \quad B = y_{cp} - y_{cq} \quad C = z_{cp} - z_{cq}$$

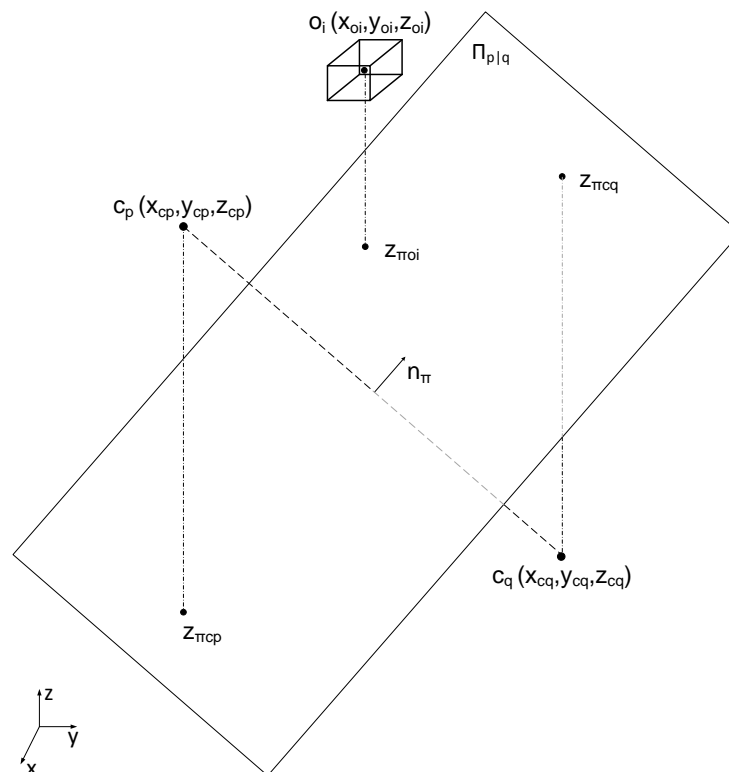
Kako normala leži na ravnini ona s njom dijeli koeficijente  $A, B$  i  $C$  iz kojih se može izračunati jednadžba simetralne ravnine  $\Pi_{p|q}$ , kako je pokazano jednadžbom (4.8).

$$\Pi_{p|q} \dots Ax + By + Cz + D = 0 \quad (4.8)$$

$$D = -\frac{1}{2}[(x_{cp}^2 - x_{cq}^2) + (y_{cp}^2 - y_{cq}^2) + (z_{cp}^2 - z_{cq}^2)]$$

Sada je potrebno utvrditi u kojem poluprostoru se nalazi MBR objekta  $o_i$ . Ako se MBR objekta  $o_i$  nalazi s iste strane ravnine kao središte grozda  $c_p$ , onda se grozd  $c_q$  odbacuje kao kandidat za objekt  $o_i$ . Također vrijedi i obrnuto. U slučaju kada MBR objekta  $o_i$  siječe simetralnu ravninu ne može se odbaciti niti jedan od grozdova, nego se nastavljaju uspoređivati s preostalim grozdovima kandidatima za objekt  $o_i$ . Na kraju svih usporedbi ako je ostalo više kandidata za njih je potrebno izračunati ED do objekta  $o_i$  i grozdu s najmanjom ED dodijeliti objekt.

Postupak odbacivanje grozda kao kandidata za objekt  $o_i$  započinje projiciranjem koordinata središta grozdova  $c_p$  i  $c_q$  te projiciranjem  $MBB_i$  na simetralnu ravninu  $\Pi_{p|q}$ . Za svaki MBB potrebno je projicirati osam koordinata vrhova. Na taj način provjerava se siječe li simetralna ravnina MBB objekta  $o_i$ . Na slici 4.3 prikazan je konstrukcija simetralne ravnine te projiciranje koordinata središta grozdova. Na slici 4.3 MBB objekta  $o_i$  projiciran je samo



Slika 4.3: Prikaz  $\mathbb{R}^3$  prostora podjeljenog simetralnom ravninom između središta grozdova  $c_p$  i  $c_q$ .

jednom (centralnom) točkom zbog jednostavnijeg prikaza, no u proračunima je potrebno provjeriti svaki od osam koordinata vrhova.

Uvjeti za uspješno odbacivanje grozda  $c_q$  prikazani su jednadžbom (4.9).

$$\begin{aligned} & (z_{\pi cp} < z_{cp} \text{ and } z_{\pi oix} < z_{oix}) \vee (z_{\pi cp} > z_{cp} \text{ and } z_{\pi oix} > z_{oix}) \\ & \forall x \in MBB_i = \{1, 2, \dots, 8\} \end{aligned} \quad (4.9)$$

Iz gornje jednadžbe vidljivo je da se uspoređuju samo  $z$  koordinate objekta i središta grozdova s njihovim projekcijama. Tako je sa  $z_{cp}$  označena  $z$  koordinata središta grozda  $c_p$ , a sa  $z_{\pi cp}$  njezina projekcija na simetralnoj ravnini  $\Pi_{p|q}$ . Sve  $z$  koordinate osam vrhova MBB-a objekta  $o_i$  označuju se sa  $z_{oix}$  gdje  $x$  može biti 1 do 8 pa se tako i njihove projekcije na simetralnoj ravnini  $\Pi_{p|q}$  označuju sa  $z_{\pi oix}$  gdje  $x$  odgovara broju vrha.

Za uklanjanje grozda  $c_q$  kao kandidat za objekt  $o_i$  treba bit ispunjen uvjet koji je dan jednadžbom (4.9). Postoje dva slučaja kada će uvjet biti ispunjen. Prvi slučaj definiran je prvim uvjetom iz jednadžbe (4.9) i prikazan je slikom 4.3. U ovom slučaju MBB objekta  $o_i$  se u cijelost nalazi u poluprostoru  $HS_{p|q}$  koji se nalazi iznad simetralne ravnine  $\Pi_{p|q}$ . Koordinata projekcije  $z_{\pi cp}$  izračunava se tako da se u jednadžbu simetralne ravnine uvrste koordinate  $x_{cp}$  i  $y_{cp}$  grozda  $c_p$ . Isto tako se za svih osam vrhova MBB-a uvrštava  $x$  i  $y$  koordinata objekta  $o_i$  u jednadžbu simetralne ravnine i kao rezultat se dobiju koordinate  $z_{\pi oi1}$ ,  $z_{\pi oi2}$ , i tako sve do  $z_{\pi oi8}$ . Nakon izračuna projekcije grozdova i vrhova MBB-a objekta na simetralnu ravninu provjerava se je li zadovoljen uvjet za odbacivanje grozdova prema jednadžbi (4.9). U ovom slučaju taj je prvi uvjet zadovoljen jer se vidi da je  $z_{cp}$  veće od  $z_{\pi cp}$  kao što su i  $z_{oix}$  koordinate MBB-a objekta  $o_i$  veće od njihovih projekcija na simetralnoj ravnini  $z_{\pi oix}$ . Zato se može u ovom slučaju sa sigurnošću reći da objekt  $o_i$  ne pripada grozdu  $c_q$  te se on može odbaciti kao kandidat. Drugi je slučaj kada se MBB objekta  $o_i$  nalazi u cijelost u poluprostoru  $HS_{p|q}$  koji se nalazi ispod simetralne ravnine  $\Pi_{p|q}$ . Koordinate projekcije računaju se na isti način kao i u prvom slučaju te uspoređuju s koordinatama objekta  $o_i$  i središta grozda  $c_p$  prema drugom uvjetu iz jednadžbe (4.9). U ovom slučaju će taj drugi uvjet biti zadovoljen i grozd  $c_q$  će se opet moći odbaciti kao kandidat za objekt  $o_i$ . U slučaju kada simetralna ravnina siječe MBB objekta koji se razvrstava, taj objekt nije moguće razvrstati niti odbaciti bilo koji od promatranih grozdova. Za ispravno razvrstavanje potrebno izračunati ED od grozdova do objekta i dodijeliti ga grozdu s najmanjim ED-om.

## 4.2. Kombinirana metoda razvrstavanja objekata s trodimenzionalnom prostornom neodređenosti

Kako je već ranije ustanovljeno, MinMax metodu karakterizira brzo odbacivanje grozdova kao kandidata za neki udaljeni objekt, a nedostatak joj je nedovoljan broj odbačenih grozdova, nasuprot njoj metoda simetralnih ravnina ima veći broj odbačenih grozdova od MinMax metode što rezultira manjim brojem računanja ED-a, no sam proces odbacivanja može imati ukupno vremensko trajanje razvrstavanja duže od onog dobivenog MinMax metode ovisno o parametrima kao što su broj objekata, uzoraka, grozdova i veličina područja neodređenosti. Ideja je bila razviti novu Kombiniranu metodu za odbacivanja središta grozdova i razvrstavanje objekta koji sadrže neodređenost u svojim prostornim podacima koja će iskoristiti prednosti postojećih metoda s ciljem smanjenja ukupnog vremenskog trajanja postupka razvrstavanja.

Kombinirana metoda koristi dijelove algoritma MinMax i metode simetralnih ravnina. Pseudo kôd za razvrstavanje i odbacivanje središta grozdova kao kandidata Kombiniranom metodom prikazan je algoritmom 4.1. Zbog jednostavnosti računanja minimalnih i maksimalnih euklidskih udaljenosti i brzine odbacivanja udaljenih grozdova od promatranog objekta  $o_i$  prvo se provodi postupak MinMax metode te se odbacuju svi grozdovi čija je

Algoritam 4.1: Pseudo kôd Kombinirane metode za razvrstavanje objekata s neodređenosti.

1. **Ulaz:** skup objekata  $O$  i početne pozicije grozdova  $C$
2. **for all**  $c_j \in C$  i objekt  $o_i$  i  $C = C_k$
3.     izračunati  $MinD(o_i, c_j)$  i  $MaxD(o_i, c_j)$  prema (4.2) i (4.3)
4.     izračunati  $MinMaxD(o_i)$
5.     **for all**  $c_j \in C$
6.         **if**  $MinD(o_i, c_j) \geq MinMaxD(o_i)$
7.             ukloni  $c_j$  iz skupa mogući kandidata  $C_k$  za objekt  $o_i$
8.         **end if**
9.     **end for all**
10. **end for all**
11. izračunaj simetralne ravnine za  $C_k$
12. **for all**  $c_p \in C_k$  i  $c_q \in C_k$ ;  $c_p \neq c_q$
13.     **if**  $MBB_i \subseteq HS_{p|q}$  prema (4.9)
14.         ukloni  $c_q$  kao kandidata za  $o_i$
15.         ukloni  $c_q$  iz for petlje
16.     **else**  $MBB_i \subseteq HS_{q|p}$  prema (4.9)
17.         ukloni  $c_p$  kao kandidata za  $o_i$
18.         ukloni  $c_p$  iz for petlje
19.     **end if**
20. **end for all**
21. **for all** ostale nedodjeljene objekte i grozdove
22.     izračunati  $ED$  prema (3.4)
23. **end for all**

minimalna udaljenost veća od najmanje maksimalne udaljenosti za objekt  $o_i$ . U slučaju da se odbace svi grozdovi osim jednoga postupak razvrstavanje je završen. U izvedenim pokusima čiji su rezultati izloženi u sljedećem potpoglavlju to nije bio čest slučaj. U slučaju da se ne odbace svi grozdovi MinMax dijelom algoritma, ostali grozdovi (koji se nalaze u skupu  $C_k$ ) odbacuju se metodom simetralne ravnine. I u ovom slučaju kao i kod SPP metode postoji uklanjanje neposredno odbačenih grozova iz petlje. Ako se ni ovim dijelom algoritma ne odbace svi grozdovi potrebno je računati ED od preostalih grozdova do objekta  $o_i$  i dodijeliti ga grozdu s najmanjom ED-om.

U sljedećem potpoglavlju provedeni su pokusi koji na zadanim skupovima parametara uspoređuju metode za razvrstavanje objekata s neodređenosti. U pokusima se uspoređuju prilagođena MinMax metoda i metoda simetralne ravnine s novom Kombiniranom metodom s ciljem dokazivanja smanjenja vremenskog trajanja procesa razvrstavanja objekata koji su prostorno neodređeni.

### 4.3. Usporedba i analiza rezultata pokusa

Pokusi su jedan od način dokazivanja da razvijena Kombinirana metoda ima kraće vrijeme trajanja procesa razvrstavanja od postojećih metoda. Metode s kojima se uspoređuje ova nova metoda su one koje su se do sada pokazale kao najbolji izbor za određeni skup objekata i broj grozdova. Metode sa najkraćim vremenom izvođenja do sada su MinMax metoda i metoda simetralnih ravnina. MinMax metoda ima brži proces odbacivanja grozdova nego metoda simetralnih ravnina, no broj uspješno odbačenih grozdova veći je kod metode simetralnih ravnina. Uz ove činjenice, na ukupno vrijeme trajanja razvrstavanja, kod ove dvije metode, jako utječe i računanje očekivane udaljenosti (ED-a). Ako računanje ED-a zahtijeva više vremena, a to je u slučaju kada je broj uzoraka za modeliranje funkcije gustoće vjerojatnosti (PDF-a) velik, metoda simetralnih ravnina imati će kraće vrijeme razvrstavanja od MinMax metode. Ako je broj uzoraka manji, potrebno je manje vremena za računanje ED-a i preporučuje se koristiti MinMax metodu jer će potrebno vrijeme za izračun simetralnih ravnina i projekcija na njih biti veće od vremena koje je potrebno za izračun minimalnih i maksimalnih udaljenosti.

U pokusima se uspoređuju sve tri metode s ciljem dokazivanja koja metoda daje bolje rezultate za slučajeve kada se mijenja broj objekata, veličina minimalnog područja nesigurnosti, broj uzoraka i broj grozdova. Za sve pokuse korišten je osnovni skup objekata



koji se sastoji od  $n$  objekata. Svi objekti smješteni su u trodimenzionalni prostoru veličine  $100 \times 100 \times 100$ . Svaki objekt opisan je s minimalnim područjem neodređenosti MBB. Svakom neodređenom objektu duljina svih stranica MBB-a se generirana slučajnim odabirom, s tim da je maksimalna duljina stranice ograničena duljinom  $a$ . MBB svakog objekta podijeljen je na  $\sqrt[3]{s} \times \sqrt[3]{s} \times \sqrt[3]{s}$  uzoraka, gdje je  $s$  ukupan broj uzoraka koji se koristi za predstavljanje PDF-a. Svaki uzorak ima vjerojatnost da se objekt nalazi u tom uzorku. Zbroj vjerojatnosti svih uzoraka jednak je jedan. Objekti su raspoređeni po ukupnom području skupa objekata i među njima se mogu odabrati predstavnici za početni skup od  $k$  grozdova.

Svi pokusi ponovljeni su najmanje deset puta zbog povećanja točnosti rezultata, a kao reprezentativni rezultat uzima se srednja vrijednost izmjerenih rezultata. Rezultati razvrstavanja različitih metoda uspoređuju se nakon svakog pokusa kako bi se osigurala točnost razvrstavanja. Pokusi su provedeni na osobnom računaru (Intel Core i7-870 2.93GHz procesor; 4GB radne memorije) u programskom paketu MATLAB 2009b. Kao osnovni skup parametara izabrane su vrijednosti prikazane u tablici 4.1. Vrijednosti osnovnog skupa parametara temelje se na pokusima provedenim u [43]. Ideja je bila provesti pokuse nad osnovnim skupom parametara i nakon toga u ostalim pokusima mijenjati jedan od parametara dok ostali parametri zadržavaju vrijednosti iz osnovnog skupa.

Tablica 4.1: Vrijednosti osnovnog skupa parametara

Parametar	Kratice	Vrijednost
Broj objekata s neodređenosti	$n$	500
Maksimalna duljina stranice MBB-a	$a$	6
Broj uzoraka u MBB-u	$s$	15625
Broj grozdova	$k$	49

Predloženim pokusima nastojalo se provjeriti koja od metoda ima najbolja svojstva u pojedinim primjenama. Pokusima se treba dokazati kako se upotrebom Kombinirane metode dobiju bolji rezultati, tj. kraće vrijeme potrebno za razvrstavanje nego MinMax metodom ili metodom simetralnih ravnina.

### 4.3.1. Analiza rezultata pokusa s osnovnim skupom parametara

U ovom potpoglavlju provedeni su pokusi s osnovnim skupom parametara, koji se prikazanim u tablici 4.1. Svi pokusi provedeni su deset puta, a prikazani rezultati u tablici 4.2

su srednja vrijednost izmjerenih rezultata. Rezultati razvrstavanja sve tri metode se uspoređuju nakon svakog pokusa kako bi se osiguralo da je razvrstavanje objekata u grozdove točno. U tablici, uz naziv metode prikazano je vrijeme potrebno za proces razvrstavanja u sekundama i broj računanja očekivane udaljenosti po objektu u svakoj iteraciji (eng. Number of Expected Distance Calculations per Object per Iteration - NED). NED se upotrebljava kao mjera uspješnosti, tj. taj broj govori koliki je broj računanja ED-a po jednom objektu u svakoj iteraciji. Na primjer za *uk-means* metodu, koja ne koristi odbacivanje grozdova, NED je jednak broju grozdova jer se računa ED od objekta do svakog grozda. Za odbačene grozdove ne računa se ED što smanjuje NED. Iz tablice 4.2 je vidljivo da je NED kod promatranih metoda znatno manji od broja grozdova pa su oni znatno brži od *uk-means* metoda. Iz tablice 4.2 također se vidi da je NED kod metode simetralnih ravnina i Kombinirane metode jednak, što je bilo i za pretpostaviti poznavajući činjenice o načinu odbacivanja. U ta dva slučaja NED je smanjen s 49 (upotrebom *uk-means* metode) na 0,46 što predstavlja smanjenje od 99%.

Tablica 4.2: Vremena trajanja procesa razvrstavanja i NED za pojedinu metodu koristeći vrijednosti osnovnog skupa parametara

Korištena metoda razvrstavanja	Vrijeme izvođenja [s]	NED
MinMax 3D	85,23	1,46
Metoda simetralnih ravnina	96,57	0,46
Kombinirana metoda	33,13	0,46

U nekim primjenama dubinske analize podataka vrijeme je kritična varijabla i potrebno je doći do rezultata u što kraćem vremenu. Kako je razvrstavanje bitan dio te analize cilj svake nove metode je skraćivanje vremena potrebnog za proces razvrstavanja. Iz rezultata dobivenih ovim pokusom vidljivo je da Kombinirana metoda ima najmanje vrijeme potrebno da se razvrsta 500 objekata u 49 grozdova u  $\mathbb{R}^3$  geometriskom prostoru veličine  $100 \times 100 \times 100$ .

Dodatni primjer da se za različite parametre dobiju različite vrijednosti vremena trajanja razvrstavanja vidljiv je u sljedećem primjer. U tablici 4.2 vidljivo je da najkraće vrijeme ima nova Kombinirana metoda. Drugo najkraće vrijeme ima MinMax metoda i s najlošijim vremenom je metoda simetralnih ravnina. No, ako se poveća broj uzoraka u pojedinom MBB-u sa 15625 na 19683 onda se dobiju vremena prikazana u tablici 4.3. I dalje nova Kombinirana metoda ima najkraće vrijeme razvrstavanja, ali ovaj put najduže vrijeme razvrstavanja ima MinMax metoda.

Tablica 4.3: Vremena trajanja procesa razvrstavanja i NED za pojedinu metodu koristeći sljedeće parametre ( $n=500$ ,  $a=6$ ,  $s=19683$ ,  $k=49$ )

Korištena metoda razvrstavanja	Vrijeme izvođenja [s]	NED
MinMax 3D	115,31	1,44
Metoda simetralnih ravnina	109,42	0,44
Kombinirana metoda	45,78	0,44

U oba pokusa rezultat NED veći je kod MinMax metode, nego kod ostale dvije metode. Razlog toga detaljno je objašnjen u ranijim poglavljima, a može se zaključiti da odbacuje manje grozdova jer ima slabija svojstva koja koristi pri odbacivanju. Zbog toga se pri razvrstavanju MinMax metodom mora računati veći broj ED-a po objektu u svakoj iteraciji pa je i njegovo vrijeme izvođenja duže kada je MBB objekta opisan s većim brojem uzoraka.

### 4.3.2. Analiza rezultata pokusa s promjenom broja objekata

Pokusi u ovom potpoglavlju provedeni su s ciljem izračunavanja vremena trajanja procesa razvrstavanja pri različitom broju objekata koje treba razvrstata. Razvrstavanje je provedeno sa sve tri promatrane metode i dobiveni rezultati prikazani su u tablici 4.4. Prvi pokus ima 250, drugi 500, teći 750, četvrti 1000 i peti 1500 objekata. Ostali parametri jednaki su onima iz osnovnog skupa u tablici 4.1. Svaki pokus ponovljen je deset puta i na slici 4.4 prikazane su prosječne vrijednosti vremena trajanja razvrstavanja.

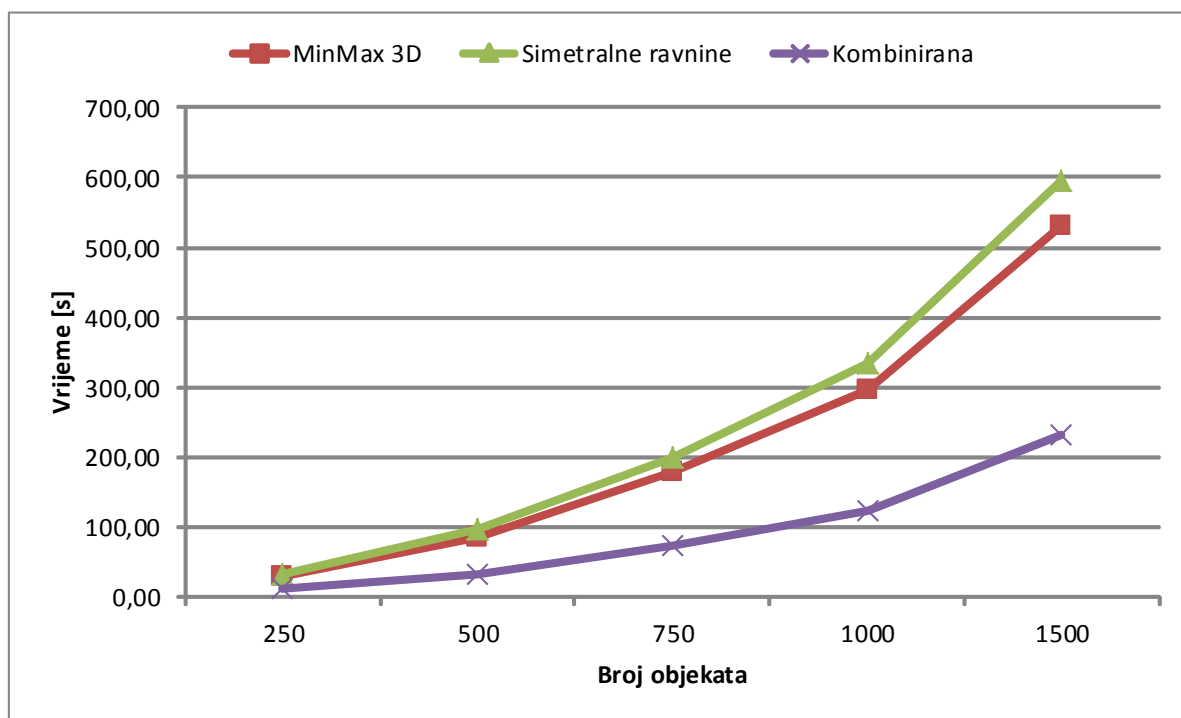
Tablica 4.4: Vremena trajanja procesa razvrstavanja i NED za pojedinu metodu koristeći različit broj objekata  $n$  i ostale varijable iz osnovnog skupa parametara

Metoda razvrstavanja	Vrijeme izvođenja [s] za $n=$					avg. NED
	250	500	750	1000	1500	
MinMax 3D	27,57	85,23	177,16	294,91	532,38	1,46
Simetralnih ravnina	30,74	96,57	200,10	335,16	596,20	0,47
Kombinirana	9,92	33,13	72,02	121,94	232,28	0,47

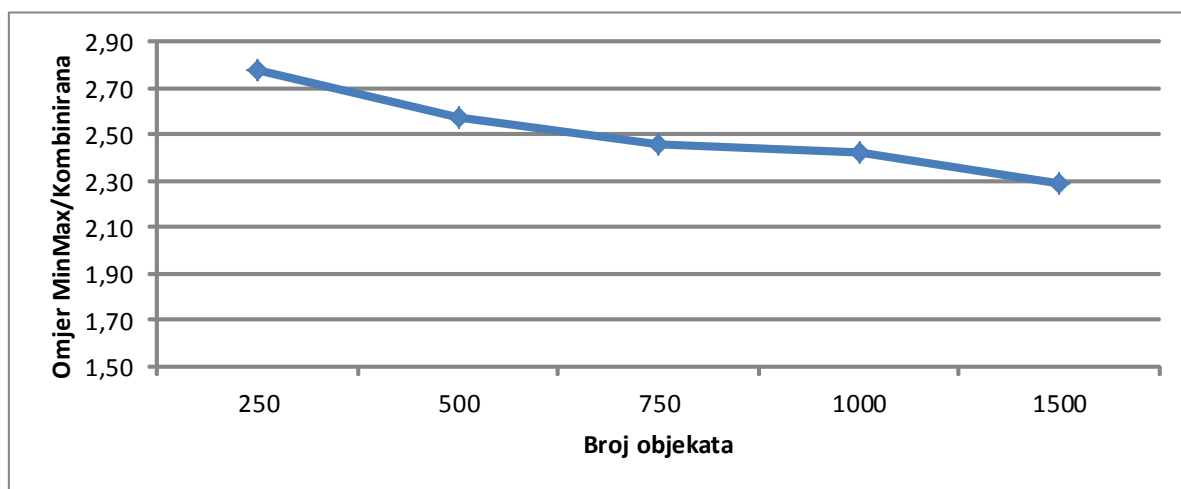
Iz tablice 4.4 i slike 4.4 vidljivo je da Kombinirana metoda ima kraće vrijeme razvrstavanja od ostale dvije metode. Kako se povećava broj objekata iz pokusa u pokus tako se povećava i vrijeme izvođenja procesa razvrstavanja. Povećanjem broja objekata povećava

se i broj relacija objekt - grozd koje treba izračunati. Primjećuje se da metoda simetralnih ravnina i Kombinirana metoda imaju jednak NED, jer imaju jednak broj odbačenih grozdova, no Kombinirana metoda ima znatno kraće vrijeme trajanja razvrstavanja zbog upotrebe bržih procesa odbacivanja grozdova.

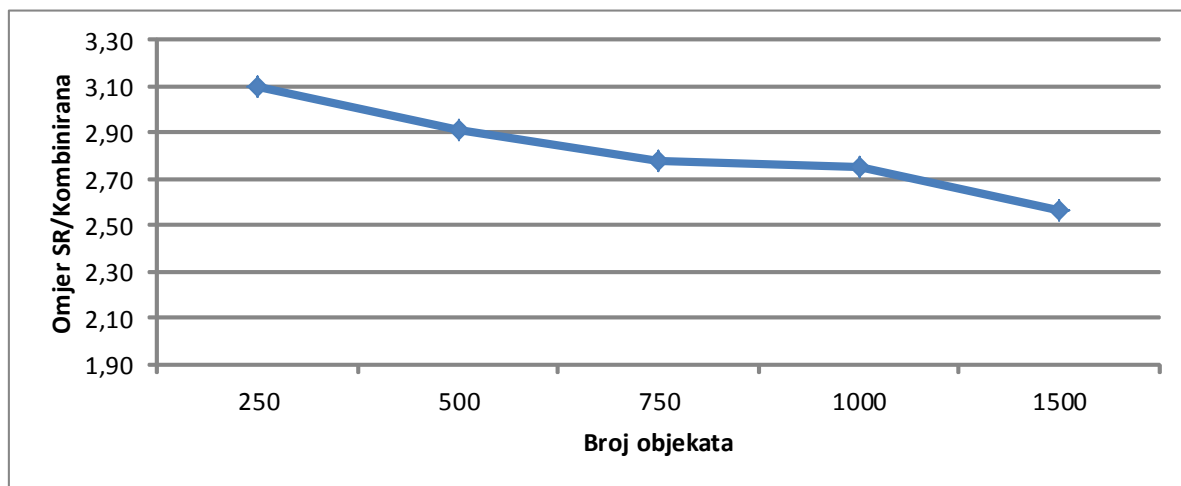
Na slikama 4.5 i 4.6 prikazani su omjeri vremena potrebnog da se proces razvrstavanja završi. Na slici 4.5 prikazan je omjer MinMax i Kombinirane metode, a na slici 4.6 prikazan je omjer metode simetralnih ravnina i Kombinirane metode. Sa slika je vidljivo da se s



Slika 4.4: Grafički prikaz vremena trajanja razvrstavanja za sve tri metode pri promjeni broja objekata.



Slika 4.5: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode pri promjeni broja objekata.



Slika 4.6: Grafički prikaz omjera vremena izvođenja metode simetralnih ravnina i Kombinirane metode pri promjeni broja objekata.

povećanjem broja objekata smanjuju međusebni omjeri, što znači da Kombinirana metoda brže razvrstava manje skupove objekata, no Kombinirana metoda ima kraća vremena izvođenja u čitavom rasponu broja objekata, što ju čini najboljom metodom za primjenama s bilo kojim brojem objekata.

### 4.3.3. Analiza rezultata pokusa s promjenom maksimalne duljine stranice MBB-a

Pokusi u ovom potpoglavlju provedeni su s ciljem izračunavanja vremena trajanja procesa razvrstavanja pri različitim maksimalnim duljinama stranice MBB-a, tj. različite veličine neodređenosti objekata. Razvrstavanje je provedeno sa sve tri promatrane metode i dobiveni rezultati prikazani su u tablici 4.5. U prvom pokus je maksimalna stranica  $a$  MBB-a je 2, drugom 4, tećem 6, četvrtom 10 i u petom 12 jedinica u  $\mathbb{R}^3$  geometrijskom prostoru

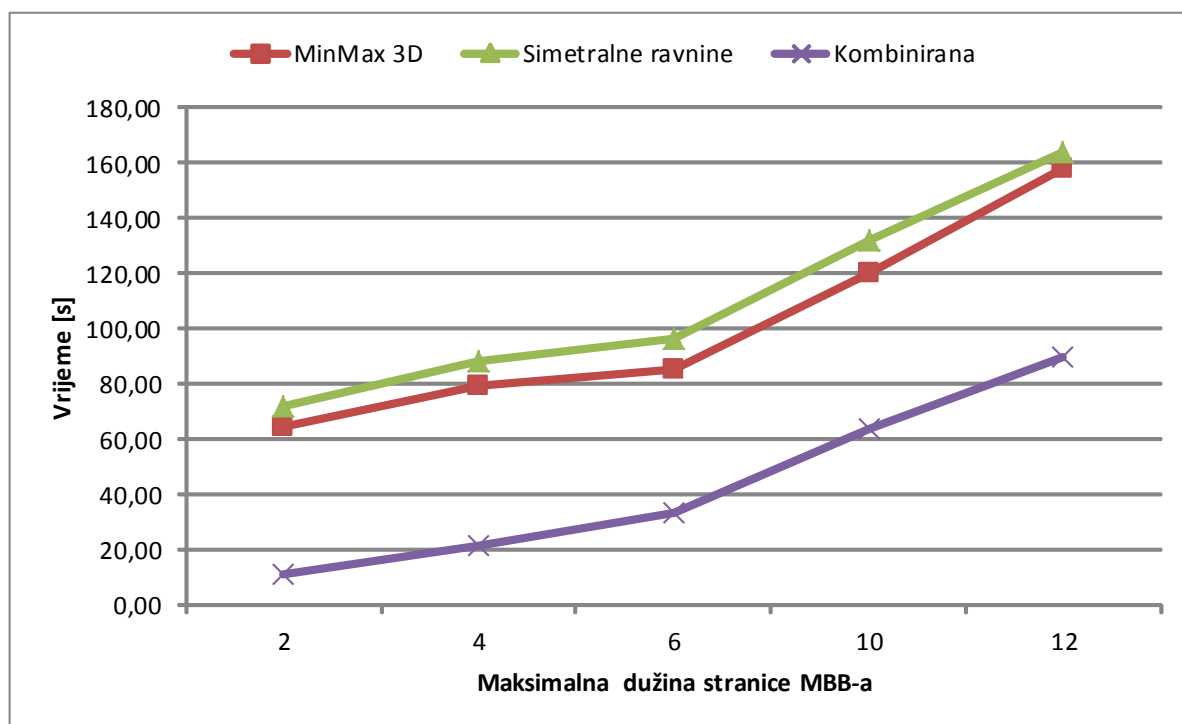
Tablica 4.5: Vremena trajanja procesa razvrstavanja i NED za pojedinu metodu koristeći različite maksimalne duljine stranice  $a$  i ostale varijable iz osnovnog skupa parametara

Metoda razvrstavanja	Vrijeme izvođenja [s] za $\max(a)=$					avg. NED
	2	4	6	10	12	
MinMax 3D	64,69	78,95	85,23	120,50	158,11	1,59
Simetralnih ravnina	71,83	88,43	96,57	131,83	163,66	0,55
Kombinirana	11,06	21,36	33,13	63,74	89,61	0,55

veliĉine  $100 \times 100 \times 100$  jedinica. Ostali parametri jednaki su onima iz osnovnog skupa u tablici 4.1. Svaki pokusi ponovljen je deset puta i na slici 4.7 prikazane su prosjeĉne vrijednosti vremena trajanja razvrstavanja.

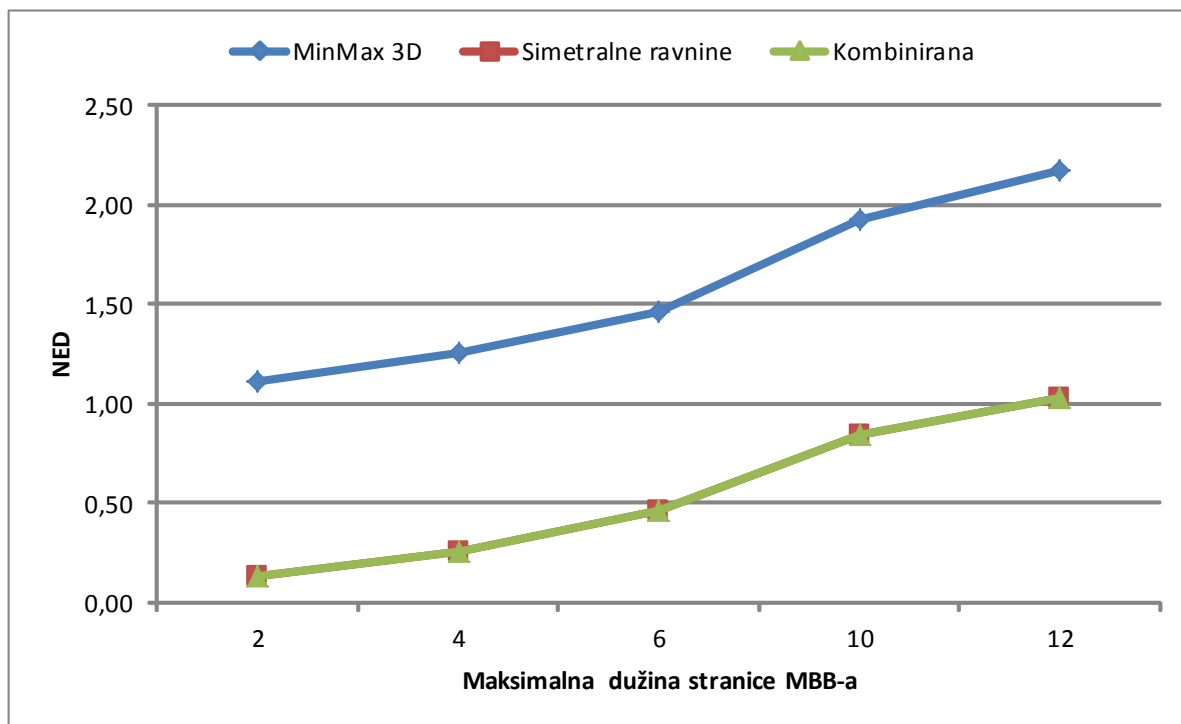
Povećanjem duljine stranice MBB-a povećava se minimalno područje neodređenosti. Što je veća duljina stranice  $a$  to je veća vjerojatnost da će simetralna ravnina presijecati MBB, što direktno dovodi do manjeg broja odbaĉenih grozdova pomoću metode simetralnih ravnina. Što se više povećava duljina stranice  $a$  tako se povećava i maksimalna udaljenost od objekta do grozda što opet dovodi do manjeg broja odbaĉenih grozdova pomoću MinMax metode. Sa slike 4.8 vidljivo je da će doći do porasta broja raĉunanja ED-a zbog rasta NED-a s porastom duljine maksimalne stranice.

Na slici 4.7 vidljivo je kako je Kombiniranoj metodi potrebno kraće vrijeme za razvrstavanje objekata s neodređenosti, od preostale dvije metode, za različite duljine stranica MBB-a. Zanimljiva ĉinjenica je omjer vremena izvođenja izmeĉu Kombinirane metode i ostale dvije metode. Na slici 4.9 prikazan je MinMax/Kombinirana omjer vremena potrebnog da se proces razvrstavanja završi, a na slici 4.10 prikazan je omjer vremena za Simetralnu/Kombiniranu metoda. S obje slike može se zakljuĉiti da omjeri opadaju s povećanjem duljine stranice i ustaljuju se na omjeru 1,75. S daljnjim povećanjem duljine stranice  $a$  za oĉekivati je da će taj omjer ostati nepromijenjen.

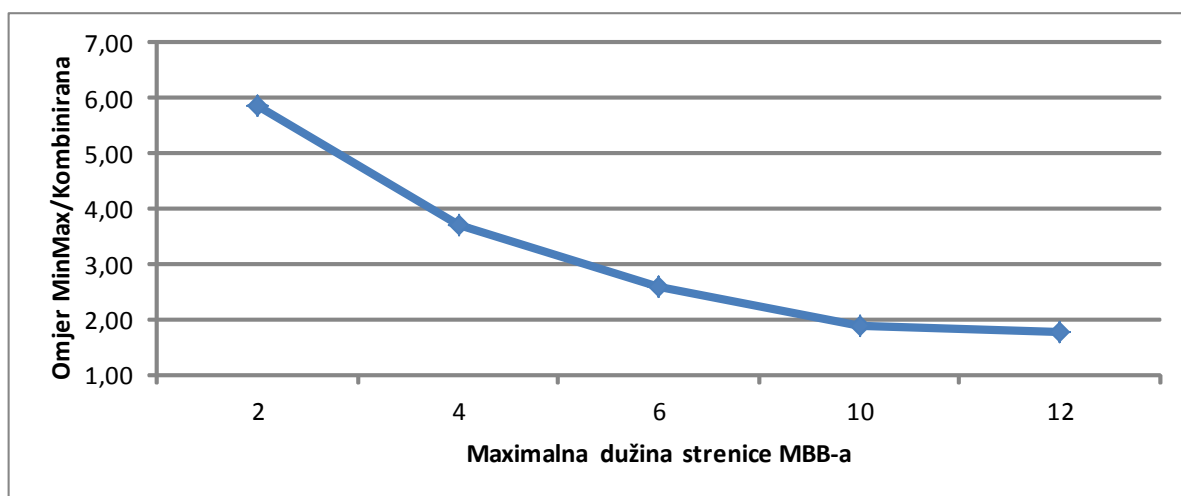


Slika 4.7: Grafiĉki prikaz vremena trajanja razvrstavanja za sve tri metode pri promjeni maksimalne duljine stranice  $a$  MBB-a.

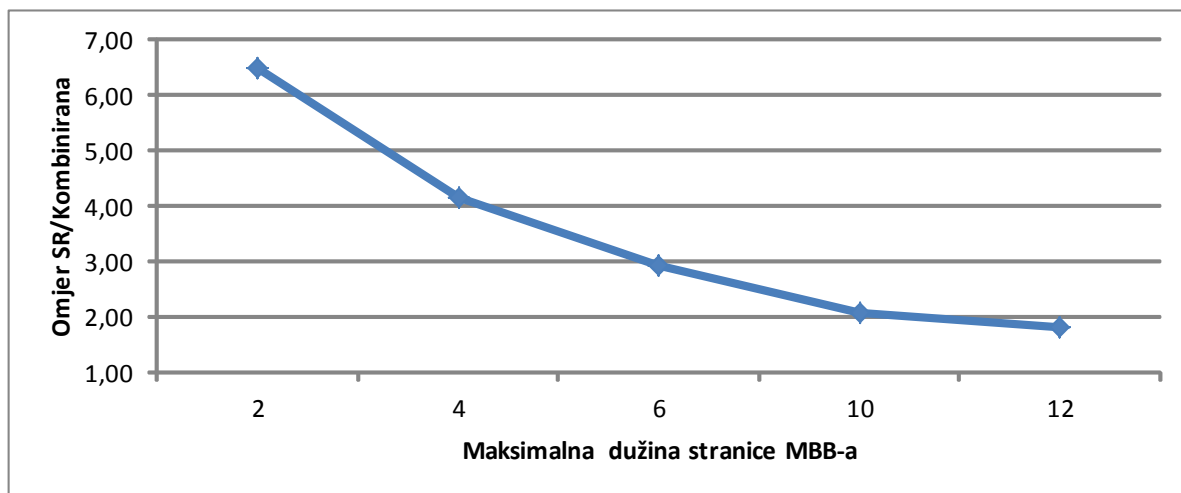
U oba slučaja iznimka su najmanje duljine stranice  $a$  MBB-a. Kod malih duljina MinMax metoda ima lošija svojstva odbacivanja grozdova i NED je osam puta veći, nego kod Kombinirane metode kao što je prikazano slikom 4.11. I na ovoj slici vidljiv je trend smanjenja omjera i njegovog ustaljenja s povećanjem duljine stranice MBB-a. Na samom kraju može se zaključiti kako će Kombinirana metoda imati najkraće vrijeme razvrstavanja za sve veličine MBB-a.



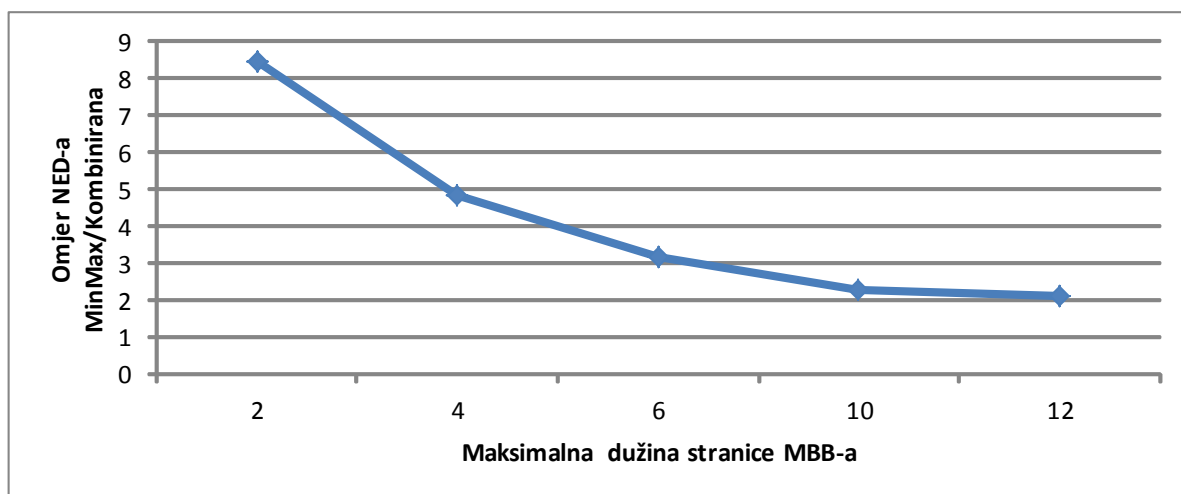
Slika 4.8: Prikaz promjene NED-a za sve tri metode pri promjeni maksimalne duljine stranice MBB-a.



Slika 4.9: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode pri promjeni maksimalne duljine stranice MBB-a.



Slika 4.10: Grafički prikaz omjera vremena izvođenja metode simetralnih ravnina i Kombinirane metode pri promjeni maksimalne duljine stranice MBB-a.



Slika 4.11: Grafički prikaz omjera NED-a za MinMax i Kombiniranu metodu pri promjeni maksimalne duljine stranice MBB-a.

#### 4.3.4. Analiza rezultata pokusa s promjenom broja uzoraka u MBB-u

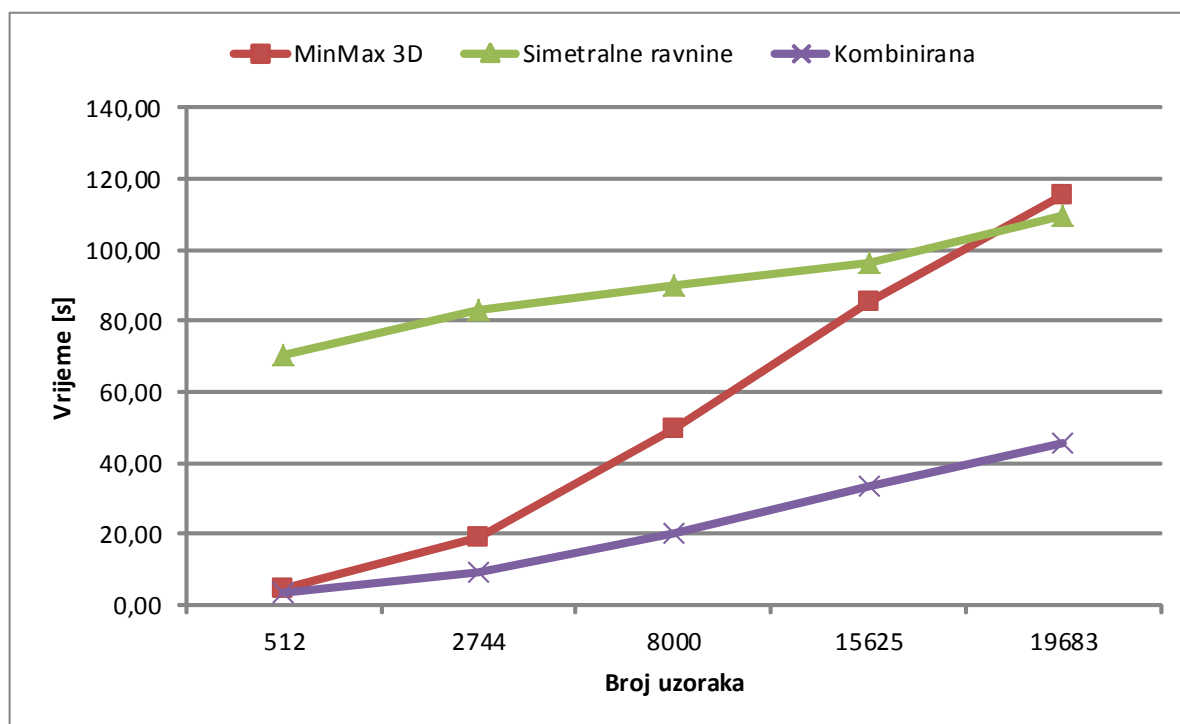
Pokusi u ovom potpoglavlju provedeni su s ciljem izračunavanja vremena trajanja procesa razvrstavanja pri različitom broju uzoraka MBB-a. Što je veći broj uzoraka koji se koristi za modeliranje PDF-a, moguće je preciznije opisati kolika je vjerojatnost da se objekt nalazi u nekoj točki MBB-a. Razvrstavanje je provedeno sa sve tri promatrane metode i dobiveni rezultati prikazani su u tablici 4.6. U prvom pokus MBB podijeljen je na 512, drugom na 2744, trećem na 8000, četvrtom na 15625 i u petom na 19683 uzoraka. Ostali parametri jednaki su onima iz osnovnog skupa u tablici 4.1. Svaki pokus ponovljen je deset puta i na slici 4.12 prikazane su prosječne vrijednosti vremena trajanja razvrstavanja.



Tablica 4.6: Vremena trajanja procesa razvrstavanja i NED za pojedinu metodu koristeći različit broj uzoraka u MBB-u i ostale varijable iz osnovnog skupa parametara

Metoda razvrstavanja	Vrijeme izvođenja [s] za $s =$					avg. NED
	512	2744	8000	15625	19683	
MinMax 3D	4,54	18,85	49,52	85,23	115,31	1,44
Simetralnih ravnina	70,35	83,30	89,86	96,57	109,42	0,44
Kombinirana	3,60	9,19	20,24	33,13	45,78	0,44

Broj uzoraka  $s$  koji se nalaze unutar nekog MBB-a, ne utječe na broj odbačenih grozdova, tj NED, ali zato utječe na vrijeme potrebno za računanje ED-a. Za svaki objekt koji nije dodijeljen grozdu na kraju procesa odbacivanja mora se izračunati udaljenost između svakog uzorka i promatranih središta grozdova. Zato je jasno da će s povećanjem broja uzoraka računanje ED-a trajati duže. Poznato je da metoda simetralnih ravnina i Kombinirana metoda imaju isti NED, tako da će se kod ove dvije metode uvijek računaju isti broj ED-a, no zbog bržeg procesa odbacivanja grozdova Kombinirana metoda imati će kraće vrijeme potrebno za razvrstavanje objekata. Nasuprot ove dvije metode, MinMax metoda odbacuje manji broj grozdova pa je broj računanja ED-a u njezinom slučaju veći. Zato je sa slike 4.12



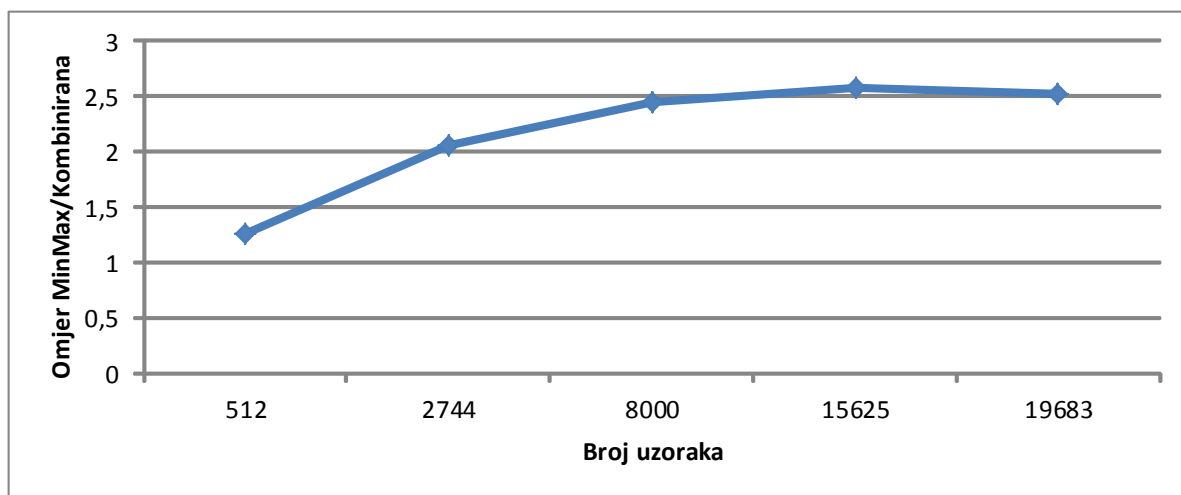
Slika 4.12: Grafički prikaz vremena trajanja razvrstavanja za sve tri metode pri promjeni broja uzoraka  $s$  koji se nalaze unutar MBB-a.

vidljivo da broj uzoraka ima veći utjecaj na vrijeme izvođenja MinMax metode. U dosadašnjim pokusima MinMax metoda postizala je bolje rezultate od metode simetralnih ravnina, no s povećanjem broja uzoraka iskazuje se njezina slabost, a to su ED izračuni.

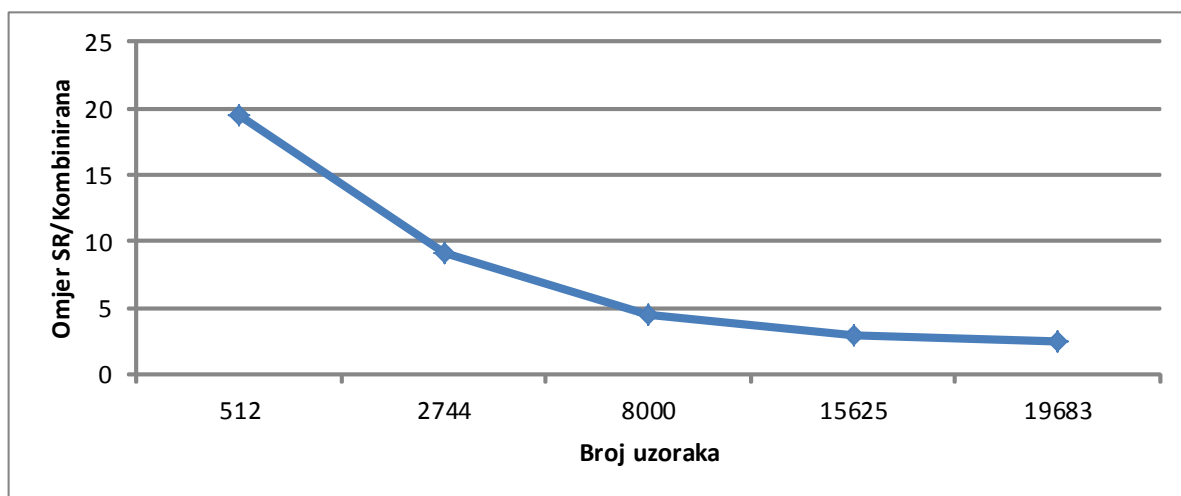
Na slikama 4.13 i 4.14 prikazani su omjeri vremena potrebnog da se proces razvrstavanja završi. Na slici 4.13 prikazan je omjer MinMax i Kombinirane metode, a slici 4.14 omjer metode simetralnih ravnina i Kombinirane metode.

### 4.3.5. Analiza rezultata pokusa s promjenom broja grozdova

Pokusi u ovom potpoglavlju provedeni su s ciljem izračunavanja vremena trajanja procesa razvrstavanja pri različitom broju grozdova u koje treba razvrstati objekte.



Slika 4.13: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode pri promjeni broja uzoraka unutar MBB-a.



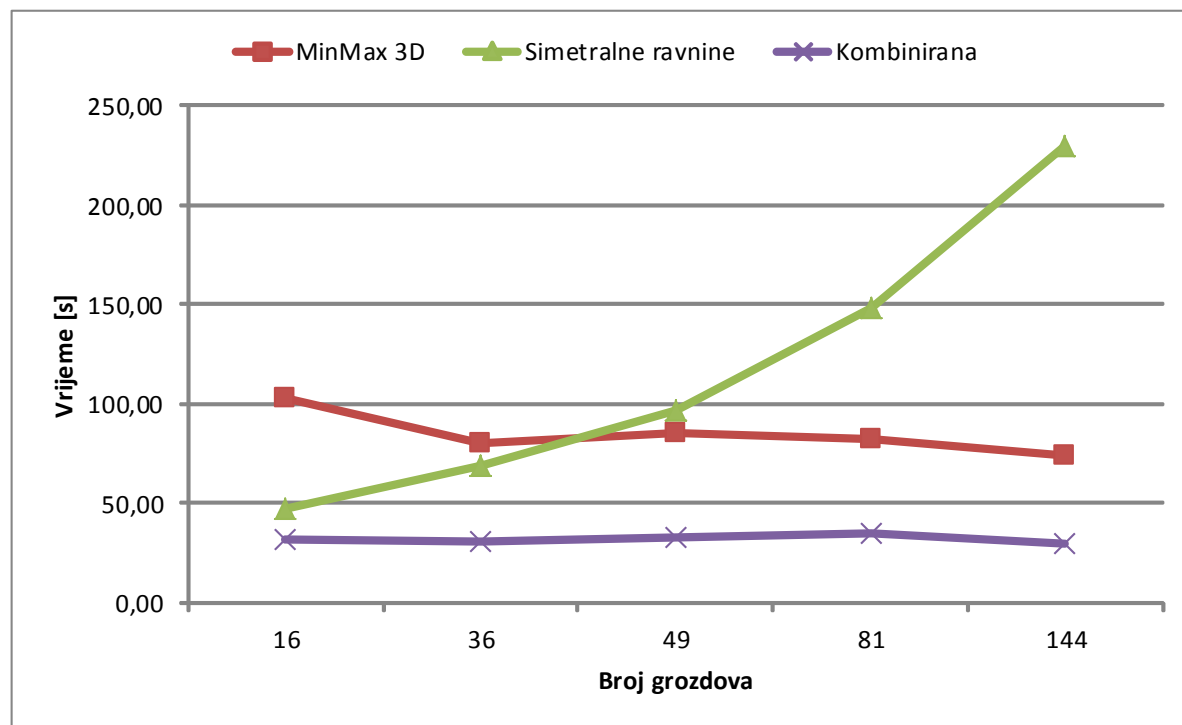
Slika 4.14: Grafički prikaz omjera vremena izvođenja metode simetralnih ravnina i Kombinirane metode pri promjeni broja uzoraka unutar MBB-a.

Razvrstavanje je provedeno sa sve tri promatrane metode i dobiveni rezultati prikazani su u tablici 4.7. U prvom pokus je zadano 16, drugom 34, trećem 49, četvrtom 66 i u petom 144 grozda. Ostali parametri jednaki su onima iz osnovnog skupa u tablici 4.1. Svaki pokus ponovljen je deset puta i na slici 4.15 prikazane su prosječne vrijednosti vremena trajanja razvrstavanja.

Tablica 4.7: Vremena trajanja procesa razvrstavanja i NED za pojedinu metodu koristeći različit broj grozdova  $k$  i ostale varijable iz osnovnog skupa parametara

Metoda razvrstavanja	Vrijeme izvođenja [s] za $k=$					avg. NED
	16	34	49	66	144	
MinMax 3D	103,07	79,83	85,23	82,07	74,08	1,56
Simetralnih ravnina	47,54	68,43	96,57	148,04	230,03	0,47
Kombinirana	31,87	30,46	33,12	34,90	30,17	0,47

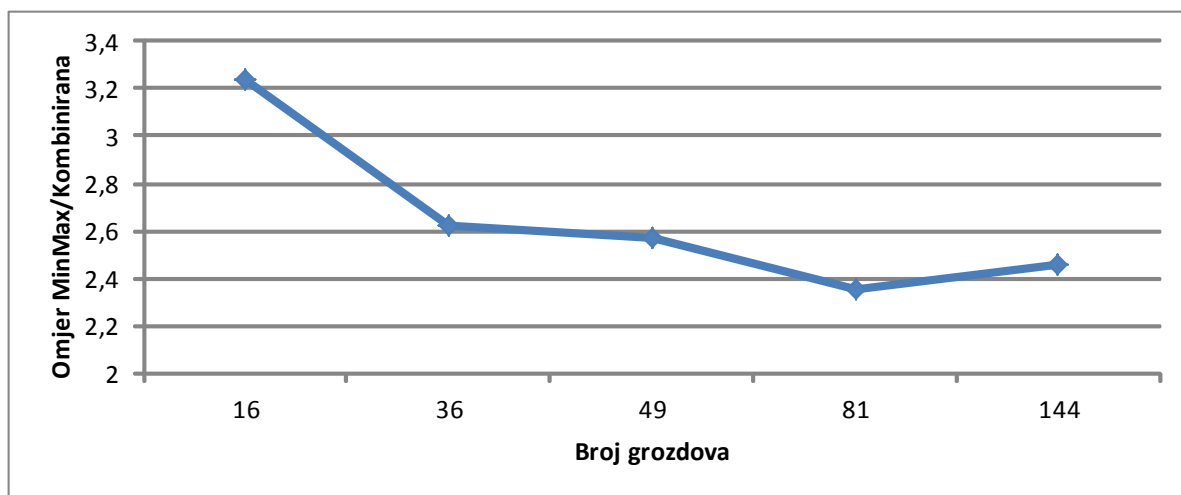
Iz tablice i slike vidljivo je da najkraće vrijeme izvođenja razvrstavanja ima Kombinirana metoda. S povećanjem broja grozdova povećava se i broj relacija između objekata i grozdova koje treba usporediti. Sa slike 4.15 vidi se da promjena broja grozdova ne



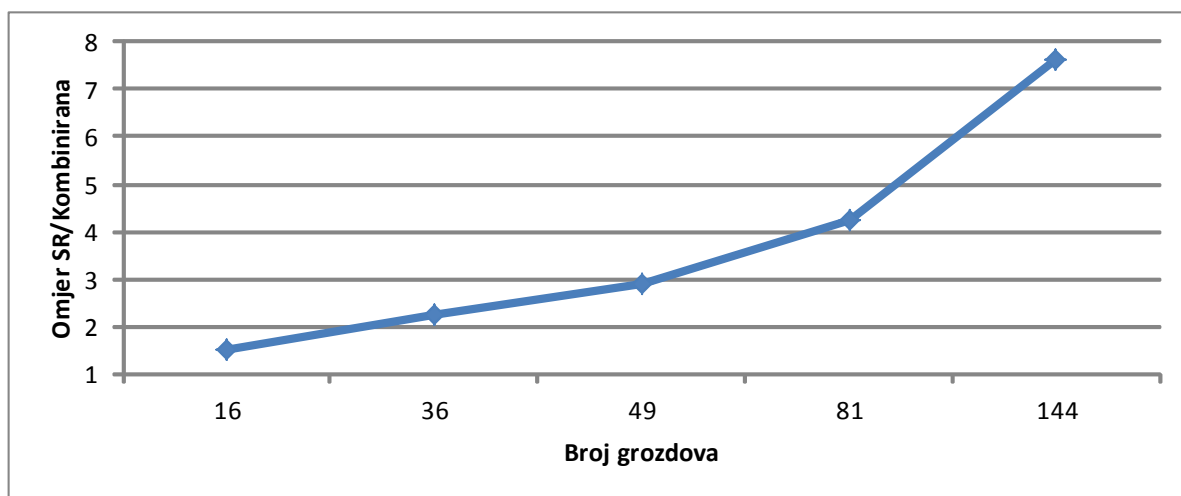
Slika 4.15: Grafički prikaz vremena trajanja razvrstavanja za sve tri metode pri promjeni broja grozdova.

utječe na povećanje vremena kod MinMax metode. Kako Kombinirana metoda iskoristava prednosti obje metode, tako promjena broja grozdova ne utječe na nju, a opet ima kraće vrijeme izvođenja jer odbacuje više grozdova. Gotovo eksponencijalni porast vremena izvođenja kod metode simetralnih ravnina može se objasniti svim potrebnim izračunima pri povećavanju broja grozdova. Na slikama 4.16 i 4.17 prikazani su omjeri vremena potrebnog da se proces razvrstavanja završi. Na slici 4.16 prikazan je omjer MinMax i Kombinirane metode, a slici 4.17 omjer metode simetralnih ravnina i Kombinirane metode. Sa slike 4.16 vidljiv je približno konstantni omjer vremena izvođenja, osim za slučaj s 16 grozdova, dok je na sljedećoj slici vidljiv gotovo eksponencijalni rast koji je već primijećen na slici 4.15.

Što je broj grozdova veći, grozdovi su i međusobno bliži te je manja vjerojatnost za

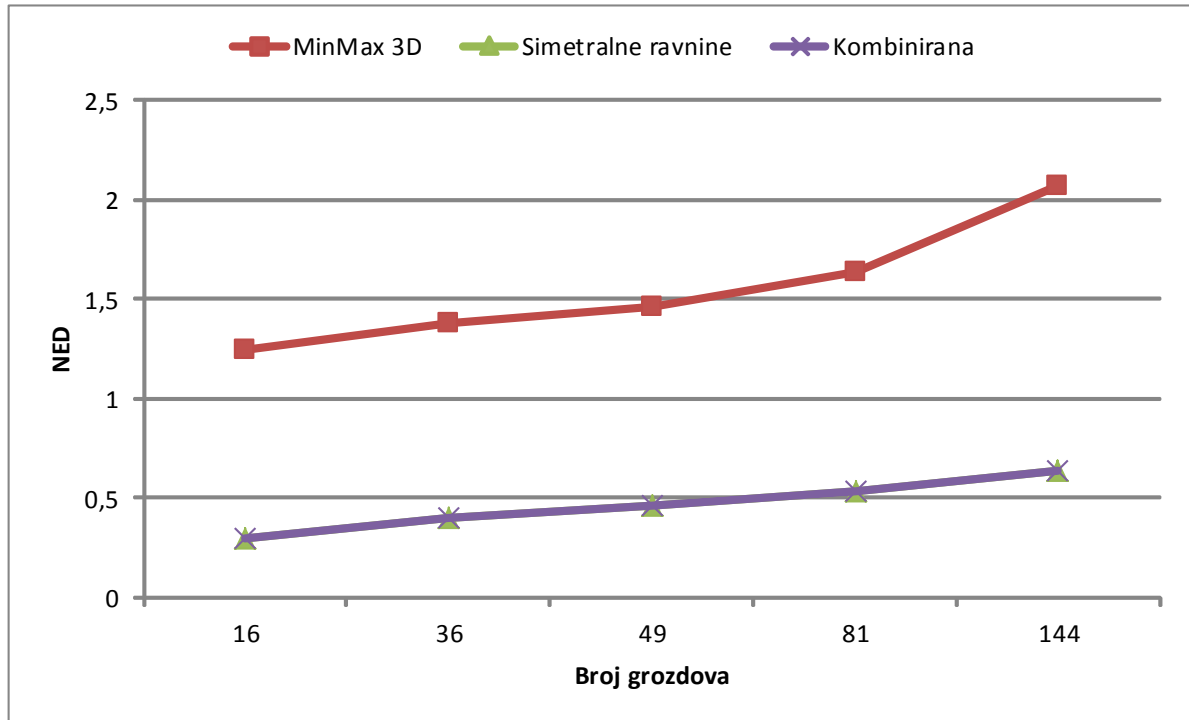


Slika 4.16: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode pri promjeni broja grozdova.



Slika 4.17: Grafički prikaz omjera vremena izvođenja metode simetralnih ravnina i Kombinirane metode pri promjeni broja grozdova.

uspješnim odbacivanjem grozdova. To se najbolje vidi u povećavanju NED-a s povećavanjem broja grozdova, prikazano na slici 4.18. Najveći NED ima MinMax metoda pri pokusu sa 144 grozda. U tom slučaju prosječni NED je jednak 2,07 što je i dalje veliko poboljšanje u odnosu na *uk-means* metodu kojoj je NED jednak 144.



Slika 4.18: Prikaz promjene NED-a za sve tri metode pri promjeni broja grozdova.

#### 4.4. Zaključak o provedenim pokusima

U ovom poglavlju provedeni su pokusi koji uspoređuju vremena potrebna za razvrstavanje objekata s prostornom neodređenosti u sve tri dimenzije geometrijskog prostora  $\mathbb{R}^3$  za MinMax metodu, metodu simetralnih ravnina i novo predloženu Kombiniranu metodu. Prvo je odabran osnovni skup parametara i provedeni su pokusi nad njim za sve tri metode. Zatim su provedeni pokusi u kojima se mijenja jedan od parametara, dok ostali parametri zadržavaju vrijednosti iz osnovnog skupa. U svim pokusima Kombinirana metoda imala je najmanje vrijeme trajanja procesa razvrstavanja uz najmanji NED. Prikazanim omjerima između Kombinirane i ostalih metoda dobivena su smanjenja vremena trajanja razvrstavanja od 1,20 do čak 19,50 puta.

Iz prikazanih rezultata vidi se da Kombinirana metoda koristi prednosti obje metode te na taj način, prvo, vrlo brzo, odbaci dio udaljenih grozdova koristeći prednosti jednostavnog računanja euklidske udaljenosti (potpoglavlje 4.3.2) te kada ostane mali broj grozdova (potpoglavlje 4.3.5) ostale odbaci korištenjem metode simetralnih ravnina i velikog NED-a koji je njezina prednost. Tako se smanjuje ukupno vrijeme potrebno za razvrstavanje objekata s neodređenosti jer se u prvom dijelu ove metode odbacuju grozdovi na jednostavniji način (nego u metodi simetralnih ravnina). U drugom dijelu se odbacuju dodatni grozdovi (koji ne bi bili odbačeni MinMax metodom) te time skraćuje vrijeme koje bi inače bilo potrošeno na izračunavanje ED-a.

## 5. MODEL SKLADIŠTA ZA VERIFIKACIJU RAZVIJENE NOVE METODE RAZVRSTAVANJA

Metode razvrstavanja objekata koje su predstavljene i opisane u prethodnom poglavlju mogu se upotrijebiti u različitim sustavima. Objekti takvog sustava, koji se analizira, sadrže neku vrstu neodređenosti i predstavljaju problem pri razvrstavanju klasičnim metodama. Razvrstavanjem objekata sustava nastoji se otkriti korisne podatke te ih grupirati u grozdove na takav način da su objekti unutar istog grozda, po bitnim svojstvima, međusobno sličniji od objekata izvan grozda. Otkrivanjem korisnih podataka može se opisati trenutno stanje i predvidjeti buduće stanje sustava. Analizom različitih slučajeva koji su se dogodili u sustavu, tijekom vremena, moguće je predvidjeti buduća stanje sustava u sličnim slučajevima. Ta predviđanja mogu biti korisna u pripremi sustava kako bi što djelotvornije mogao odgovoriti na različite zahtjeve u tijeku rada.

U ovoj disertaciji promatrani sustav je skladište robe. Za to skladište napravljena je baza podataka koja sadrži podatke o lokaciji proizvoda u tom skladištu. U bazi su spremljene sve koordinate paleta na kojima se nalaze proizvodi, broj i veličina proizvoda te način i redoslijed pristupa svakom proizvodu. Također, spremljene su lokacije i mogući smjerovi gibanja skladišnih poslužitelja. Zbog lakšeg razvrstavanja palete su podijeljene na uzorke veličine proizvoda koji se na njima nalaze i tako predstavljaju minimalno područje neodređenosti, tj. MBB jednog objekta. Veličina svakog MBB-a određena je brojem i veličinom pojedinog proizvoda koji se nalazi na paleti. Neodređenost proizlazi iz činjenice da veličina i broj proizvoda na različitim paletama nisu jednaki. Svaka paleta predstavlja objekt s neodređenosti u ovom sustavu. Svaki proizvod na paleti predstavlja jedan uzorak funkcije gustoće vjerojatnosti – PDF, a broj proizvoda na paleti određuje broj uzoraka kojim je PDF predstavljen. Vjerojatnost da se objekt nalazi u nekom uzorku jednaka je vjerojatnosti postojanja proizvoda u tom uzorku.

U ovoj disertaciji napravljen je model skladišta kako bi se koristio za ispitivanja. Model skladišta sadrži bazu podataka u koju se spremaju podaci o proizvodima, narudžbama i zadacima sustava. Podaci iz baze podataka mogu se analizirati u svrhu smanjenja vremena potrebnog za formiranje narudžbe ili reorganizaciju razmještaja paleta u skladištu zbog smanjenja vremena potrebnog za formiranje narudžbe. Na osnovu postojećih podataka o lokaciji nekog proizvoda razvrstavanjem se može dobiti informacija koji je poslužitelj najpogodniji za preuzimanje narudžbe i napraviti predviđanje stanja sustava nakon njezinog izvršenja. Pomoću predviđanja pretpostaviti će se buduće potrebe sustava, što može za

rezultat imati znatnu uštedu sredstva za održavanje sustava i vremena potrebnog kako bi se obavila neka narudžba.

Početne informacije u bazi podataka opisuju trenutno stanje u skladištu. Razvrstavanjem proizvoda u grozdove, čija su središta poslužitelji, može se pronaći onaj poslužitelj koji će u najkraćem vremenu pružiti naručeni proizvod. S druge strane kada nema narudžbe, razvrstavanjem postojećih podataka o lokaciji proizvoda mogu se odrediti lokacije na kojima trebaju biti središta grozdova, tj. poslužitelji. Tako sustav može iskoristiti predviđanje i rasporediti poslužitelje u središte grozdova kako bi što djelotvornije mogli umanjiti buduće vrijeme reakcije za neku narudžbu.

Cilj ovog poglavlja napraviti je model skladišta i u njemu upotrijebiti novo predstavljenu Kombiniranu metodu za razvrstavanje objekata s prostornom neodređenosti i usporediti je s drugim metodama. Cilj je pokazati da će se upotrebom Kombinirane metode razvrstavanja smanjiti vrijeme odabira poslužitelja koji je u mogućnosti najefikasnije preuzeti narudžbu kao i vremena pozicioniranja istih poslužitelja, kako bi se smanjilo potrebno vrijeme reakcije prilikom sljedeće narudžbe. U sljedećim poglavljima bit će izloženi različiti sustavi skladišta, model odabranog skladišta i rezultati sustavne analize prostorno-vremenskih značajki sustava. Pod vremenskim značajkama podrazumijevaju se različiti scenariji u kojima se mijenjaju zahtjevi u sustavu. Budući da su potrebe u sustavu ovisne o različitom dobu dana, danima u tjednu ili tijekom godine predstavljeni su scenariji normalnih i povećanih zahtjeva u skladištu. Na temelju provedenih analiza predložiti će se načini za poboljšanje djelovanja sustava za prikupljanje narudžbe.

## 5.1. Logistika skladišta

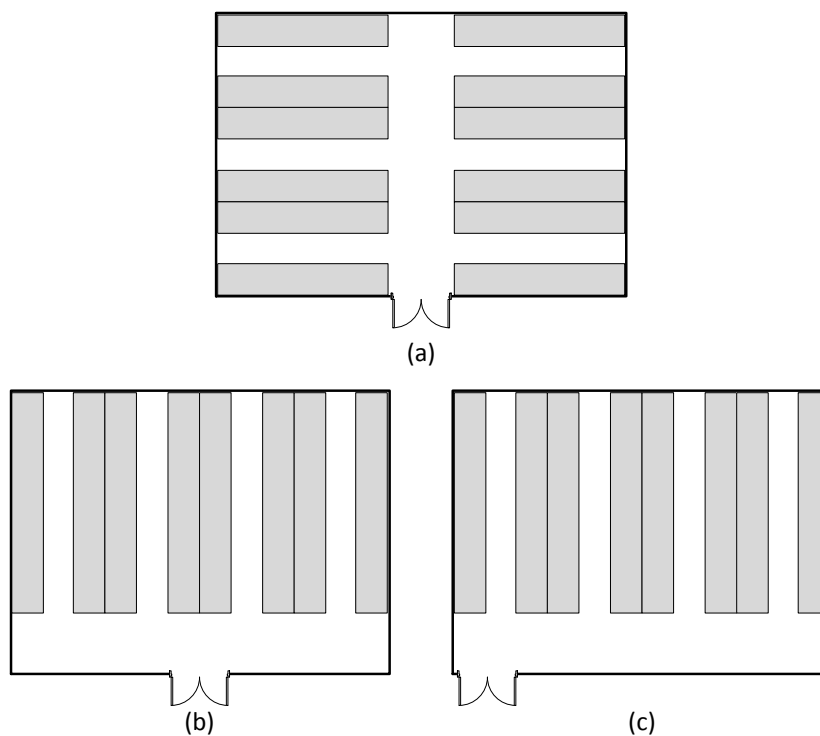
Skladište, u širem smislu, ograđeni je ili neograđeni prostor te pokriveni ili nepokriveni prostor koji se koristi za čuvanje sirovina, poluproizvoda ili gotovih proizvoda kojima se odvijaju poslovni procesi u poduzeću [51]. Skladišta su jedan od najvažnijih dijelova mreže logističkog lanca (lanca nabave). S logističkog stajališta skladište je čvor u logističkoj mreži na kojem se roba prihvaća i/ili prosljeđuje u nekom drugom smjeru unutar mreže. Logistika je proces planiranja, implementacije i kontrole efikasnog i efektivnog tijeka i skladištenja materijala, usluga i povezanih informacija od točke izvora do točke potrošnje u svrhu zadovoljenja zahtjeva kupca [52]. Ciljevi skladištenja dijele se na:



- osiguranje kontinuiteta procesa reprodukcije tako da se sve odvija planski i sustavno po unaprijed navedenom redu,
- omogućavanje kontinuitet procesa reprodukcije skladištenja uz najniže troškove skladištenja i uz najmanja moguća financijska sredstva angažirana u zalihe,
- sprječavanje rasipavanja, kvara, loma i ostalih gubitka na vrijednosti zaliha,
- održavanje stalne kvalitete zaliha robe čuvanjem, zaštitom i održavanjem fizičko-kemijskih svojstava robe,
- racionalno ubrzavanje tijekom robe kako bi se proces poslovanja skratio i tako ubrzao koeficijent obrtaja sredstava vezanih za zalihe i
- svojim poslovanjem utjecati na povećanje konkurentne sposobnosti radne organizacije.

Iz ovih ciljeva vidljivo je da za skladište, ali i cijeli proces nabave, efikasnost, tj. što kraće vrijeme isporuke robe, jedan od glavnih prioriteta.

Mnogi čimbenici imaju utjecaj na efikasnost poslovanja skladišta, no jedan od glavnih čimbenika je razmještaj (eng. Layout) robe u skladištu. Skladište s dobrim razmještajem može značajno smanjiti cijenu poslovanja. Skladišta su većinom pravokutnog tlocrta i obično je razmještaj robe u njima podijeljen u tri tipa [53]. Prvi tip čine dva bloka regala s glavnim prolazom koji vodi do glavne ulazno/izlazne točke skladišta, kao što je prikazano na slici 5.1(a). Drugi i treći tip imaju samo jedan blok regala koji su smješteni okomito u odnosu na



Slika 5.1: Tri tipa razmještaja u skladištu.

glavnu ulazno/izlaznu točku skladišta, s tim da je jednom tipu ta točka smještena na sredini, a drugom u donjem lijevom kutu, kako je prikazano slikama 5.1(b) i 5.1(c).

### 5.1.1. Razne podjele skladišta

Aktivnosti koje se obavljaju u skladištu dijele se u četiri najvažnije:

- prijam,
- skladištenje,
- komisioniranje i
- izdavanje robe.

Aktivnost prijama (eng. Receiving) počinje kada roba stigne u skladište na mjesto pružanja robe. Sljedeća je aktivnost skladištenje robe (eng. Storage), tj. transport do mjesta odlaganja i odlaganje robe unutar prostora skladišta. Komisioniranje ili proces odabira narudžbe (eng. Order Picking) proces je pronalaženja i izuzimanja određenih (naručenih) proizvoda iz skladišta u određenim količinama koje je odredio kupac. I na kraju, aktivnost izdavanja robe (eng. Shipping) bavi se premještanjem robe od mjesta odlaganja u skladištu do odjela za pakiranje i slanje gdje se komisionirana narudžba spaja u jedinicu za otpremu. Ova disertacija analizira aktivnosti procesa komisioniranja i skladištenja. U ovom poglavlju fokus će biti na ubrzavanju odluke koji će od poslužitelja biti najbolji izbor za izvršenje komisioniranja, a u sljedećem poglavlju bit će dani prijedlozi kako izmijeniti razmještaj robe u skladištu kako bi se još više ubrzao proces komisioniranja.

Skladišta se mogu podijeliti i prema stupnju mehanizacije na:

- niskomehanizirana skladišta,
- viskomehanizirana skladišta,
- automatizirana skladišta i
- robotizirana skladišta

Niskomehanizirana skladišta ili tzv. klasična podrazumijevaju da zaposlenici rukuju robom na najjednostavnije načine, kao što su npr. ručna kolica. U viskomehaniziranim skladištima zaposlenici koriste različite uređaje kao što su viljuškari, skladišna dizala i sl. Skladišta gdje su zaposlenici zaduženi samo za poslove nadzora nad poslovnim procesima koji se odvijaju pomoću računala, a upravljanje robom je automatizirano, nazivaju se automatizirana skladišta. I na kraju najviši stupanj mehanizacije su robotizirana skladišta. To su skladišta gdje se upravljanje robom vrši pomoću robota i računala, bez ikakve uloge zaposlenika.

Po vrsti robe koja se skladišti skladišta se mogu podijeliti na:

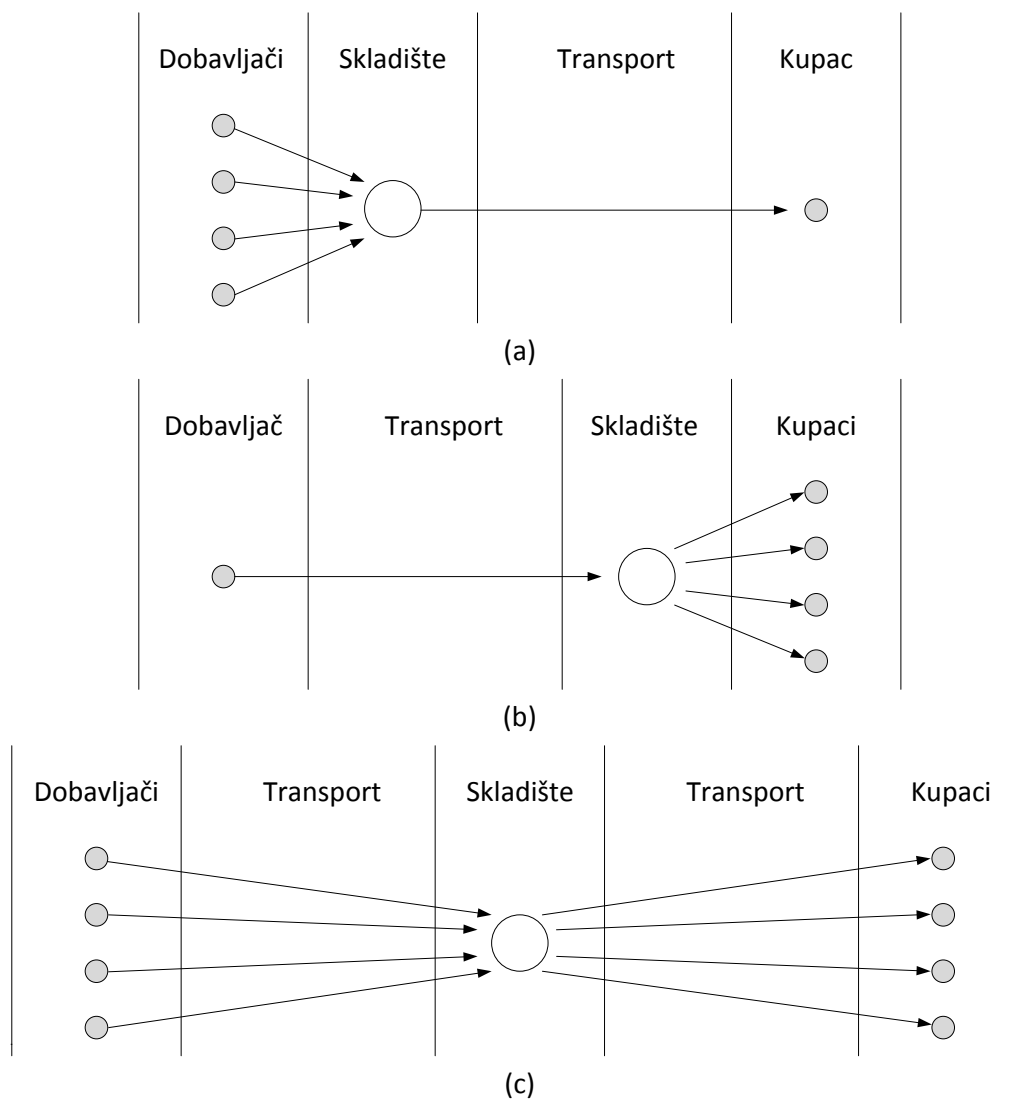
- skladištenje sirovina,
- skladištenje poluproizvoda,
- skladištenje gotovih proizvoda,
- skladištenje alata i
- skladištenje rezervnih dijelova.

Još jedna važna podjela skladišta je prema njihovoj ulozi u logističkom lancu:

- mjesto objedinjavanja transporta - konsolidacija (eng. Consolidation) slika 5.2(a)
- mjesto objedinjavanja transporta - pregrupiranje (eng. Break-bulk) slika 5.2(b)
- mjesto miješanja proizvoda (eng. Mixed warehouse, make-bulk/break-bulk) 5.2(c)

Podjela skladišta po izvedbi objekta:

- horizontalna ili niska (visine do 7m) i
- okomita ili visoka (visine preko 12m).



Slika 5.2: Uloga skladišta u logističkom lancu.

### 5.1.2. Skladišni prostor

U današnje vrijeme grade se visokoregalna skladišta čija visina može biti i do 50 metara. Regali su izrađeni od čelika, s policama koje je moguće prilagoditi visini robe. Izrađuju se u tri ili više etaže, a u običnim skladištima visina regala je od dva do šest metara.

Vrste regala su:

- polični regali,
- paletni regali,
- konzolni regali,
- prolazni regali,
- protočni regali,
- pokretni regali,
- prijevozni regali.

Razmak između regala ovisi o vrsti robe u skladištu. Većina robe se, u današnje vrijeme, dostavlja na euro-paletama. Te palete imaju dimenzije  $120 \times 80 \times 14$  cm, prikazano na slici 5.3, izrađene su većinom od drveta i pružaju mogućnost lakog transporta proizvoda koji se na njima nalaze. Skladištenje takvih paleta obavlja se na paletnim regalima, a između dvije palete u skladištu treba ostaviti razmak od 10 cm.

Pri projektiranju zgrade, u kojoj će se nalaziti skladište, važno je držati se uputa. Zbog opasnosti od požara potrebno je osigurati 60 cm široke prolaze između složene robe i zida. Širinu prolaza treba prilagoditi načinu manipuliranja robom u skladištu. Svi pregradni zidovi trebaju biti montažni, a nosači što niži i čelični. Vrata i prozori sigurni i zaštićeni zbog



Slika 5.3: Euro paleta.

mogućih provala. Osim ovih projektantskih smjernica, kod projektiranja skladišnog prostora jako je važno znati na koji način će biti pakiran veći dio robe kako bi se što više povećala iskoristivost i efikasnost skladišta. Potrebno je na sve načine smanjiti troškove skladištenja. To se postiže brzim rukovanjem robom, očuvanjem vrijednosti robe, zaštitom od krađe, brzim snalaženje i fleksibilnošću u rasporedu robe.

Osim regala u skladištu susrećemo i sljedeću opremu:

- oprema za odlaganje materijala,
- transportna sredstva,
- pomoćna skladišna oprema i
- dodatnu opremu.

Pod opremu za odlaganje materijala osim regala spadaju i različiti ormari, postolja i sl. Transportna sredstva mogu biti ručna (kao što su različite prikolice i kolica ili ručni viljuškari), motorna (viličari ili vozila za prijevoz), zatim različiti transporteri (s trakama, valjcima ili elevatori) ali i poluautomatski ili automatski uređaji za slaganje i izuzimanje robe s palete. U pomoćnu skladišnu opremu ubrajaju se informacijsko-komunikacijski sustav, sredstva za rastavljanje tereta, oprema za mjerenje (kao što su vage, posude i sl) i oprema za pakiranje. Dodatnu opremu u skladištu čine protupožarni uređaji, sigurnosno-zaštitni uređaji, uređaji za klimatizaciju, sanitarno-higijenski uređaji, uređaji za održavanje čistoće i sredstva zaštite na radu [52].

U ovoj disertaciji cilj je iskoristiti i poboljšati informacijsko-komunikacijska sustav skladišta. Kvalitetan skladišni sustav informacija i komunikacijskih veza važan je za brzinu obrtaja uskladištene robe, iskorištenost skladišnih kapaciteta, smanjenje razine skladišnih troškova i kvalitetu isporuke. Načini primjene računala u skladištu su u svrhu osiguranja potrebnih informacija o redosljedu narudžbi, lokaciji proizvoda, izuzimanja narudžbi iz skladišta i kontroli pokreta automatiziranih strojeva.

### **5.1.3. Pozicioniranje pojedine robe u skladištu**

Postoje propisana pravila o pozicioniranju robe u skladištu. Tako, ako se u skladištu nalazi zapaljiva, eksplozivna ili otrovana roba mora se skladištiti u specijalnim skladištima. Zatim su tu pravila o rukavanju s robom koja traži posebnu njegu, lomljiva je ili lako pokvarljiva. No tu se nalaze i pravila kao što su:

- teža roba stavlja se na niže, a lakša na više police regala,
- robu koja se učestalije izdaje smjestiti bliže mjestu izdavanja,

- na ulazu postaviti plan skladišnog prostora i
- roba se skladišti prema nomenklaturi i svaki prostor u skladištu potrebno je označiti.

Prema [54] postoje tri sustava pozicioniranja robe u skladištu:

1. Svaka pojedina roba ima točno određeno mjesto u skladištu.
2. Kombinacija fiksnog i slobodnog korištenja mjesta unutar određenog prostora.
3. Kaotično skladištenje.

U prvom sustavu pozicioniranja robe u skladištu dodijeljuju se fiksno mjestu u skladištu gdje se onda skladište sve količine te robe. Koje mjesto u skladištu odabrati ovisi o određenim kriterijima kao što su: količina isporuke, koeficijent obrtaja, količine zaliha itd. Odluku donosi skladištar ili računalo, tj. informacijsko-komunikacijski sustav skladišta. Pri odabiru lokacija za ovaj način pozicioniranja robe posebno je važno paziti na prostor za maksimalne zalihe. Ukoliko je taj prostor nedovoljno velik da primi novu robu, onda se višak smješta na udaljeno, neraspoređeno mjesto. Na taj način roba više nije pod kontrolom i ova vrsta pozicioniranja robe gubi smisao. To znači da u skladištu uvijek treba imati prostor koji se može iskoristiti u slučaju proširenja asortimana. Nakon što se popuni i sav taj prostor, treba napraviti novu podjelu i presložiti robu. Ovaj način pozicioniranja robe često se temelji na iskustvu i uspješno se primjenjuje na male asortimane robe, npr. skladište rezervnih dijelova za vlastite potrebe. Prednost ovog sustava pozicioniranja je u tome što je robu moguće brzo pronaći i bez upotrebe računala ili drugih organizacijskih sredstava, a nedostatak je što se relativno slabo koristi raspoloživi skladišni prostor.

Drugi sustav pozicioniranja je kombinacija fiksnog i slobodnog korištenja mjesta unutar određenog prostora. Skladište se podijeli na dio za fiksno skladište pojedinih vrsta robe i manji dio koji se koristi za slobodno pozicioniranje robe. Prednosti ovog sustava je da se roba s velikim koeficijentom obrtaja može smjestiti bliže izlazu pomoću slobodnog raspoređivanja, bez obzira na vrstu robe.

Zadnji sustav pozicioniranja je kaotično skladištenje ili potpuno slobodno korištenje prostora. Primjenjuje se u velikim skladištima gdje računalo upravlja smještanjem i izdavanjem robe. Ovaj način pozicioniranja moguć je jedino uz potpuno razrađen i proveden sustav označavanja skladišnih mjesta, popraćen s adekvatnom skladišnom opremom. Raspored robe provodi se prema svim kriterijima racionalnog smještaja, što znači da se roba iste vrste može rasporediti na različita skladišna mjesta. Prednosti ovog sustava skladištenja u odnosu na fiksno pozicioniranje su brojne. Kao prvo ne postoji potreba za pričuvnim prostorom niti potreba za premještanjem robe. Mjesto skladištenja može se odrediti prema potrebama u

trenutku skladištenja i zbog računalne evidencije upravljanje procesima skladištenja i izdavanja robe jednostavno je, te je rizik neovlaštenog i neevidentiranog izdavanja robe vrlo malen. Može se zaključiti da je iskoristivost skladišnog prostora bolja, nego kod drugih sustava pozicioniranja robe. Dok su nedostaci prepozanti u tome što ako nije poznata oznaka robe ili oznaka mjesta skladištenja robu nije moguće pronaći u kratkom vremenu, a proces skladištenja i izdavanja robe ne mogu se ekonomično odvijati bez računala.

## 5.2. O modelu skladišta

Kako bi se razvijena Kombinirana metoda mogla usporediti s ostalim metodama u problemima razvrstavanja objekata s neodređenosti u nekom skladištu bilo je potrebno napraviti model skladišta nad kojim će se izvoditi pokusi. Istraživanjem problematike skladištenja odlučeno je da se napravi model skladišta sa sljedećim karakteristikama:

- razmještaj u dva bloka regala s glavnim prolazom koji vodi do glavne ulazno/izlazne točke skladišta,
- prema stupnju mehanizacije ovaj model odgovara robotiziranom skladištu,
- vrsta robe koja se skladišti su gotovi proizvodi,
- proizvodi se nalaze na euro paletama i smještaju na paletne regale,
- po izvedbi objekta radi se o horizontalnom skladištu do 7m visine,
- po ulozi u logističkom lancu ovo skladište može bit mjesto objedinjavanja transporta ili mjesto miješanja proizvoda,
- sustav pozicioniranja robe koji se koristi u modelu skladišta je kaotično skladištenje.

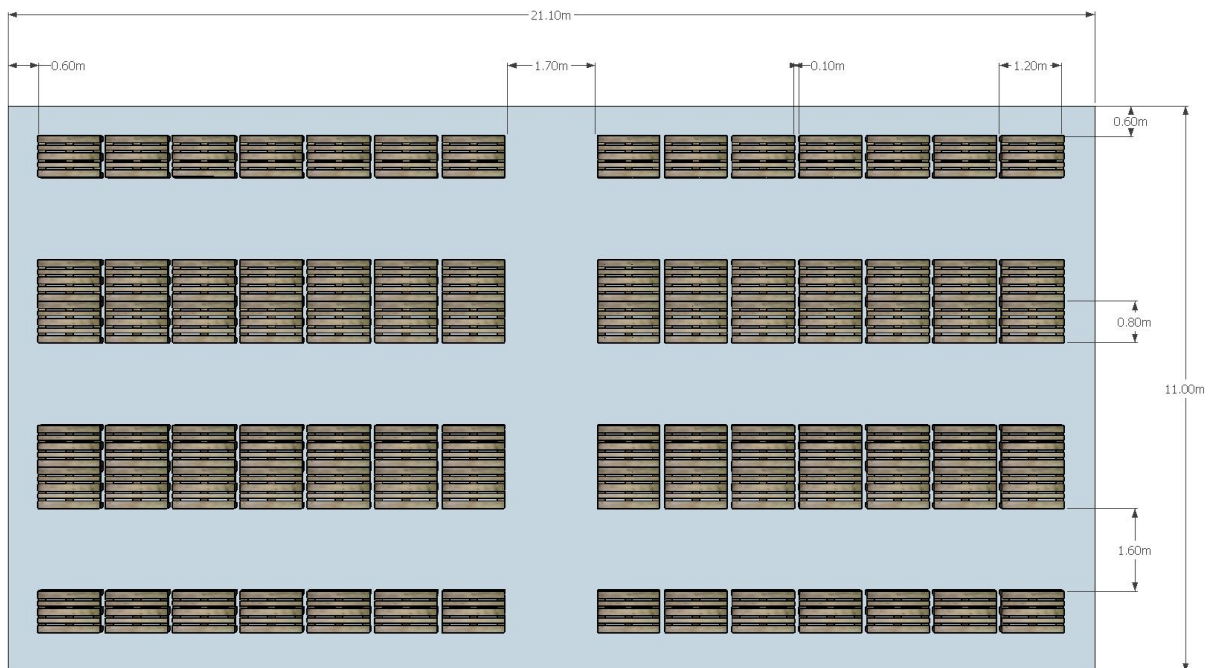
### 5.2.1. Dimenzije i kapacitet skladišta

Odabrani model skladišta je pravokutnog tlocrta, prikazan na slici 5.4, podijeljen u dva paralelna bloka podijeljena glavnim prolazom sa četiri reda paletnih regala (dva jednostruka i dva dvostruka). Unutarnja površina skladišta je  $21,1 \times 11$  m, a visina skladišta je 7 m. Svaki regal ima tri nivoa i duljina mu je jednaka duljini reda, kao što je prikazano slikom 5.5. U svaki nivo regala stane sedam paleta, što čini ukupno 21 paleta po regalu. U cijelom skladištu ima 12 regala, tj. kapacitet skladišta je 252 palete. Svaki regal je širok 0,8 m i dugačak 9,1 m. Visina regala od poda je 6 m, s policom debljine 5 cm na svakih 1,95 m. Svaki regal se nalazi na udaljenosti od 0,6 m od najbližeg zida skladišta. U regale se odlažu proizvodi na

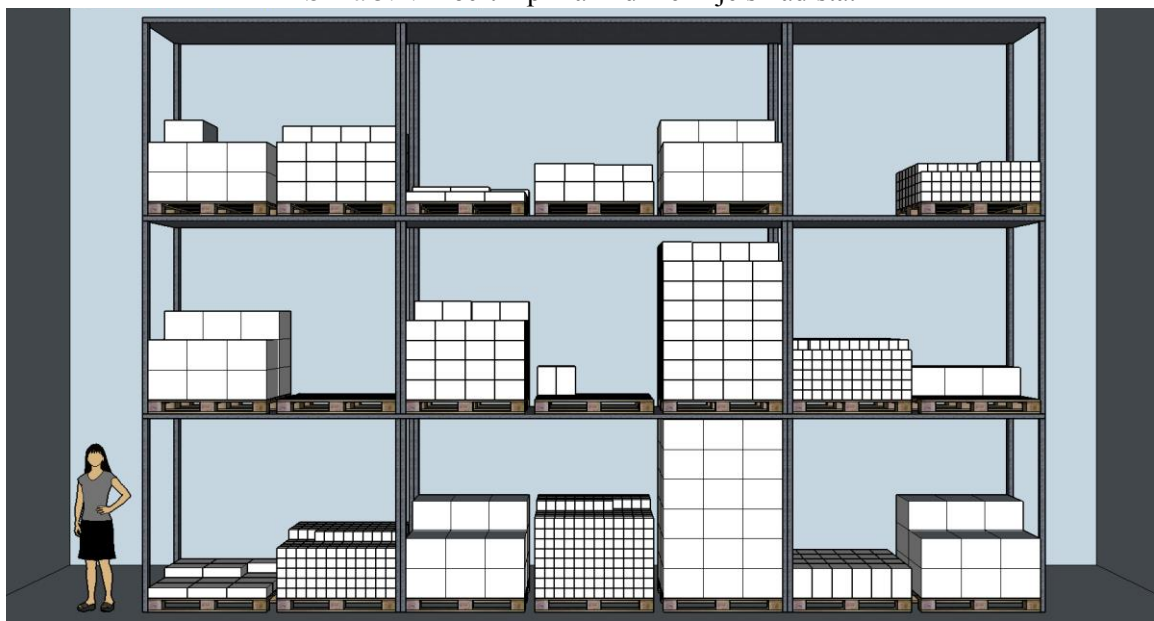
standardnim euro-paletama dimenzija  $120 \times 80 \times 14 \text{ cm}$  do visine od  $180 \text{ cm}$ . Glavni prolaz širok je  $1,7 \text{ m}$ , dok su prolazi između paletnih regala široki  $1,6 \text{ m}$ , za neometano mimoilaženje dva poslužitelja. Razmak između dviju susjednih paleta je  $10 \text{ cm}$ .

### 5.2.2. Proizvodi i poslužitelji

Broj proizvoda na paleti ovisi o njihovoj veličini. Zbog jednostavnosti odabrani proizvodi su paketi s cjelobrojnim veličinama stranice od  $10 \text{ cm}$  do  $40 \text{ cm}$ , no u njihovom



Slika 5.4: Tlocrtni prikaz i dimenzije skladišta.



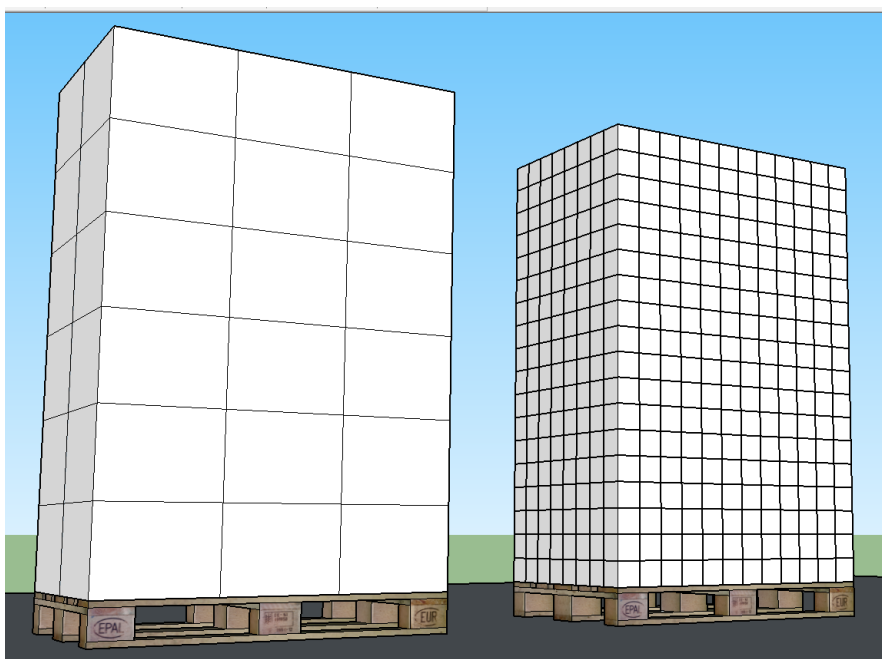
Slika 5.5: Skica jednog paletnog regala s tri nivoa.



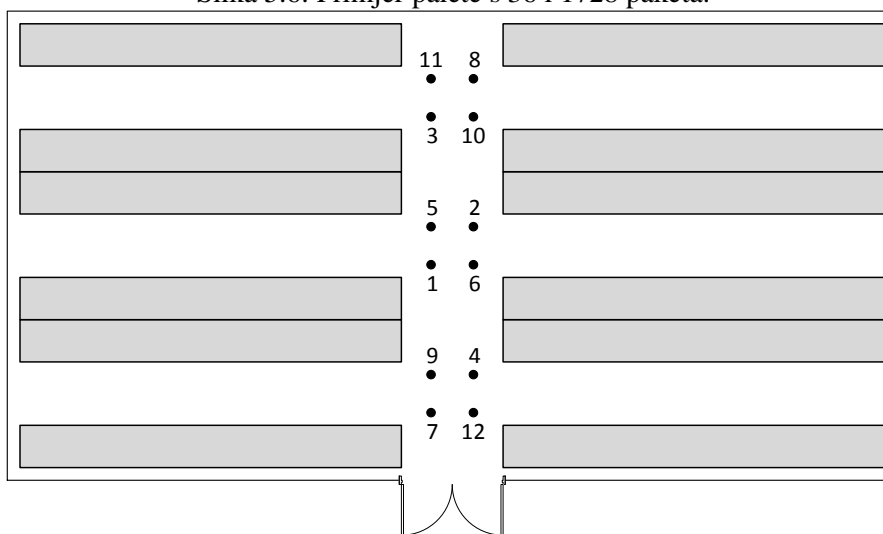
kreiranju u modelu se zapravo koriste omjeri sa stranicama palete. Tako po dužini paleta može imati 3, 4, 6 ili 12, po širini 2, 4 ili 8, a u visini 6, 9 ili 18 paketa. Kombinacijom ovih omjera može se dobiti ukupni broj proizvoda, a to je od 36 do 1728 proizvoda po paleti, kao što je prikazano na slici 5.6.

U bazi podataka, koja je napravljena za ovaj model, spremljene su koordinate svih paleta u skladištu. Svaka paleta opisana je MBB-om i brojem uzoraka koji je jednak maksimalnom broju paketa koji se mogu nalaziti na toj paleti. Broj paketa koji se i nalaze na paleti jednak je broju uzoraka kojima je vjerojatnost veća od nule. Veličina pojedinog uzorka MBB-a ovisi o veličini paketa koji se nalazi na paleti.

Za početni razmještaj u modelu odabran je slučajni odabir paleta, na kojima je slučajnim



Slika 5.6: Primjer palete s 36 i 1728 paketa.



Slika 5.7: Početni razmještaj poslužitelja.

odabirom određena veličina paketa i isto tako slučajnim odabirom određen je broj paketa koji se nalazi na paleti. Taj razmještaj sprema se kao početno stanje - PS od kojega počinje svaki pokus.

Poslužitelji u modelu zamišljeni su kao roboti, tj. automatski uređaji, za slaganje i izuzimanje robe s palete. Radi se o najvišem stupanj mehanizacije skladištu. U ovom modelu skladišta upravljanje robom vrši se pomoću robota i računala bez ikakve uloge zaposlenika. Poslužitelji primaju zadatak iz centralnog računala. Računalo prima naredžbu i na temelju trenutnog stanja u skladištu, različitim metodama, odlučuje koji je od poslužitelja najbolji kandidat za određeni zadatak. Cilj ove disertacije je razvijenim postupkom razvrstavanja smanjiti vrijeme potrebno za donošenje te odluke. Početni razmještaj poslužitelja u modelu skladišta ovisi o broj poslužitelja, a shema rednih mjesta dana je slikom 5.7. Postoje dvije vrste robota. Prva vrsta su roboti koji skupljaju proizvode narudžbe i isporučuju ih do odjela za pakiranje i slanje. Ova vrsta robota može se gibati u svim smjerovima na xy-ravnini (pri  $z = 0$ ) te ima robotsku ruku s hvataljkom koja se može gibati po z-osi koja izuzima pakete s palete. Svaki robot ima ograničeni kapacitet, tj. može prikupiti samo određeni broj proizvoda prije nego što ih isporuči odjelu za pakiranje i slanje. Ova se vrsta robota podrazumijeva kada je riječ o poslužiteljima. Druga vrsta robota su oni koji donose palete robe u skladište i odvoze prazne iz njega.

### **5.3. Verifikacija metoda u modelu skladišta**

U ovom poglavlju napravljeni su pokusi u kojima se u skladištu obavljaju uobičajeni zadaci koji su podjednako raspoređeni po čitavom skladištu. U njima se uspoređuju vremena trajanja razvrstavanja između MinMax i Kombinirane metode. Prvi je pokus razvrstavanje različitog broja poslužitelja pri PS-u. Drugi je pokus obrada narudžbe pri prvom scenariju, dok je treći pokus obrada narudžbe pri drugom scenariju.

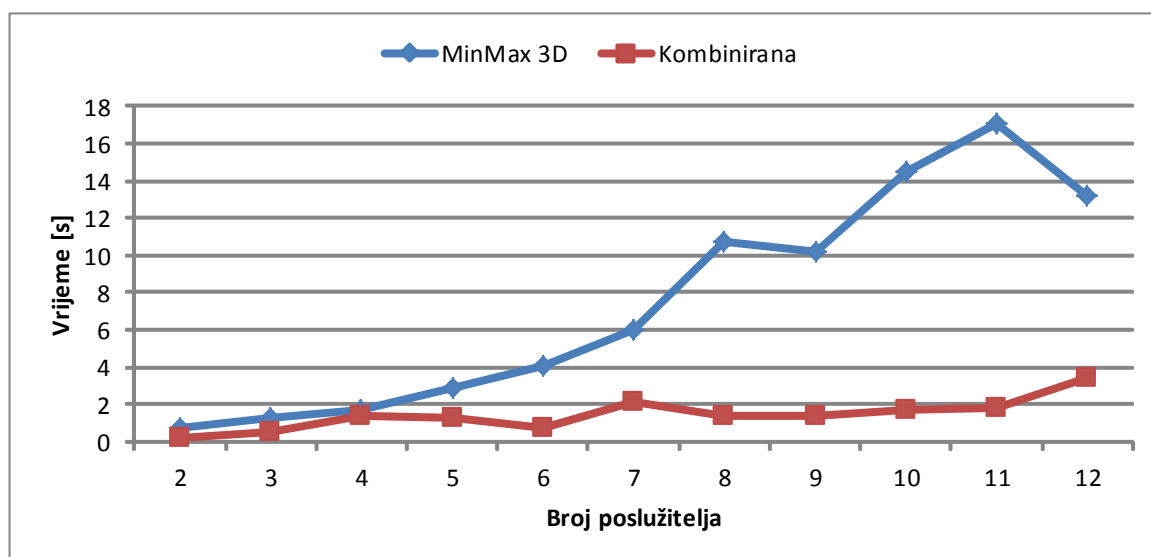
#### **5.3.1. Razvrstavanje početnih pozicija poslužitelja**

Pokusi u ovom potpoglavlju provedeni su s ciljem izračunavanja vremena trajanja procesa razvrstavanja pri različitom broju poslužitelja u modelu skladišta. Kako je objašnjeno u prethodnom potpoglavlju, raspored poslužitelja ovisi o njihovu broju. Njihove pozicije prikazane su na slici 5.7 i potrebno je prije prve narudžbe pozicionirati poslužitelje u centre grozdova, koji se izračunavaju iz PS-a. Ovaj pokus napravljen je za sve kombinacije

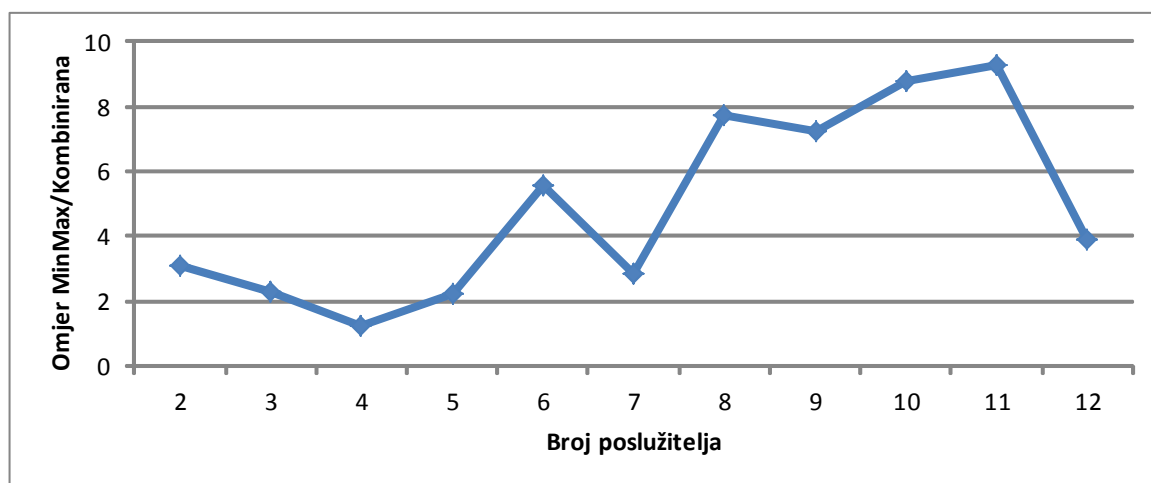
poslužitelja (od dva do dvanaest), uz nepromijenjen skup PS-a. Pokusi su izvedeni i uspoređeni su rezultati za dvije metode.

Izabrane su MinMax i Kombinirana metoda. Kako se radi o skladištu gdje se poslužitelji mogu kretati isključivo između redova paletnih regala, ne može se upotrijebiti euklidska udaljenost kao mjera različitosti, nego je upotrebljena udaljenost gradskih blokova, tj. Manhattan udaljenost. Na slici 5.8 prikazana su vremena trajanja razvrstavanja za spomenute metode, a na slici 5.9 omjer vremena razvrstavanja. Na slici 5.10 prikazana je promjena NED-a za obje metode i vidljivo je da je razlika velika, a u prosjeku je za oko 5,2 puta NED manji kod Kombinirane metode.

Kao rezultat ovog pokusa dobivene su nove „optimalne“ pozicije za smještaj poslužitelja. Na slici 5.11 dan je rezultat razmještaja za šest poslužitelja. Slika je dana kao tlocrt skladišta, točke predstavljaju najbližu moguću poziciju poslužitelja u odnosu na središta



Slika 5.8: Grafički prikaz vremena trajanja razvrstavanja za PS 2 do 12 poslužitelja.

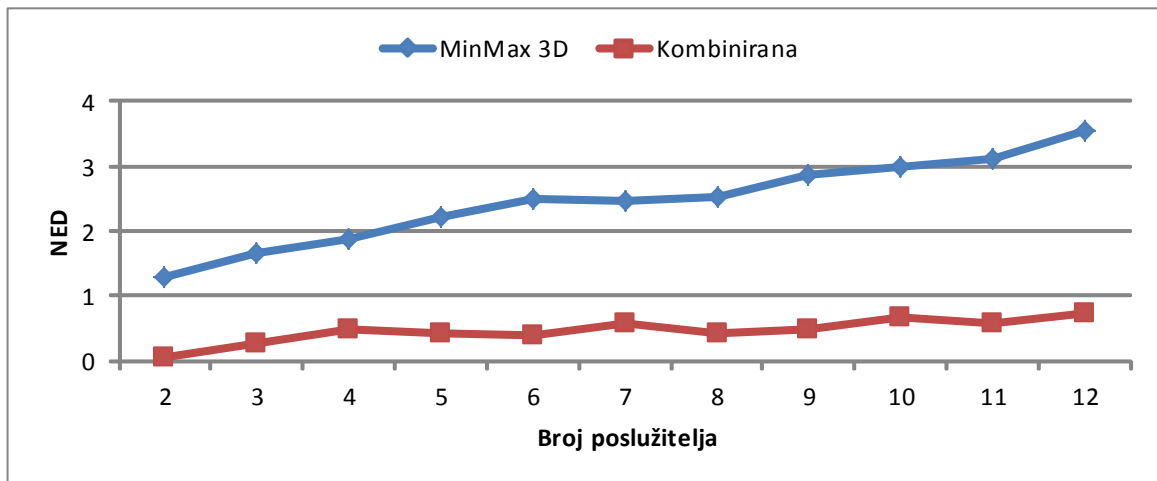


Slika 5.9: Omjer vremena trajanja razvrstavanja za MinMax i kombiniranu metodu.

grozdova u xy-ravnini. Visina na kojoj se nalazi hvataljka poslužitelja je između 291cm i 323cm.

### 5.3.2. Proces razvrstavanja pri normalnom opterećenju

U ovom poglavlju napravljeni su pokusi u kojima se u skladištu obavljaju uobičajeni zadaci, koji su podjednako raspoređeni po čitavom skladištu te su u ovisnosti o stanju na skladištu i narudžbi. Pokusi se izvode po prvom scenariju, s ciljem izračunavanja vremena trajanja procesa razvrstavanja pri različitom broju poslužitelja (grozdova) u kojem treba razvrstati objekte. Prvi scenarij osmišljen je kao normalno opterećenje skladišta gdje poslužitelji nakon svake izvršene narudžbe imaju dovoljno vremena izračunati i otići u



Slika 5.10: Grafički prikaz NED-a za PS 2 do 12 poslužitelja.

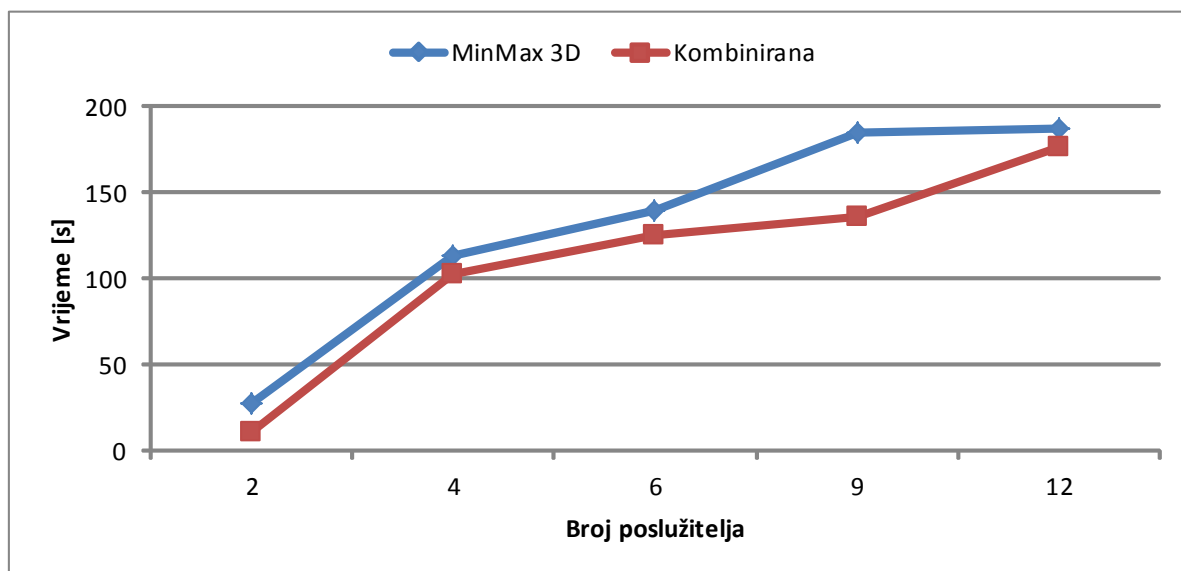


Slika 5.11: Razmještaj poslužitelja u skladištu prije prve narudžbe.

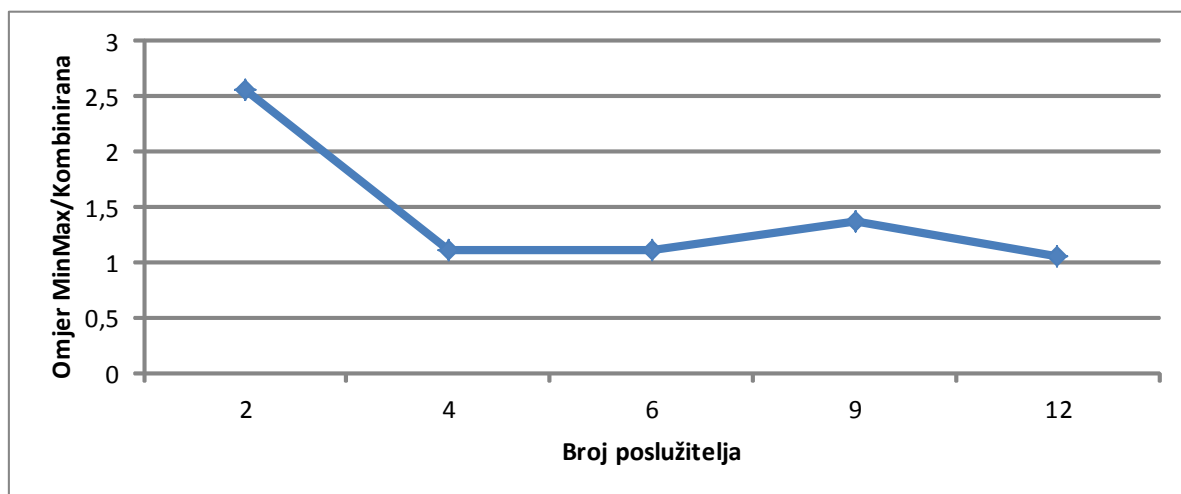
„optimalni“ položaj prije nove narudžbe. Pokusi su napravljeni za dva, četiri, šest, devet i dvanaest poslužitelja. Trajanje je svakog pokusa 100 narudžbi. Razvrstavanje je provedeno s prilagođenom MinMax i Kombiniranom metodom, za Manhattan udaljenost, i dobiveni rezultati prikazani su na slici 5.12. Sa slike je vidljivo da najkraće vrijeme izvođenja razvrstavanja ima Kombinirana metoda.

Na slici 5.13 prikazan je omjer MinMax i Kombinirane metode za prvi scenarij. Sa slike 5.13 vidljiv je približno konstantni omjer vremena izvođenja.

Što je broj grozdova veći, grozdovi su međusobno bliži te je manja vjerojatnost za uspješnim odbacivanjem grozdova. Ovo se najbolje vidi u povećavanju NED-a s povećavanjem broja grozdova, prikazano na slici 5.14. Omjer NED-a između ove dvije



Slika 5.12: Grafički prikaz vremena trajanja razvrstavanja za MinMax i kombiniranu metodu pri promjeni poslužitelja za prvi scenarij.

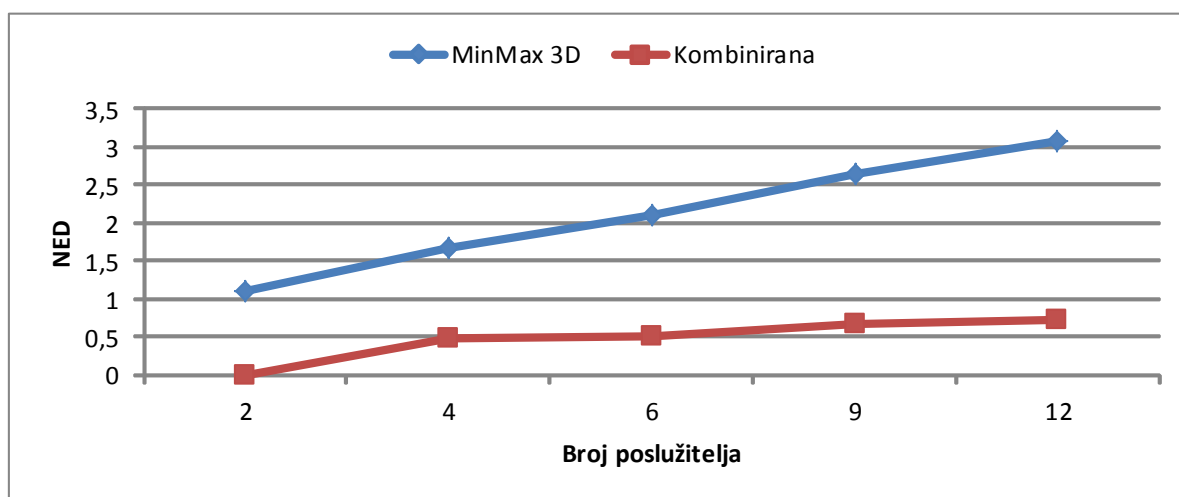


Slika 5.13: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode za prvi scenarij.

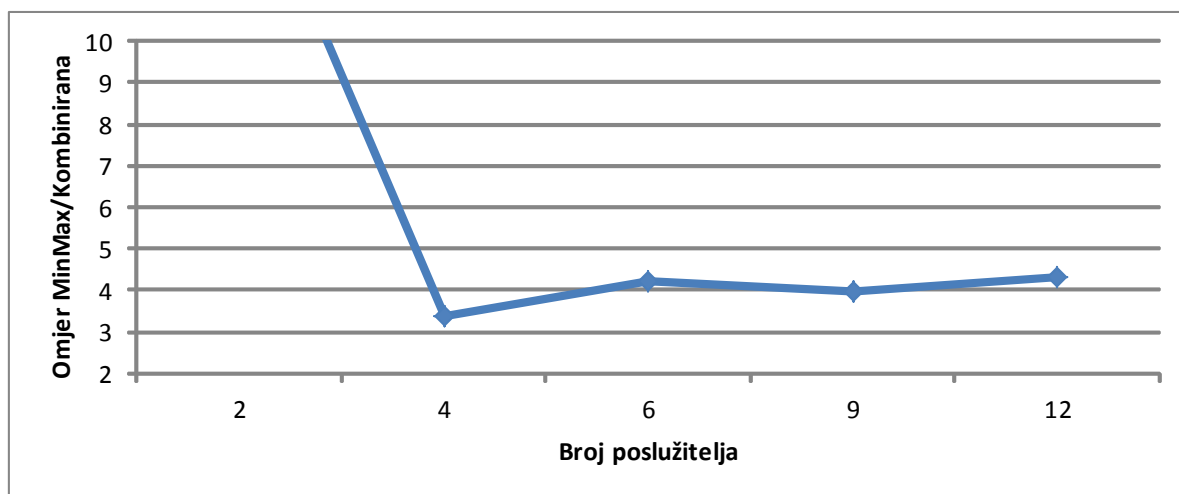
metode približno je jednak četiri, osim u slučaju s dva poslužitelja kada je NED za Kombiniranu metodu jednak nula. Kako dijeljenje s nulom kao rezultat daje beskonačan broj, na slici 5.15, rezultat s dva poslužitelja prikazan je samo kao jako veliki broj izvan prikazanog područja na grafu.

### 5.3.3. Proces razvrstavanja pri povećanom opterećenju

U ovom poglavlju napravljeni su pokusi u kojima se u skladištu obavljaju uobičajeni zadaci, koji su podjednako raspoređeni po čitavom skladištu te su u ovisnosti o stanju na skladištu i narudžbi. Pokusi se izvode po drugom scenariju, s ciljem izračunavanja vremena trajanja procesa razvrstavanja, pri različitom broju poslužitelja u kojem treba razvrstati objekte. Drugi scenarij osmišljen je kao povećano opterećenje skladišta gdje poslužitelji



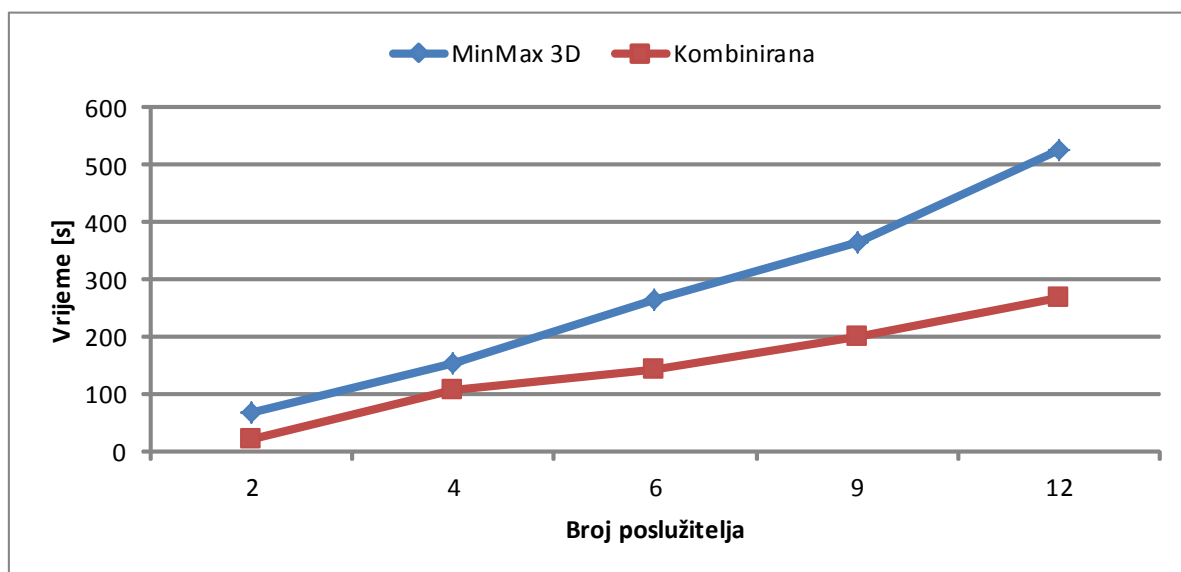
Slika 5.14: Grafički prikaz NED-a za MinMax i kombiniranu metodu pri promjeni poslužitelja za prvi scenarij.



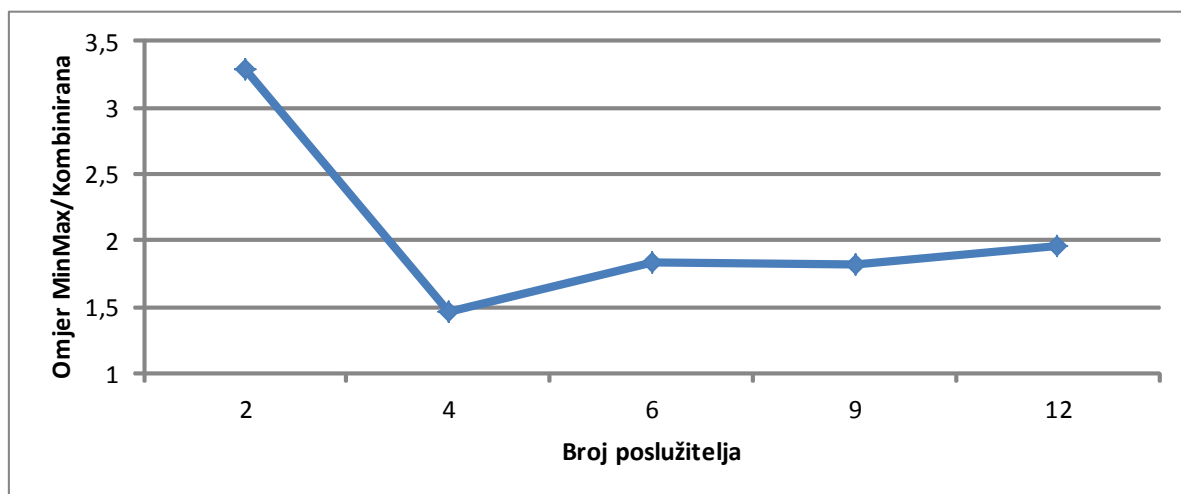
Slika 5.15: Grafički prikaz omjera NED-a MinMax i Kombinirane metode za prvi scenarij.

nakon svake izvršene narudžbe nemaju dovoljno vremena izračunati i otići u „optimalni“ položaj prije nove narudžbe, nego se razvrstavanje provodi prema trenutnim pozicijama poslužitelja. Pokusi su napravljeni za dva, četiri, šest, devet i dvanaest poslužitelja. Trajanje svakog pokusa je 100 narudžbi. Razvrstavanje je provedeno s prilagođenom MinMax i Kombiniranom metodom za Manhattan udaljenost i dobiveni rezultati prikazani su na slici 5.16. Vidljivo je kako najkraće vrijeme izvođenja razvrstavanja ima Kombinirana metoda.

Na slici 5.17 prikazan je omjer MinMax i Kombinirane metode za prvi scenarij. Sa slike 5.17 vidljiv je približno konstantni omjer vremena izvođenja i povećanje omjera u odnosu na prvi scenarij.

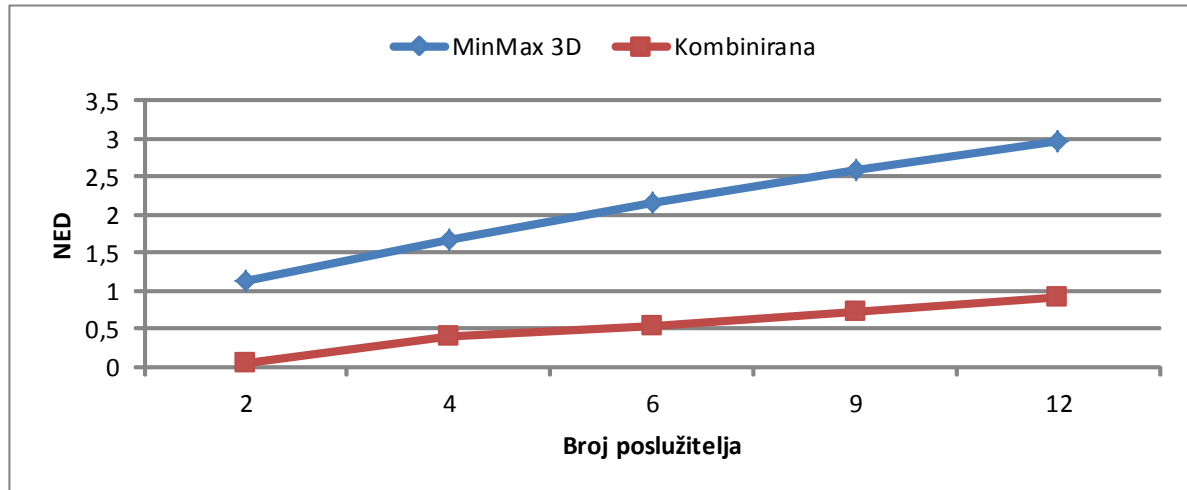


Slika 5.16: Grafički prikaz vremena trajanja razvrstavanja za MinMax i Kombiniranu metodu pri promjeni poslužitelja za drugi scenarij.

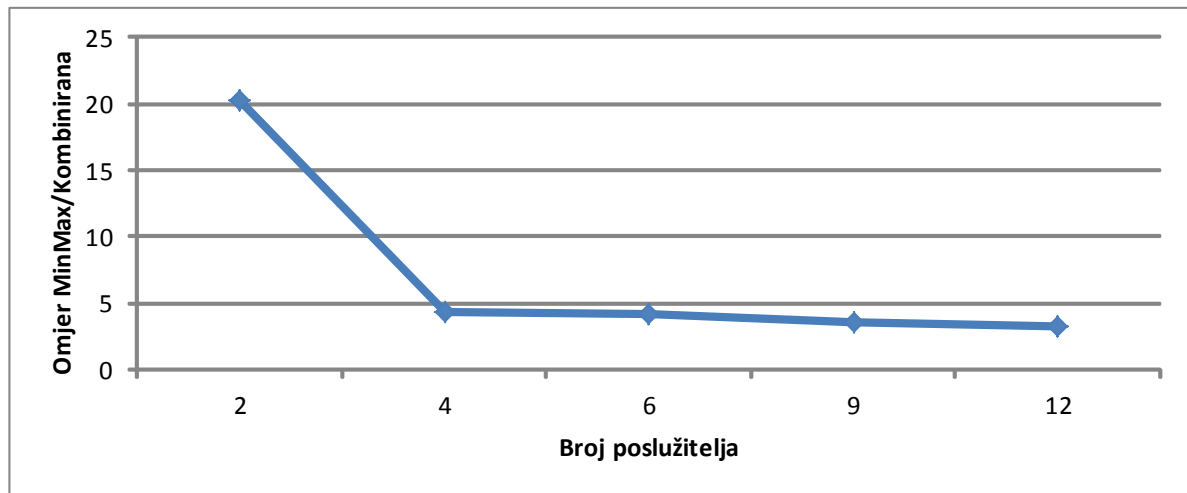


Slika 5.17: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode za drugi scenarij.

I ovdje se potvrđuje da se s povećanjem broja grozdova smanjuje vjerojatnost za uspješno odbacivanje grozdova. To se vidi u povećavanju NED-a s povećavanjem broja grozdova, prikazano na slici 5.18. Omjer NED-a između ove dvije metode postupno se smanjuje s povećanjem broja poslužitelja, a u prosjeku je oko 3,5, kako je prikazano slikom 5.19.



Slika 5.18: Grafički prikaz NED-a za MinMax i Kombiniranu metodu pri promjeni poslužitelja za drugi scenarij.



Slika 5.19: Grafički prikaz omjera NED-a MinMax i Kombinirane metode za drugi scenarij.



## 6. PRILAGOĐAVANJE MODELA SKLADIŠTA S CILJEM OPTIMIZACIJE POSLUŽIVANJA

U prethodnom poglavlju predstavljen je model klasičnog skladišta kojeg se može pronaći u raznim poduzećima. Nad tim modelom napravljeni su različiti pokusi u cilju smanjivanja potrebnog vremena za izbor optimalnog poslužitelja za izvršenje komisioniranja. U pokusima su uspoređene MinMax i Kombinirana metoda za razvrstavanje objekata kroz dva scenarija. U oba slučaja Kombinirana metoda davala je bolje rezultate, tj. kraće vrijeme potrebno za postupak razvrstavanja i manji NED.

U ovom poglavlju cilj je prilagoditi sam model skladišta kako bi se ubrzao procesa za izbor optimalnog poslužitelja i skratilo vrijeme za komisioniranje narudžbe. Skraćivanje vremena potrebnog za komisioniranje narudžbe proporcionalno je putu koji poslužitelj mora prijeći pri izuzimanju naručene robe. Ideja je izbjeći potrebu da se poslužitelji gibaju isključivo pravocрно kroz skladište te na taj način skratiti ukupni prijeđeni put i pojednostaviti proces razvrstavanja zamjenjivanjem udaljenosti gradski blokova, tj. Manhattan udaljenosti, kao mjere različitosti, s Euklidskom udaljenosti.

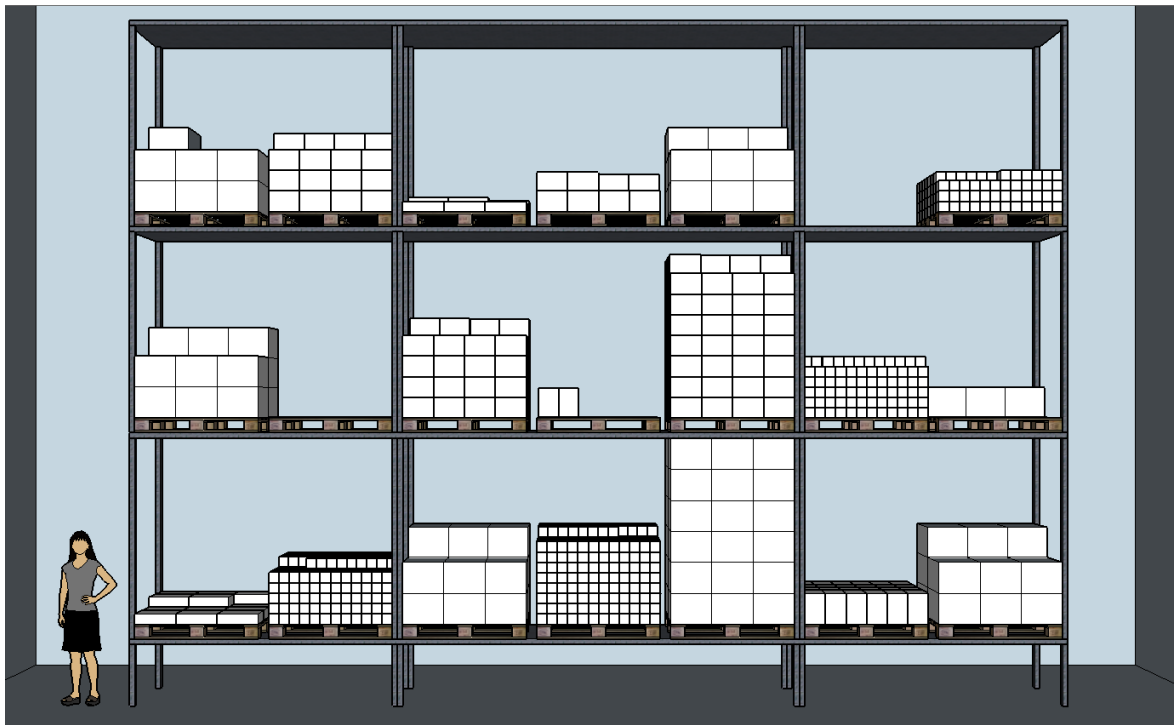
### 6.1. Prilagođeni model skladišta

Model skladišta iz prethodnog poglavlja imao je razmještaj s glavnim prolazom, između dva bloka regala, koji vodi do glavne ulazno/izlazne točke skladišta. Proizvodi su se nalazili na euro paletama koje su smještene na paletne regale. Sustav pozicioniranja robe u skladištu je kaotično skladištenje.

Glavne karakteristike modela skladišta neće se mijenjati kako bi se rezultati pokusa iz ovog poglavlja mogli usporediti s onima iz prethodnog. Glavna izmjena u modelu skladišta je pozicioniranje paletnih regala u odnosu na z-os. Tako je i dalje model skladišta pravokutnog tlocrt, podijeljen u dva paralelna bloka, podijeljena glavnim prolazom sa četiri reda paletnih regala, kao što je prikazano slikom 5.4. Unutarnja površina skladišta je ostala  $21,1 \times 11$  m, ali je visina skladišta povećana za 0,65 m i sada iznosi 7,65 m.

Regali imaju i dalje tri nivoa s duljinom jednakom duljini reda, no prva polica nalazi se na visini od 0,6 m, kao što je prikazano slikom 6.1. Visina regala od poda je 6,65 m, s policom debljine 5 cm na svakih 1,95 m. Regali se nalaze na visini od 0,6 m kako bi omogućili nesmetani prolaz poslužiteljima. Potreba za glavnim prolazom i prolazima između redova i dalje postoji zbog potrebe obnavljanja zaliha u skladištu. Postoji mogućnost

sužavanja prolaza između redova na širinu dovoljnu za prolazak jednog robota koji donosi pune palete s proizvodima i odvozi prazne iz skladišta, sa čime bi se stvorilo dovoljno mjesta za još jedan red s dvostrukim regalom i povećao kapacitet skladišta za 84 palete ili 33,3%, kao što je prikazano slikom 6.2. To nije napravljeno u prilagođenom modelu jer bi povećanje broja objekata utjecalo na trajanje razvrstavanja i ne bi bilo moguće usporediti rezultate s onima iz prošlog poglavlja.



Slika 6.1: Skica jednog paletnog regala s tri nivoa odmaknutog od poda na visinu 0,6 m.



Slika 6.2: Tlocrtni prikaz moguće prilagodbe skladišta smanjenjem širine prolaza između redova.

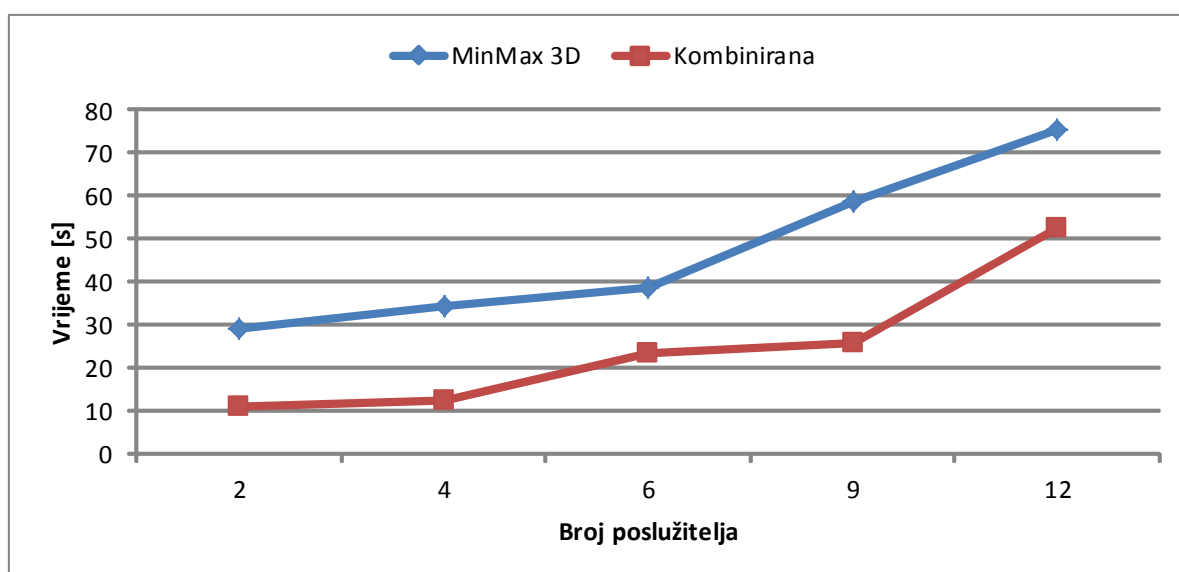
## 6.2. Proces razvrstavanja u prilagođenom modelu skladišta

Kako bi se mogla napraviti usporedba između dva modela skladišta potrebno je napraviti pokuse u kojima se u prilagođenom modelu skladišta obavljaju uobičajeni zadaci koji su podjednako raspoređeni po čitavom skladištu. U svim pokusima uspoređuju se vremena trajanja razvrstavanja između MinMax i Kombinirane metode. Prvi je pokus obrada narudžbe pri prvom scenariju, dok je drugi pokus obrada narudžbe pri drugom scenariju. Narudžbe iz prvog i drugog scenarija jednake su onima iz prethodnog poglavlja

### 6.2.1. Proces razvrstavanja pri normalnom opterećenju

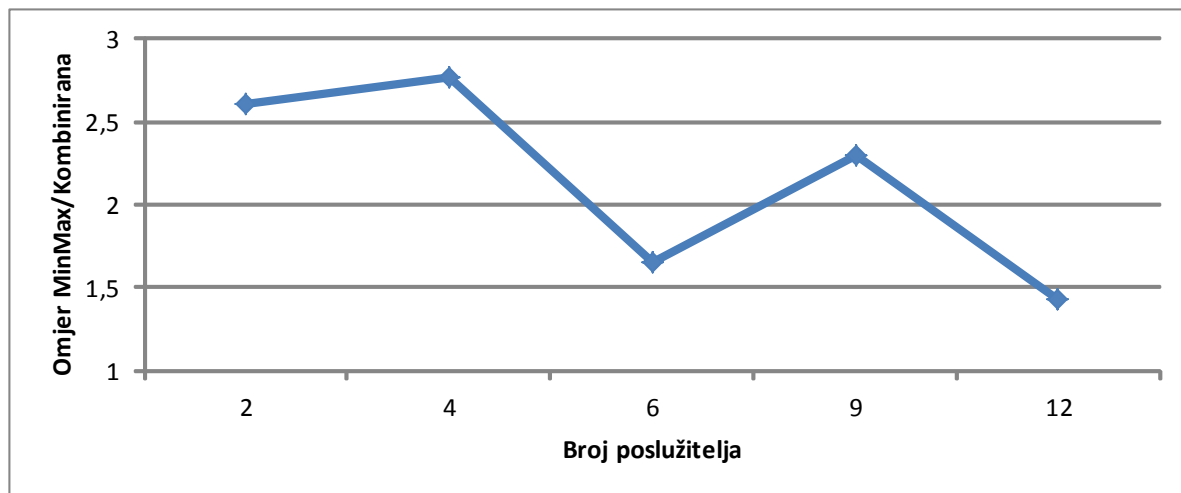
U ovom poglavlju napravljeni su pokusi koji se izvode po prvom scenariju, s ciljem izračunavanja vremena trajanja procesa razvrstavanja pri različitom broju poslužitelja u kojem treba razvrstati objekte. Ovaj scenarij osmišljen je kao normalno opterećenje skladišta gdje poslužitelji nakon svake izvršene narudžbe imaju dovoljno vremena izračunati i otići u „optimalni“ položaj prije nove narudžbe. Pokusi su napravljeni za dva, četiri, šest, devet i dvanaest poslužitelja. Trajanje svakog pokusa je 100 narudžbi. Razvrstavanje je provedeno s MinMax i kombiniranom metodom, a dobiveni rezultati prikazani su na slici 6.3. Sa slike je vidljivo, kao što je i očekivano, da najkraće vrijeme izvođenja razvrstavanja ima Kombinirana metoda.

Na slici 6.4 prikazan je omjer MinMax i Kombinirane metode za prvi scenarij. Sa slike 6.4 vidljivo je prosječno ubrzanje postupka od 2,2 puta.

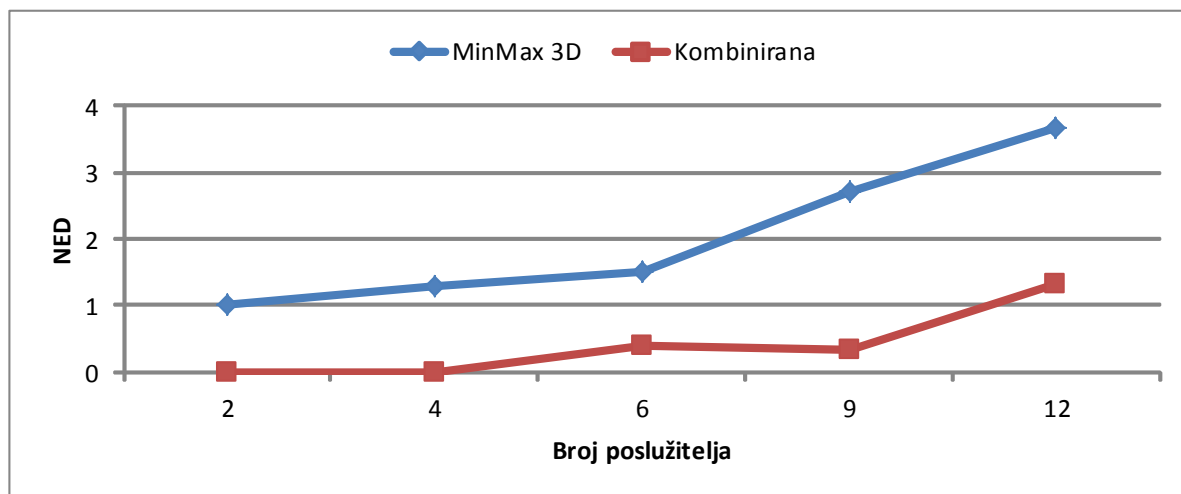


Slika 6.3: Grafički prikaz vremena trajanja razvrstavanja za MinMax i Kombiniranu metodu pri promjeni poslužitelja za prvi scenarij u prilagođenom modelu skladišta.

Kako je već ranije objašnjeno u potpoglavljima 4.3.5 i 5.3.2, povećanjem broja grozdova, grozdovi su međusobno bliži te je manja vjerojatnost za uspješno odbacivanje grozdova. To se vidi i na slici 6.5. NED za obje metode raste s povećavanjem broja grozdova. Omjer NED-a između ove dvije metode različit je u svakoj od mjerenih točaka. Tako je za dva poslužitelja omjer oko 2100, četiri oko 500, šest 4, devet 8 i za dvanaest poslužitelja približno 3. Radi se o činjenici da je NED gotovo jednak nuli pri malom broju poslužitelja u Kombiniranoj metodi, a zatim se počinje približavati konstantnom omjeru.



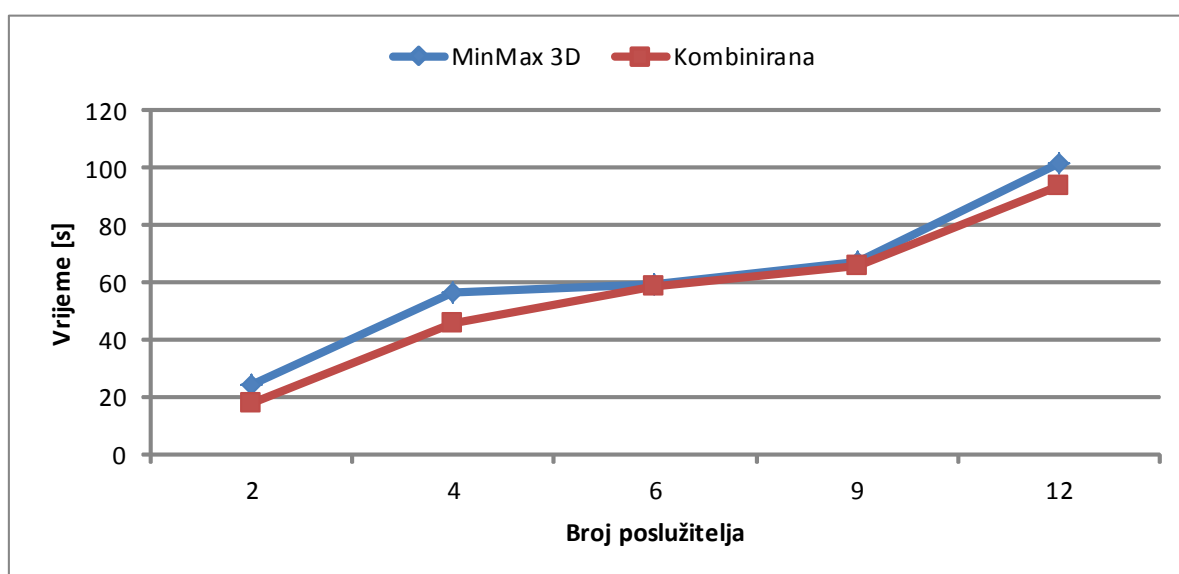
Slika 6.4: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode za prvi scenarij u prilagođenom modelu skladišta.



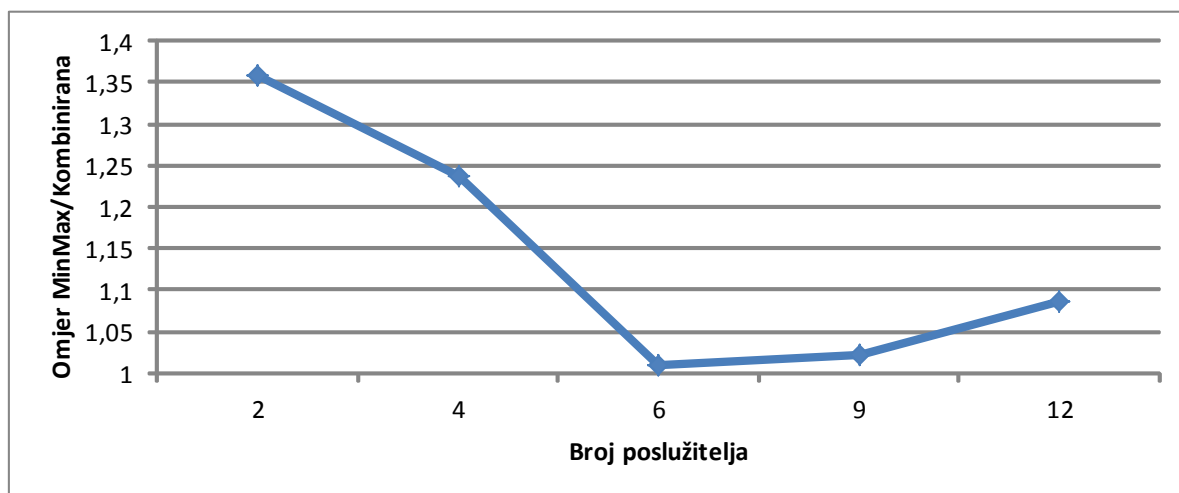
Slika 6.5: Grafički prikaz NED-a za MinMax i kombiniranu metodu pri promjeni poslužitelja za prvi scenarij u prilagođenom modelu skladišta.

### 6.2.2. Proces razvrstavanja pri povećanom opterećenju

U ovom poglavlju napravljeni su pokusi koji se izvode po drugom scenariju, s ciljem izračunavanja vremena trajanja procesa razvrstavanja pri različitom broju poslužitelja u kojem treba razvrstati objekte. Drugi scenarij osmišljen je kao povećano opterećenje skladišta gdje poslužitelji nakon svake izvršene narudžbe nemaju dovoljno vremena izračunati i otići u „optimalni“ položaj prije nove narudžbe, nego se razvrstavanje provodi prema trenutnim pozicijama poslužitelja. Pokusi su napravljeni za dva, četiri, šest, devet i dvanaest poslužitelja. Trajanje svakog pokusa je 100 narudžbi. Razvrstavanje je provedeno MinMax i Kombiniranom metodom, a dobiveni rezultati prikazani su na slici 6.6. Sa slike je vidljivo da



Slika 6.6: Grafički prikaz vremena trajanja razvrstavanja za MinMax i Kombiniranu metodu pri promjeni poslužitelja za drugi scenarij u prilagođenom modelu skladišta.



Slika 6.7: Grafički prikaz omjera vremena izvođenja MinMax i Kombinirane metode za drugi scenarij u prilagođenom modelu skladišta.

Kombinirana metoda ima kraće vrijeme izvođenja od MinMax metode.

Na slici 6.7 prikazan je omjer vremena izvođenja za MinMax i Kombinirane metode za drugi scenarij. Sa slike 6.7 vidljivo je da prosječno ubrzanje postupka nije jako veliko u odnosu na MinMax metodu, no s povećanjem broja objekata, kako je dokazano u potpoglavlju 4.3.2., tj. povećanjem broja paleta i proizvoda u skladištu ovaj bi omjer rastao.

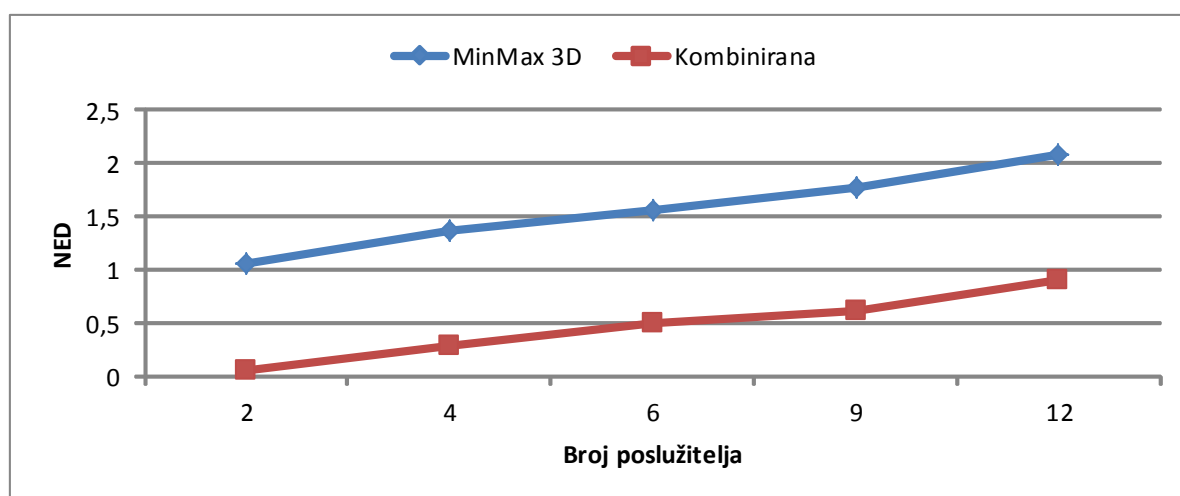
I u ovom se slučaju potvrđuje da se s povećanjem broja grozdova smanjuje vjerojatnost za uspješno odbacivanje grozdova. NED raste s povećanjem broja grozdova, kako je prikazano na slici 6.8. Omjer NED-a između ove dvije metode postupno se smanjuje s povećanjem broja poslužitelja i to od 16,7 do 2,3.

### 6.3. Usporedba rezultata klasičnom i prilagođenog modela skladišta

U ovom poglavlju dana je usporedba rezultata trajanja procesa razvrstavanja Kombinirane metode u klasičnom i prilagođenom modelu skladišta. Svi rezultati prikazani su tablično i prikazan je graf njihovog omjera. U prvom potpoglavlju uspoređeni su rezultati razvrstavanja pri normalnom opterećenju skladišta, dok su u drugom potpoglavlju uspoređeni rezultati pri povećanom opterećenju.

#### 6.3.1. Usporedba rezultata pri normalnom opterećenju

U ovom potpoglavlju provedena je usporedba vremena trajanja procesa razvrstavanja za 100 narudžbi, po prvom scenariju, u klasičnom i prilagođenom modelu skladišta. Vremena



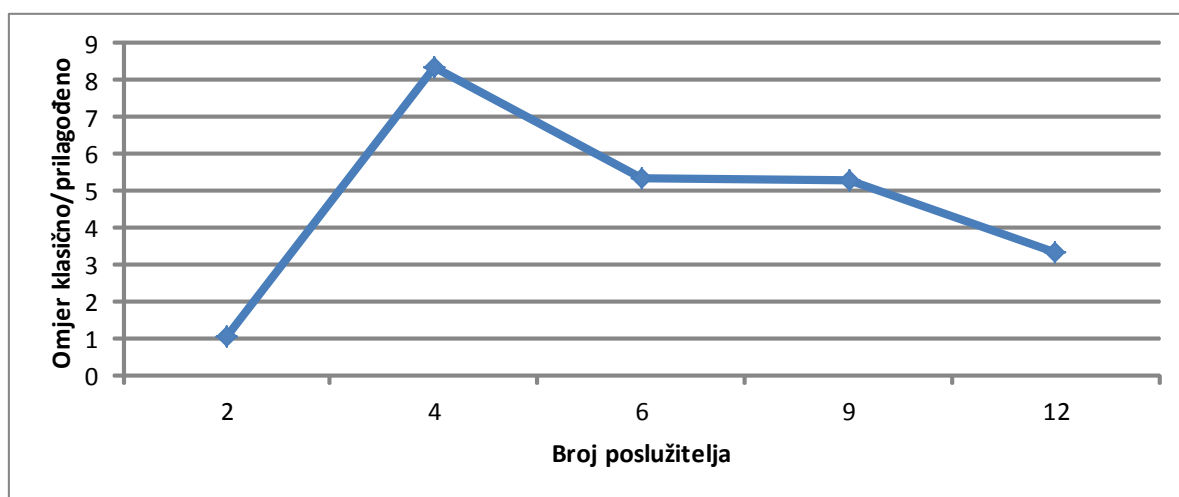
Slika 6.8: Grafički prikaz NED-a za MinMax i Kombiniranu metodu pri promjeni poslužitelja za drugi scenarij u prilagođenom modelu skladišta.

Tablica 6.1: Vremena trajanja procesa razvrstavanja Kombiniranom metodom za prvi scenarij u klasičnom i prilagođenom modelu skladišta

Model skladišta	Vrijeme izvođenja [s] za različiti broj poslužitelja				
	2	4	6	9	12
Klasični model (poglavlje 5)	11,07	103,01	125,12	135,54	176,76
Prilagođeni model (poglavlje 6)	10,80	12,36	23,33	25,65	52,67

trajanja procesa razvrstavanja prikazana su u tablici 6.1. Iz tablice je vidljivo da se prilagodbom modela skladišta znatno smanjilo vrijeme potrebno za razvrstavanje objekata u skladištu. Omjer vremena izvođenja prikazan je slikom 6.9. Sa slike je vidljivo da je za dva poslužitelja omjer 1, no za ostale poslužitelje omjer je između 8,5 i 3 puta. Iz navedenog je moguće zaključiti da je prosječno ubrzanje postupka od 4,6 puta.

Ako se usporede vrijednosti NED-a za ova dva modela skladišta može se zaključiti da je NED manji u prilagođenom modelu skladištu, pogotovo u pokusima sa četiri poslužitelja. U tom slučaju NED prilagođenog modela, nakon prvog razvrstavanja s PS podacima, jednak je nula, za razliku od NED-a klasičnog modela skladišta koji je jednak 0,49. U ostalim slučajevima omjer NED-a je u prosjeku 1,25.



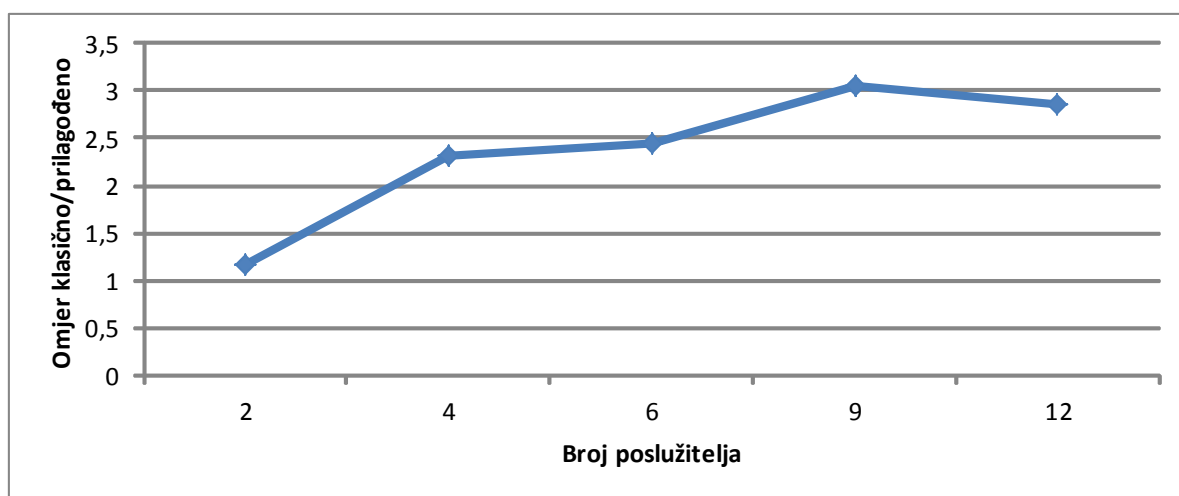
Slika 6.9: Grafički prikaz omjera vremena, izvođenja Kombinirane metode za prvi scenarij, klasičnog i prilagođenog modela skladišta.

### 6.3.2. Usporedba rezultata pri povećanom opterećenju

Nakon analize usporedbe vremena za prvi scenarij, u ovom potpoglavlju provedena je usporedba trajanja procesa razvrstavanja po drugom scenariju, u klasičnom i prilagođenom modelu skladišta za 100 narudžbi. Vremena trajanja procesa razvrstavanja prikazana su u tablici 6.2. Iz tablice je vidljivo da prilagođeni modela skladišta ima znatno kraće vrijeme potrebno za razvrstavanje objekata od klasičnog modela skladišta. Na slici 6.10 prikazan je omjer vremena izvođenja. Sa slike je vidljivo da je omjer najmanji za dva poslužitelja i približno je jednak 1. Za četiri poslužitelja omjer raste na 2,3 i zadržava tendenciju rasta za ostale brojeve poslužitelja. Može se reći da je prosječno ubrzanje postupka za drugi scenarij u prilagođenom modelu skladišta oko 2,4 puta. Vrijednosti NED-a za ova dva modela skladišta gotovo su jednaka i njihov prosječan omjer je oko 1,1.

Tablica 6.2: Vremena trajanja procesa razvrstavanja Kombiniranom metodom za drugi scenarij u klasičnom i prilagođenom modelu skladišta

Model skladišta	Vrijeme izvođenja [s] za različiti broj poslužitelja				
	2	4	6	9	12
Klasični model (poglavlje 5)	20,66	105,86	143,61	200,26	268,12
Prilagođeni model (poglavlje 6)	17,87	45,68	58,90	65,57	93,78



Slika 6.10: Grafički prikaz omjera vremena, izvođenja Kombinirane metode za drugi scenarij, klasičnog i prilagođenog modela skladišta.



## 7. ZAKLJUČAK

U disertaciji je predložen novi postupak razvrstavanja objekata, koji sadrže neodređenost u svojim prostornim podacima, kao znanstveni doprinos. Razvrstavanje objekata s neodređenosti vremenski je zahtjevan proces u kojemu je vremenski najzahtjevnija računalna operacija izračun očekivane udaljenosti (ED). Prostorna neodređenost objekta opisuje se funkcijom gustoće vjerojatnosti (PDF) koja se predstavlja kao minimalno područje neodređenosti, koje je podijeljeno na uzorke. Svakom uzorku dodijeljuje se vjerojatnost da se objekt nalazi u tome uzorku. Složenost procesa proizlazi iz računanja očekivane udaljenosti, koja se računa od svakog uzorka područja neodređenosti objekta do središta svakog grozda, što znatno usporava proces razvrstavanja. Kako bi se ubrzao (pojednostavio) proces razvrstavanja objekata s neodređenosti, razvijene su metode za odbacivanje grozdova kao kandidata za pojedini objekt bez računanja očekivane udaljenosti.

Postojeće metode razvrstavanja objekata s neodređenosti, navedene u literaturi, imaju određene nedostatke u pojedinim slučajevima. Tako MinMax metoda odbacuje mali broj grozdova u slučajevima malog broja grozdova i velikog broja uzoraka, metoda Voronojevih dijagrama i metoda simetralne podjele prostora (SPP) imaju veći broj odbačenih grozdova, ali je sam proces odbacivanja neučinkovit i ukupno vrijeme razvrstavanja može trajati duže od onog dobivenog MinMax metodom. Iz tog razloga, predstavljena je nova Kombinirana metoda koja iskorištava prednosti postojećih metoda i uklanja njihove nedostatke. Kombinirana metoda smanjuje broj računanja očekivane udaljenosti u odnosu na MinMax metodu te ubrzava proces odbacivanja grozdova u odnosu na metodu simetralne podjele prostora i Voronojevih dijagramima. Navedeno rezultira vremenski kraćim procesom razvrstavanja u čitavom rasponu parametara, kao što su broj objekata, maksimalna duljina stranice i broj uzoraka područja neodređenosti te broj grozdova. Kombinirana metoda u prvom dijelu algoritma odbacuje grozdove koristeći minimalne i maksimalne udaljenosti od grozda do objekta, a u drugom dijelu algoritma odbacuje dodatne grozdove koristeći podjelu prostora simetralnim ravninama. Pokusima je u četvrtom poglavlju dokazano da je predloženoj metodi potrebno manje vremena za proces razvrstavanja, nego postojećim metodama u čitavom opsegu svih parametara.

U ovoj disertaciji također je predstavljen model skladišta za verifikaciju rezultata predložene metode, kao drugi znanstveni doprinos. Predstavljeni model odgovara robotiziranom skladištu s razmještajem u dva bloka regala s glavnim prolazom. Napravljena je baza podataka s lokacijom i brojem paleta u skladištu. Paleta su objekti s prostornom

neodređenosti koja je predstavljena brojem paketa (proizvoda), koji se na njima nalaze. Zadatak ovog modela bio je opisivanje trenutnog stanja skladišta i predviđanje budućeg stanja s ciljem skraćivanja vremena razvrstavanja proizvoda i pronalaženja optimalnih pozicija za poslužitelje. Osmišljena su dva scenarija pristizanja narudžbi. Rezultati pokusa, u petom poglavlju, pokazuju kako Kombinirana metoda znatno skraćuje vrijeme potrebno za izbor optimalnog poslužitelja, za izvršenje komisioniranja.

Kao treći znanstveni doprinos u ovoj disertaciji, predstavljena je prilagodba modela skladišta novo razvijenoj metodi u svrhu optimiranja (optimizacije) posluživanja. Napravljene su izmjene u razmještaju skladišta kojima se omogućuje upotreba brzih metoda za razvrstavanje podataka o proizvodima i narudžbi te se u može povećati kapacitet skladišta, za jednu trećinu, bez promjene prostorne kvadrature samog objekta. Rezultati pokusa, u šestom poglavlju, pokazuju kako Kombinirana metoda u prilagođenom skladištu skraćuje vrijeme potrebno za izbor optimalnog poslužitelja čak do 8 puta u omjeru s klasičnim modelom skladišta. Prosječni omjer vremena, u čitavom opsegu svih parametara u oba scenarija, iznosi 3,5. Predloženim prilagodbama modela skladišta i korištenjem Kombinirane metode za razvrstavanje ubrzava se proces izbor optimalnog poslužitelja te smanjuju troškovi sustava.

## 8. LITERATURA

- [1] M. J. A. Berry i G. S. Linoff, *Data Mining Techniques*, Wiley, 1997.
- [2] D. Hand, H. Mannila i P. Smyth, *Principles of Data Mining*, Cambridge: The MIT Press, 2001.
- [3] M. Vranić, »Dubinska analiza podataka – određeni postupci u području deskriptivne analize,« Fakultet elektrotehnike i računarstva, Zagreb, 2009.
- [4] Grupa autora, *Rudarenje podataka: različiti aspekti informacijskog društva*, Split: Ekonomski fakultet Split, 2011.
- [5] Grupa autora, »CRISP-DM 1.0 Step-by-step data mining guide,« Tehnički izvještaj, The CRISP-DM consortium, 2000.
- [6] A. Azevedo i M. F. Santos, »KDD, SEMMA and CRISP-DM: a parallel overview,« u *IADIS European Conf. Data Mining*, Amsterdam, 2008.
- [7] E. Cox, *Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration*, Elsevier Inc., 2005.
- [8] A. K. Jain i R. C. Dubes, *Algorithms for clustering data*, Englewood Cliffs, New Jersey: Prentice Hall, 1988.
- [9] B. Everitt, *Cluster Analysis*, New York: John Wiley & Sons, Inc., 1974.
- [10] G. Gan, C. Ma i J. Wu, *Data Clustering: Theory, Algorithms, and Applications*, Philadelphia: Society for Industrial and Applied Mathematics, 2007.
- [11] Y. Fujikawa i T. B. Ho, »Cluster-Based Algorithms for Dealing with Missing Values,« u *Paper presented at the meeting of the PAKDD*, 2002.
- [12] J. A. Hartigan, »Representation of similarity matrices by trees,« *Journal of the American Statistical Association*, svez. 62, br. 320, p. 1140–1158, 1967.
- [13] C. Aggarwal, J. Wolf, P. Yu, C. Procopiuc i J. Park, »Fast algorithms for projected clustering,« u *In Proceedings of the 1999 ACM SIGMOD international conference on management of data*, Philadelphia, 1999.
- [14] L. Orloci, »An agglomerative method for classification of plant communities,« *Journal of Ecology*, svez. 55, pp. 193-206, 1967.
- [15] P.-N. Tan, M. Steinbach i V. Kumar, *Introduction to Data Mining*, Addison Wesley, 2005.

- [16] B. Duran i P. Odell, *Cluster Analysis: A Survey*, Berlin: Springer Berlin Heidelberg, 1974.
- [17] J. MacQueen, »Some methods for classification and analysis of multivariate observations,« u *In Proceedings of the 5th Berkeley symposium on mathematical statistics and probability*, Berkeley, 1967.
- [18] Z. Huang, »Extensions to the k-means algorithm for clustering large data sets with categorical values,« *Data Mining and Knowledge Discovery*, svez. 2, br. 3, p. 283–304, 1998.
- [19] S. Selim i M. Ismail, »k-means-type algorithms: A generalized convergence,« *IEEE Transactions on Pattern Analysis*, svez. 6, br. 1, pp. 81-86, 1984.
- [20] J. M. Peña, J. A. Lozano i P. Larrañaga, »An empirical comparison of four initialization methods for the K-Means algorithm,« *Pattern Recognition Letters*, svez. 20, br. 10, pp. 1027-1040, 1999.
- [21] S. S. Khan i A. Ahmad, »Cluster center initialization algorithm for K-means clustering,« *Pattern Recognition Letters*, svez. 25, br. 11, pp. 1293-1302, 2004.
- [22] V. Faber, »Clustering and the continuous k-means algorithm,« *Los Alamos Science*, svez. 22, p. 138–144, 1994.
- [23] S. J. Phillips, »Acceleration of K-Means and Related Clustering Algorithms,« u *ALENEX - Lecture Notes in Computer Science*, 2002.
- [24] D. Pelleg i A. W. Moore, »Extending k-means with efficient estimation of the number of clusters,« u *International conference on machine learning*, San Francisco, 2000.
- [25] C. H. Cheng, A. W.-C. Fu i Y. Zhang, »Entropy-based Subspace Clustering for Mining Numerical Data,« u *ACM SIGKDD international conference on knowledge discovery and data mining*, New York, 1999.
- [26] D. Comaniciu i P. Meer, »Mean shift analysis and applications,« u *ICCV - IEEE International Conference on Computer Vision*, Kerkyra, Greece, 1999.
- [27] Z. Huang, »A fast clustering algorithm to cluster very large categorical data sets in data mining,« u *Research Issues on Data Mining and Knowledge Discovery*, Tucson, Arizona, 1997.
- [28] Z. Huang, »Extensions to the k-means algorithm for clustering large data sets with categorical values,« *Data Mining and Knowledge Discovery*, svez. 2, pp. 283-304,

- 1998.
- [29] D. Wishart, »k-means clustering with outlier detection, mixed variables and missing values,« u *Exploratory Data Analysis in Empirical Research*, New York, 2002.
- [30] M. Ester, H.-P. Kriegel, J. Sander i X. Xu, »A density-based algorithm for discovering clusters in large spatial databases with noise,« u *Proc. of 2nd International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, 1996.
- [31] M. Dash, H. Liu i X. Xu, »'1 + 1 > 2': Merging Distance and Density Based Clustering,« u *DASFAA - International Conference on Database Systems for Advanced Applications*, Hong Kong, China, 2001.
- [32] L. J. Hubert, »Some applications of graph theory to clustering,« *Psychometrika*, svez. 39, br. 3, p. 283–309, 1974.
- [33] P. Grabusts i A. Borisov, »Using grid-clustering methods in data classification,« u *Parallel Computing in Electrical Engineering*, Warsaw, Poland, 2002.
- [34] L. Xiao i E. Hung, »An Efficient Distance Calculation Method for Uncertain Objects,« u *IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2007*, Honolulu, Hawaii, 2007.
- [35] M. Chau, R. Cheng i B. Kao, »Uncertain Data Mining: A New Research Direction,« u *Workshop on the Sciences of the Artificial*, Hualien, Taiwan, 2005.
- [36] M. Chau, R. Cheng, B. Kao i J. Ng, »Uncertain data mining: An example in clustering location data,« *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, p. 199–204, 2006.
- [37] N. N. Dalvi i D. Suciu, »Efficient query evaluation on probabilistic databases,« *The International Journal on Very Large Data Bases*, svez. 16, br. 4, pp. 523-544, 2007.
- [38] R. Cheng, D. V. Kalashnikov i S. Prabhakar, »Querying Imprecise Data in Moving Object Environments,« *IEEE Trans. Knowl. Data Eng.*, svez. 16, pp. 1112-1127, 2004.
- [39] O. Wolfson, P. A. Sistla, S. Chamberlain i Y. Yesha, »Updating and querying databases that track mobile units,« *Distributed and Parallel Databases*, svez. 7, pp. 257-387, 1999.
- [40] H.-P. Kriegel i M. Pfeifle, »Hierarchical density-based clustering of uncertain data,« u *Proc. of International Conference on Data Mining (ICDM 2005)*, Houston, Texas, USA, 2005.

- [41] H.-P. Kriegel i M. Pfeifle, »Density-based clustering of uncertain data,« u *Proc. of International Conference on Knowledge Discovery and Data Mining*, Chicago, 2005.
- [42] W. K. Ngai, B. Kao, C. K. Chui, R. Cheng, M. Chau i K. Y. Yip, »Efficient clustering of uncertain data,« u *International Conference on Data Mining, ICDM 2006*, Hong Kong, China, 2006.
- [43] B. Kao, S. D. Lee, D. W. Cheung, W.-S. Ho i K. F. Chan, »Clustering Uncertain Data using Voronoi Diagrams,« u *International Conference on Data Mining*, Pisa, Italy, 2008.
- [44] B. Kao, S. D. Lee, D. W. Cheung, F. K. F. Lee, D. W.-l. Cheung i W.-S. Ho, »Clustering Uncertain Data Using Voronoi Diagrams and R-Tree Index,« *IEEE Transactions on Knowledge and Data Engineering*, svez. 22, br. 9, pp. 1219-1233, 2010.
- [45] I. Lukić, N. Slavek i M. Köhler, »Improved Bisector Clustering Of Uncertain Data Using Sdsa Method On Parallel Processors,« *Tehnički Vjesnik - Technical Gazette*, svez. 20, pp. 255-261, 2013.
- [46] I. Lukić, M. Köhler i N. Slavek, »Improved Bisector Pruning for Uncertain Data Mining,« u *Proc. of the ITI 2012*, Caftat, 2012.
- [47] C. Elkan, »Using the triangle inequality to accelerate k-means,« u *International Conference on Machine Learning, ICML*, Washington DC, USA, 2003.
- [48] I. Lukić, »Simetralno razvrstavanje objekata uslužnih sustava,« Sveučilište J. J. Strossmayera - Elektrotehnički fakultet Osijek, Osijek, 2012.
- [49] F. Vanhoutte, »W:BLUT – CREATIVE CODING,« 15 7 2014. [Mrežno]. Dostupno: <http://www.wblut.com/2009/04/28/ooh-ooh-ooh-3d-voronoi/>.
- [50] B. Pavković i D. Veljan, *Elementarna matematika II*, Zagreb: Školska knjiga, 1995.
- [51] M. Habek, *Upravljanje zalihama i skladišno poslovanje*, Zagreb, 2002.
- [52] Z. Segetlija, *Distribucija*, Osijek: Ekonomski fakultet u Osijeku, 2006.
- [53] F. Caron, G. Marchet i A. Perego, »Layout design in manual picking systems: a simulation approach,« *Integrated Manufacturing Systems*, svez. 11, br. 2, pp. 94-104, 2000.
- [54] V. Ferišak, *Nabava, politika – strategija – organizacija – management*, Zagreb, 2002.

## Sažetak

U ovoj disertaciji predstavljene su prilagođene metode i predložena je nova Kombinirana metoda za razvrstavanje objekata koji sadrže neodređenost u svojim prostornim podacima u  $\mathbb{R}^3$  geometrijskom prostoru. Kombinirana metoda koristi prednosti i uklanja nedostatke postojećih metoda. Ona omogućuje brže razvrstavanje objekata neodređenosti koristeći jednostavne matematičke funkcije usporedbe za odbacivanje grozdova, a koristeći simetralne ravnine za usporedbu i odbacivanju grozdova kao potencijalnih kandidata odbacuje dodatni broj grozdova. Metoda znatno ubrzava proces odbacivanja grozdova, u odnosu na postojeće metode, tako što minimalizira ukupan broj objekata za koji je potrebno računati očekivanu udaljenost. Stvoren je model skladišta za verifikaciju nove metode razvrstavanja i u njemu provedeni pokusi s ciljem skraćivanja vremena razvrstavanja proizvoda i pronalaženja optimalnih pozicija za poslužitelje. Predstavljene su prijedlozi za optimiranje posluživanja prilagođavanjem modela skladišta novoj metodi. Na temelju prijedloga kreiran je novi model skladišta nad kojim su provedeni isti pokusi s ciljem dokazivanja ubrzanja procesa razvrstavanja.

**Ključne riječi:** objekti s neodređenosti, optimizacija, odbacivanje grozdova, razvrstavanje, skladište

## Abstract

In this dissertation adapted methods and new combined method for clustering spatially uncertain data in  $\mathbb{R}^3$  geometrical space are presented. The proposed combined method uses the advantages and eliminates the disadvantages of the existing methods. Combined method enables faster clustering of objects with spatial uncertainty using simple comparisons mathematical functions for cluster pruning. Pruning additional number of clusters is achieved by using bisector planes for comparison and rejection of clusters as potential candidates. The new method significantly speeds up the process of pruning clusters, compared to existing methods. This is achieved by minimizing the total number of objects for which it is necessary to calculate the expected distance. A warehouse model for verification of new clustering method is created. Experiments are carried out with the aim of shortening the time for clustering of products and finding the optimal position for automated servers. Proposals for service optimizing by customizing warehouse model, for new method, are presented. On the basis of proposals new warehouse model is created. On the new developed model same experiments are performed with the aim of proving acceleration of the clustering process.

**Keywords:** cluster pruning, clustering, uncertain objects, optimization, warehouse



## **Životopis**

Mirko Köhler rođen je u Osijeku 4. listopada 1983. godine. Tijekom osnovne škole redovito sudjeluje na državnim natjecanjima tehničke kulture i odlučuje se upisati srednju Elektrotehničku i prometnu školu Osijek gdje maturira s odličnim uspjehom. Nakon završetka srednje škole, 2002. godine, upisuje se na Elektrotehnički fakultet u Osijeku gdje i diplomira 2007. godine. Tijekom studija radi u Hrvatskom Telekomu, a po završetku studija zapošljava se u tvrtci Zaštita-Inspekt d.o.o. 15. prosinca 2008. godine zapošljava se kao asistent na Elektrotehničkom fakultetu u Osijeku, te upisuje poslijediplomski doktorski studij, na kojem mu je mentor prof. dr. sc. Franjo Jović. Zbog mentorovog odlaska u mirovinu za novog mentora izabran je doc. dr. sc. Krešimir Nenadić, koji ga je vodio tijekom pisanja ove disertacije. Uže područje interesa i istraživanja mu je razvrstavanje objekata koji sadrže neodređenost u prostornim podacima. Ostala područja interesa su sigurnost informacije te modeliranje. Autor je više znanstvenih radova i koautor sveučilišnog udžbenika.