

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Urša Krevs

**Računalniška analiza in primerjava
stilov igre šahistov**

MAGISTRSKO DELO
ŠTUDIJSKI PROGRAM DRUGE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: doc. dr. Matej Guid

Ljubljana, 2015

Rezultati magistrskega dela so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavljanje ali izkoriščanje rezultatov magistrskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorja.

IZJAVA O AVTORSTVU MAGISTRSKEGA DELA

Spodaj podpisana Urša Krevs sem avtorica magistrskega dela z naslovom:

Računalniška analiza in primerjava stilov igre šahistov

S svojim podpisom zagotavljam, da:

- sem magistrsko delo izdelala samostojno pod mentorstvom doc. dr. Mateja Guida,
- so elektronska oblika magistrskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko magistrskega dela,
- soglašam z javno objavo elektronske oblike magistrskega dela v zbirki "Dela FRI".

V Ljubljani, 26. marca 2015

Podpis avtorja:

*Zahvaljujem se moji družini in Igorju, ki so mi ves ta čas stali ob strani.
Zahvaljujem se tudi mentorju in vsem v laboratoriju za umetno inteligenco,
ki so mi pomagali pri tem delu.*

Kazalo

1	Uvod	1
1.1	Pregled sorodnega dela	3
1.2	Struktura magistrskega dela	4
2	Podatki	7
2.1	Splošno o podatkih	7
2.2	Atributi podatkov	9
2.3	Profil igralca	19
3	Metode	23
3.1	Računanje podobnosti	23
3.2	Hierarhično razvrščanje	25
3.3	Metoda voditeljev	26
3.4	Povprečna medsebojna informacija	27
3.5	Entropija	27
3.6	Standardni odklon	28
4	Rezultati	29
4.1	Iskanje podobnih šahistov	29
4.2	Povprečna medsebojna informacija atributov	39
4.3	Razvrščanje svetovnih šahovskih prvakov v skupine	41
5	Zaključek	45
A	Primerjava igralcev s svetovnimi šahovskimi prvaki	53

Seznam uporabljenih kratic

Kratica	V izvirniku	Slovensko
FEN	Forsyth-Edwards Notation	Forsyth-Edwardsov zapis
ECO	Encyclopaedia of Chess Openings	Enciklopedija šahovskih otvoritev
FIDE	Fédération Internationale des Échecs	Svetovna šahovska federacija

Povzetek

Računalniški šahovski programi dandanes že izjemno dobro ocenjujejo šahovske pozicije. Raziskave so pokazale, da je šahiste s pomočjo računalniškega šahovskega programa možno razvrščati glede na kvaliteto igre.

V magistrskem delu *Računalniška analiza in primerjava stilov igre šahistov* se posvetimo vsebinski analizi partij, pri kateri si pomagamo z ocenami računalniškega šahovskega programa in atributi, ki jih določimo za vsako potezo. Atributi, ki smo jih uporabili, omogočajo računalniško primerjavo stilov igre in so hkrati razumljivi šahistom.

S pomočjo izbranih atributov smo gradili profil, s katerim smo definirali stil igranja šahistov. Kvaliteto pridobljenih profilov smo evalvirali z avtomatskim prepoznavanjem šahista v množici šahistov. Pri tem smo profil šahista, v množici in izven nje, gradili z analizo različnih partij. Glede na rezultate testiranja smo spreminjali in dopolnjevali strukturo profila. Pri tem smo si pomagali tudi z računanjem povprečne medsebojne informacije atributov.

Z izbranim, najbolj primernim profilom smo šahistom iskali podobne svetovne šahovske prvake.

Svetovne šahovske prvake smo tudi razvrščali v skupine, glede na naš opis stila s profilom. Na podlagi znanih stilov teh šahistov smo ugotavljali uspešnost razvrščanja.

S pomočjo razvitih profilov lahko šahistu delno avtomatiziramo in olajšamo analizo njegovih partij. Menimo tudi, da bi bilo postopek analize možno uporabiti v nekaterih drugih domenah.

POVZETEK

Ključne besede: šah, računalniška analiza šahovskih partij, stil igranja, odkrivanje skupin.

Abstract

Today's computer chess programs are very good at evaluating chess positions. Research has shown that we can rank chess players by the quality of their game play, using a computer chess program.

In the master's thesis *Computer analysis and comparison of chess players' game-playing styles*, we focus on the content analysis of chess games using a computer chess program's evaluation and attributes we determined for each individual position. We defined meaningful attributes that can be used for computer analysis and are also comprehensible to a chess player.

Using the attributes, we built profiles with which we defined chess players' styles. We evaluated the quality of the profiles by automatically identifying a chess player in a set of chess players. The profile of the chess player in the set and the one outside of it was built using different chess games. Using the result of the analysis we refined the profiles structure. In doing so we were aided by the information gain of each attribute.

The most suitable profile was used for searching for world chess champions with the most similar style to a chosen chess player.

We also sorted the world chess champions into groups according to their style of play. Because these players are well known, we compared our groups with the chess players' actual styles and determined how successful we were.

By using the developed profiles, we can help partly automate and ease a chess player's analysis of chess games. We believe that the methods used in building the profiles and for the subsequent analysis could be applied to other domains.

ABSTRACT

Keywords: chess, computer analysis of chess games, playing style, group discovery.

Poglavje 1

Uvod

Šahisti neprenehno stremijo k izboljšanju svoje igre. Način, kako to dosegaajo, se v vsem tem času ni kaj dosti spremenil. Analizirajo poteze, ki so jih odigrali sami, ter preigravajo in analizirajo poteze drugih šahistov. Danes si pri analizah pomagajo s sodobnimi šahovskimi programi, kot sta npr. *Rybka* ali *Houdini*, slednjega smo v tem magistrskem delu uporabili tudi sami. Takšni programi lahko šahistu podajo številsko oceno odigrane poteze in izbor najboljših možnih potez za dano pozicijo. Še vedno pa mora večino analize opraviti sam. Pri tem mu je edina pomoč lastno znanje in pomoč drugih šahistov, mentorjev.

Z uspešno računalniško analizo bi lahko velik del tega postopka avtomatizirali. Če bi znali šahiste avtomatično ločiti po stilu, bi lahko nekemu izbranemu šahistu povedali, kateri šahist igra podobno kot on, čigave partije naj si ogleda, da mu bodo v pomoč, kdo v tistih pozicijah, kjer on igra slabo, igra dobro itd.

Prvi problem, s katerim smo se soočili, je bil prav ugotavljanje stila igre šahistov. Želeli smo tudi, da bi že sama definicija stila lahko šahistu nekaj povedala o načinu igre, ne pa da bi bil to le podatek, uporaben za primerjavo z drugimi šahisti. Zato smo se odločili za profil stila, ki je sestavljen iz več lahko razumljivih atributov. Takšen atribut je npr. prisotnost kraljice. Šahist bi lahko iz samega profile, ki definira stil, brez dodatnih analiz, ugotovil ali dela

več ali manj napak, če ima na šahovnici kraljico.

Sledila je analiza profila stila in uporabljenih atributov, ki ga definirajo. S pomočjo te analize smo ugotavljali iz katerih atributov mora biti sestavljen profil stila, da je čim bolj informativen. Šahiste smo glede na njihov stil tudi razvrščali z metodo hierarhičnega razvrščanja (vir [2]) in z metodo voditeljev (vira [9, 10]). Preizkusili smo dve meri podobnosti, kosinusno razdaljo in Pearsonov korelacijski koeficient (vir [1]). Attribute smo analizirali z računanjem povprečne medsebojne informacije (vir [3]) vsakega posameznega atributa.

Pri analizah podatkov o odigranih partijah smo uporabili dve bazi podatkov. Prva baza je vsebovala podatke o partijah 15 svetovnih šahovskih prvakov. To so bili podatki o partijah za naslov svetovnega šahovskega prvaka. Druga baza je vsebovala podatke o partijah več različno rangiranih šahistov. Ta baza je bila večja in smo jo uporabili za testiranje uspešnosti profilov, uporabljenih pri analizi stila igre šahistov. Bazo s svetovnimi prvaki smo uporabili za razvrščanje šahistov glede na njihov stil igre, saj so ti šahisti dobro poznani, zato ekspert lažje oceni ali je bilo razvrščanje uspešno. Pri iskanju šahistov s podobnim stilom igre smo uporabili obe bazi, saj smo šahistom iz druge baze iskali podobne svetovne prvake.

Pri analizi podatkov o odigranih partijah smo sestavili in preizkusili 5 različnih profilov. Profile smo testirali in ugotovili, da najboljši dosega 29% točnost določanja šahistov, vendar ta profil ni uporabljal atributov potrebnih za nadaljno analizo. Najboljši profil, ki je uporabljal attribute, je dosegel 27% točnost. Iskali smo svetovne šahovske prvake, ki so podobni igralcem v bazi ostalih igralcev in razvrščali svetovne šahovske prvake v skupine glede na profile. Pri razvrščanju svetovnih prvakov v skupine smo opazili, da smo kot podobne našli tiste svetovne šahovske prvake, ki so igrali partije drug proti drugemu, čeprav ti šahisti glede na stil drug drugemu niso podobni. Domnevamo, da so šahisti pri igranju partij drug drugemu vsilili svoj stil igre (vir [15]).

Programsko kodo za izvajanje analiz smo napisali v programskem jeziku

Python. Pri tem smo si za izvajanje nekaterih analiz (hierarhično razvrščanje in metoda voditeljev) pomagali s paketom *Orange* (vir [16]). Kot šahovski ekspert je nastopal mentor tega dela, doc. dr. Matej Guid, FIDE mojster.

1.1 Pregled sorodnega dela

Pri pripravi tega dela smo pregledali nekaj sorodnega dela, ki se nanaša na računalniško analizo šahovske igre in šahistov.

V članku *Using Heuristic-Search Based Engines for Estimating Human Skill at Chess* (vir [14]) ugotavljajo, da je možno ocene računalniških šahovskih programov, kot je npr. *Houdini*, uporabiti za rangiranje šahistov glede na kvaliteto njihove igre. Ugotavljajo tudi, da rangiranje šahistov ostane enako ne glede na moč uporabljenega šahovskega programa. Podobni tematiki se posveča tudi članek *Computer Analysis of World Chess Champions* (vir [12]), kjer glede na kvaliteto njihove igre poskušajo rangirati šahovske svetovne pravke. Mi se za razliko od njih posvečamo bolj stilu igralca kot kvaliteti njegove igre, vendar se tudi mi pri tem zanašamo na ocene računalniških šahovskih programov.

Članek *Learning Positional Features for Annotating Chess Games: A Case Study* (vir [13]) se ukvarja z avtomatičnim komentiranjem šahovskih partij. Pri tem avtorji uporabljajo preproste attribute, kot je npr. prisotnost lovskega para. Tudi mi uporabimo takšne preproste attribute, poleg tega pa dodamo še nekaj bolj kompleksnih a vseeno razumljivih atributov. Tak je npr. atribut dvojni kmetje, ki analizira kmečke strukture prisotne na šahovnici.

V članku *A Methodology for Learning Players' Styles from Game Records* (vir [11]) poskušajo s pomočjo atributov določiti stil šahista ter ta stil podati šahovskemu programu tako, da bi program igral z enakim stilom, kot izbrani šahist. Tudi oni za določanje stilov šahistov uporabljajo attribute. Atributov, ki jih uporabijo oni je več in so za šahista pri analizi igre manj pomembni, kot atributi, ki jih uporabljamo mi. Tak atribut je npr. ali je neka figura

na šahovnici na poljih $D4$, $D5$, $E4$ ali $E5$, ki je lahko za računalniško analizo sicer pomemben, vendar šahistu ne pove prav veliko. Mi smo pri naši analizi uporabljali predvsem attribute, ki so preprosto razumljivi za šahista.

Računalniška analiza človekovih napak pri igranju šaha (vir [5]) je moje diplomsko delo, ki je predhodnik tega magistrskega dela. V diplomskem delu smo se ukvarjali predvsem z analizo atributov, ki jih lahko uporabimo za analizo stilov igre šahistov. Z atributi smo poskušali sestaviti pravila, kdaj nek šahist igra dobro in kdaj slabo, v tem delu pa te attribute uporabimo za primerjavo šahistov. Z njimi poskušamo definirati stil šahista.

1.2 Struktura magistrskega dela

Magistrsko delo vsebuje 5 poglavji. Poleg poglavja Uvod (1) še poglavja Podatki (2), Metode (3), Rezultati (4) in Zaključek (5).

Poglavje Podatki vsebuje opis podatkov. Najprej smo v podpoglavju 2.1 podali nekaj splošnih podatkov o podatkovnih bazah, ki smo jih uporabili. V podpoglavju 2.2 smo opisali attribute podatkov, ki smo jih uporabili v analizi, v podpoglavju 2.3 pa profile igralcev s katerimi smo definirali stil igralca.

V poglavju 3 smo opisali metode, ki smo jih uporabljali pri analizi. V podpoglavju 3.1 smo se posvetili računanju podobnosti. Opisani sta kosinusna podobnost in Pearsonov korelacijski koeficient. Naslednji dve podpoglavji se nanašata na razvrščanje v skupine. V podpoglavju 3.2 smo opisali hierarhično razvrščanje, v podpoglavju 3.3 pa metodo voditeljev. V podpoglavju 3.4 smo opisali povprečno medsebojno informacijo, v podpoglavju 3.5 entropijo in v podpoglavju 3.6 standardni odklon.

V poglavju 4 so rezultati dela. V podpoglavju 4.1 smo opisali iskanje podobnih igralcev. Najprej smo podali rezultate evalvacije metode, nato smo komentirali nekaj zanimivih rezultatov iskanja podobnih igralcev. V podpoglavju 4.2 so rezultati računanja povprečne medsebojne informacije, s katerimi smo si pomagali pri gradnji profilov. V podpoglavju 4.3 so opisani rezultati in komentar razvrščanja velemojstrov v skupine.

V zaključku (5) smo povzeli rezultate dela in podali predloge kako bi se analizo stila igre šahistov dalo izboljšati in razširiti.

Delo ima eno prilogo: Rezultati iskanja podobnih igralcev (A).

Poglavje 2

Podatki

2.1 Splošno o podatkih

Za analizo stilov igre šahistov smo uporabili dve bazi podatkov. Eno za svetovne šahovske prvake in eno za ostale igralce, ki se po kvaliteti igre zelo razlikujejo. Analizirali smo le podatke o partijah med 15 in 30 potezo. Te poteze smo izbrali zato, da smo izločili otvoritev in celotno končnico. Otvoritve smo izločili, ker jih upoštevamo s pomočjo atributov. Končnice pa nismo analizirali, da smo ohranili preprostost analize.

2.1.1 Svetovni šahovski prvaki

Za svetovne šahovske prvake smo imeli podatke o partijah odigranih na prvenstvih za svetovne šahovske prvake. Podatki so torej o partijah, ki so jih svetovni šahovski prvaki igrali proti drugim svetovnim šahovskim prvakom ali kandidatom za prvaka. To so torej zelo zahtevne partije, na katere se šahisti dalj časa intenzivno pripravljajo, pri čemer jim je nasprotnik znan v naprej.

Uporabljali smo podatke za 15 svetovnih šahovskih prvakov. Podatki so predstavljeni v tabeli 2.1.

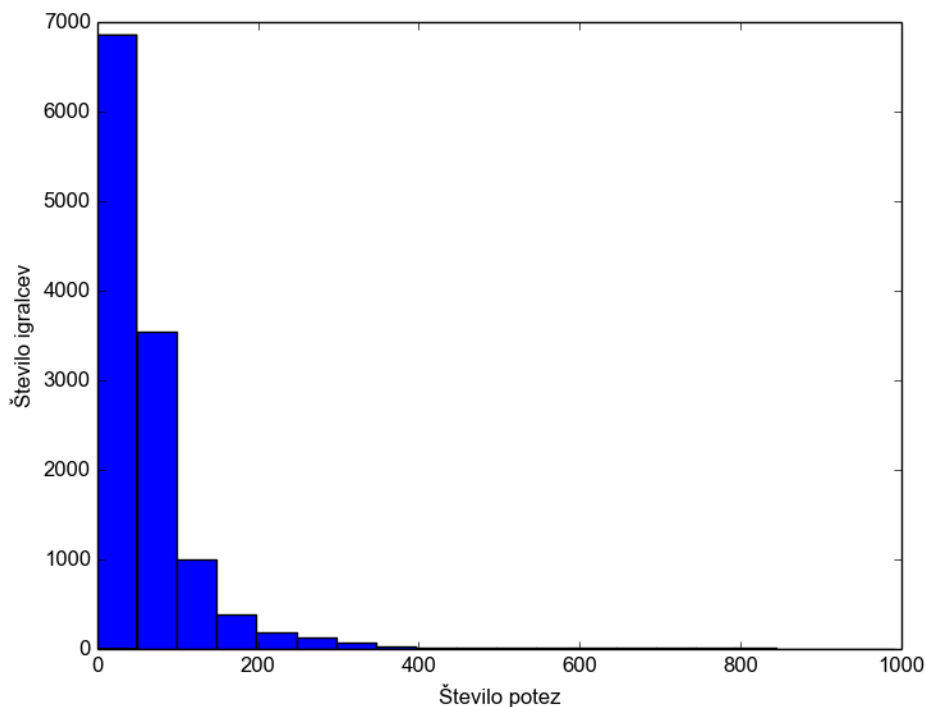
Šahist	Število potez	Število dobrih potez	Število slabih potez
Alekhine	1647	1124	257
Anand	605	482	56
Botvinnik	2176	1429	326
Capablanca	587	456	40
Euwe	828	559	169
Fischer	252	188	29
Karpov	2290	1643	245
Kasparov	2250	1668	235
Kramnik	583	428	59
Lasker	1165	790	176
Petrosian	845	592	86
Smyslov	1079	737	169
Spassky	842	559	114
Steinitz	1290	813	279
Tal	529	353	83

Tabela 2.1: Pregled podatkovne baze o svetovnih šahovskih prvakah

2.1.2 Ostali igralci

Podatki o ostalih igralcih so iz baze *The Week in Chess* za časovno obdobje od 17. 7. 2012 do 18. 6. 2013. Podatki vsebujejo veliko šahistov z zelo malo odigranimi potezami in malo šahistov z veliko potezami. Podatki vsebujejo 722035 potez za 12271 igralcev.

Podatki so predstavljeni na grafih 2.1 in 2.2.



Slika 2.1: Število potez za igralce.

2.2 Atributi podatkov

V tem poglavju so opisani atributi, ki smo jih uporabljali za analizo podatkov. Nekaj teh atributov je bilo podanih v podatkih. Definirali smo tudi nekaj dodatnih, opisnih atributov, ki so nam olajšali analizo.

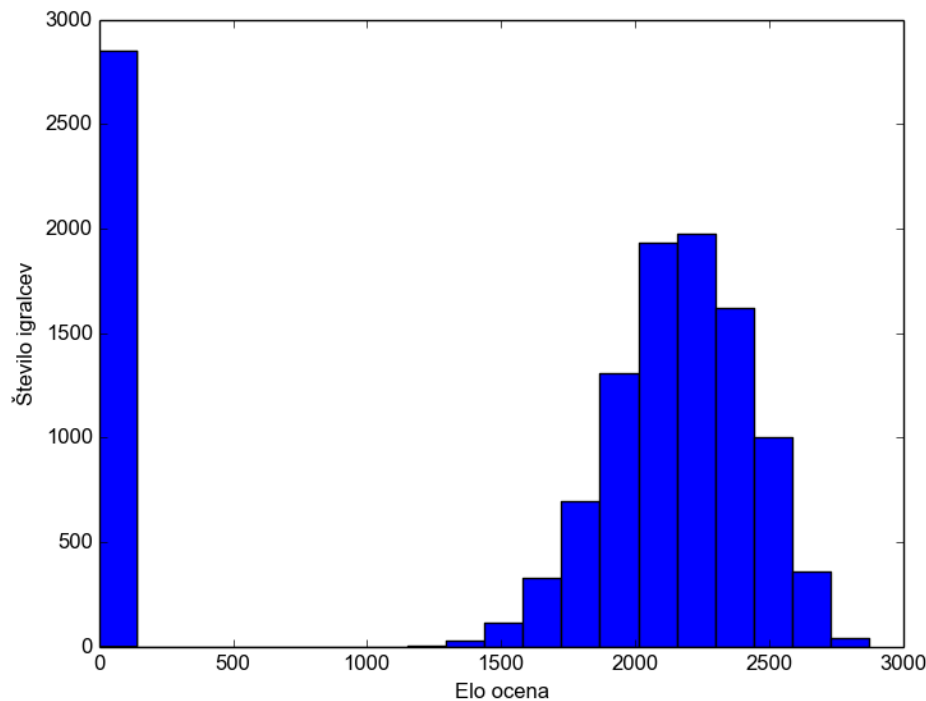
2.2.1 Atributi prisotni v podatkih

Atributi, ki so bili prisotni v podatkih so:

Player Ime šahista, ki je odigral potezo.

Event Ime dogodka, na katerem je šahist odigral potezo.

GameNo Zaporedna številka partije v kateri je bila odigrana poteza določenega šahista, na določenem dogodku.



Slika 2.2: Elo ocene igralcev

MoveNo Zaporedna številka poteze v odigrani partiji.

Side Barva figur s katerimi je igral šahist.

FEN Opis postavitve figur na šahovnici v Forsyth-Edwardsovem zapisu (glej poglavje 2.2.1.2).

MovePlayed Poteza, ki jo je šahist odigral.

MovePlayedValue Ocena odigrane poteze. Ocenjevanje potez je opisano v poglavju 2.2.1.1.

BestMove Najboljša poteza, ki bi jo šahist lahko odigral. Kako je najboljša poteza ugotovljena je opisano v poglavju 2.2.1.1.

BestMoveValue Ocena najboljše možne poteze. Ocenjevanje potez je opisano v poglavju 2.2.1.1.

ECO Opis otvoritve odigrane pri dani partiji po sistemu ECO (glej poglavje 2.2.1.3).

ELO Ocena šahista po sistemu ocenjevanja Elo (glej poglavje 2.2.1.4).

2.2.1.1 Ocena poteze

S šahovskim programom *Houdini* smo ocenjevali odigrane poteze in ugotavljali najboljšo potezo.

Houdini je trenutno eden najmočnejših in najboljših šahovskih programov (vir: [19]).

Opis šahovskega programa *Houdini* je povzet po viru [17].

Program oceni poteze na podlagi preiskovanja drevesa možnih potez. Izračuna možnosti za zmago pri vsaki potezi in oceni kako dobra je bila odigrana poteza. Poišče tudi optimalno potezo.

Poteze so bile ocenjene na globini preiskovanja 20.

Ocene so podane s stotinkami vrednosti kmeta in enoto za podajanje vrednosti imenujemo *centipawn* (slovensko centikmet). Napaka 100 centipawn je enakovredna izgubi enega kmeta.

2.2.1.2 FEN

Opis Forsyth-Edwardsovega zapisa je povzet po viru [6].

FEN, oziroma Forsyth-Edwardsov zapis sta razvila David Forsyth in Steven J. Edwards, za zapis podatkov o neki šahovski poziciji. Zapis sestavlja 6 podatkov:

Pozicije figur Kako so figure postavljene na šahovnici pri čemer so figure označene s črkami navedenimi v tabeli 2.2. Za belega se uporabljajo velike črke, za črnega pa male. Pozicije figur, so označene tako, da je za vsako vrstico na šahovnici od zgoraj navzdol za vsako polje zapisana figura, ki je na njem ali pa je označeno število praznih mest. Vrste so ločene z znakom ”/”.

Kdo je na potezi Kdo je trenutno na potezi, pri čemer z znakom w označi belega in b črnega.

Možnost rokade Kje je še možno narediti rokado.

En passant Kje, in ali je možnost poteze en passant.

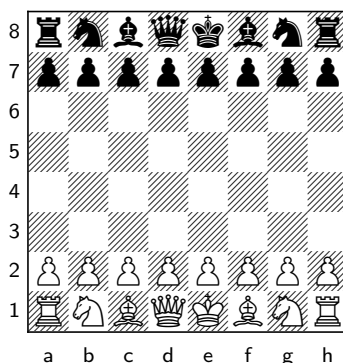
Število pol potez Število pol potez od zadnjega premika kmeta, ali zadnjega jemanja figure.

Število potez Število vseh odigranih potez.

Podatki so v zapisu ločeni s presledkom.

FEN za začetno postavitev šahovnice (glej slika 2.3) je:

`rnbqkbnr/pppppppp/8/8/8/8/PPPPPPP/RNBQKBNR w KQkq - 0 1`



Slika 2.3: Začetna postavitev šahovnice.

Za nas je najpomembnejši podatek Pozicija figur, saj z njegovo pomočjo ugotavljamo vrednosti nekaterih atributov opisanih v poglavju 2.2.2.

2.2.1.3 ECO kode

Opis ECO (ang: *Encyclopaedia of Chess Openings*) kod je povzet po viru [7].

Figura	Vrednost	Oznaka
Kmet	1	P
Skakač	3	N
Lovec	3	B
Trdnjava	5	R
Kraljica	9	Q
Kralj	/	K

Tabela 2.2: Vrednosti šahovskih figur.

Enciklopedija šahovskih otvoritev je publikacija, ki poda klasifikacijo šahovskih otvoritev. Otvoritve razdeli na 5 skupin od *A* do *E*, in te nato v podskupine od 00 do 99. S tem so sistematično klasificirane vse otvoritve. S pomočjo ECO kod smo definirali atribut *Otvoritev*, opisan v poglavju 2.2.2.5.

2.2.1.4 Elo ocenjevalni sistem

Opis Elo ocenjevalnega sistema je povzet po viru [8].

Elo ocenjevalni sistem je razvil madžarski šahist Arpad Elo. Z njim je izboljšal tedanji šahovski ocenjevalni sistem. Elo ocenjevalni sistem se danes uporablja tudi na drugih področjih, predvsem pri športih.

Pri tem sistemu vsak šahist dobi Elo oceno relativno na druge šahiste s katerimi je igral. Ocena se popravlja glede na to ali šahist izgublja ali zmaguje partije. Če šahist z nizko oceno premaga šahista z visoko oceno, potem šahist z nizko oceno pridobi veliko ratinških točk, šahist z visoko oceno pa izgubi veliko ratinških točk. Če sta šahista bolj izenačena, je tudi izmenjava ratinških točk manjša.

V naših podatkih so podane Elo ocene, kot jih je izračunala organizacija FIDE (fr: *Fédération Internationale des Échecs*). Pri sistemu Elo imajo svetovni vele mojstri običajno oceno nad 2500 ratinških točk. Največja dosežena ocena do sedaj je bila 2882 ratinških točk. Maja 2014 jo je dosegel šahist Magnus Carlsen.

2.2.2 Dodatni opisni atributi

To poglavje opisuje opisne attribute, ki smo jih definirali za lažjo analizo podatkov. Atributi kompenzacija za materialni primanjkljaj, simetrična kmečka struktura, material, prisotnost kraljice, prisotnost lovskega para, blokiran center, dvojni kmetje, osamljen damin kmet, materialno neravnovesje, relativna ocena poteze in razred so enaki, kot atributi, ki sem jih uporabila v svojem diplomskem delu *Računalniška analiza človekovih napak pri igranju šaha* (vir [5]). Njihovi opisi so povzeti po opisih v tem delu.

Dodali smo tudi 3 nove attribute: material zvezno, materialno neravnovesje zvezno in otvoritev.

Opisni atributi, ki smo jih dodali za lažjo analizo so:

Simetrična kmečka struktura Atribut imenujemo *PawnSymmetry*. Zavzame lahko dve možni vrednosti. *Yes*, kadar so kmetje postavljeni simetrično in *No* v nasprotnem primeru. Kmečka struktura je simetrična, kadar je v vsaki liniji enako število kmetov vsake barve.

Material Atribut imenujemo *Material*. Zavzame lahko dve različni vrednosti. *High*, kadar ima šahist veliko materiala in *Low* v nasprotnem primeru. Opis računanja materiala sledi v poglavju 2.2.2.4.

Material zvezno Atribut imenujemo *MaterialContinuous*. Atribut je zvezna različica atributa *Material* in zavzame celoštevilsko vrednost med 0 in 31. Opis računanja materiala sledi v poglavju 2.2.2.4.

Prisotnost kraljice Atribut imenujemo *HasQueen*. Opredelimo ga z dvema različnima vrednostima. Z *Yes*, če je šahistova kraljica prisotna na šahovnici, in *No*, če kraljice ni.

Prisotnost lovskega para Atribut imenujemo *HasBishopPair*. Zavzame lahko dve možni vrednosti. *Yes*, če ima šahist na šahovnici oba lovca, nasprotnik pa ne, in *No* v vseh ostalih primerih.

Blokiran center Atribut imenujemo *BlockedCenter*. Zavzame lahko dve možni vrednosti. *Yes*, kadar je center šahovnice blokiran, *No*, kadar ni. Opis blokirane centra sledi v poglavju 2.2.2.1.

Dvojni kmetje Podatek je predstavljen z atributom *DoublePawnsPlayer* za šahista, ki je opravil potezo in *DoublePawnsOpponent* za njegovega nasprotnika. Vsak izmed atributov ima vrednost *Yes*, če šahist, ki ga analiziramo ima dvojne kmete in *No* v nasprotnem primeru. Dvojni kmetje so prisotni, ko sta dva kmeta enake barve na isti liniji.

Osamljen damin kmet Podatek je predstavljen z atributom *IsolatedQueensPawnPlayer* za šahista, ki je opravil potezo in *IsolatedQueensPawnOpponent* za njegovega nasprotnika. Vsak izmed atributov ima vrednost *Yes*, če ima šahist, ki ga analiziramo, osamljenega daminega kmeta in *No*, v nasprotnem primeru. Opis osamljenega daminega kmeta sledi v poglavju 2.2.2.2.

Materialno neravnovesje Atribut imenujemo *Imbalance*. Zavzame lahko dve vrednosti. *Yes*, kadar neravnovesje je, in *No*, kadar ga ni. Opis materialnega neravnovesja sledi v poglavju 2.2.2.3.

Materialno neravnovesje zvezno Atribut imenujemo *ImbalanceContinuous*. Je zvezna različica atributa *Imbalance*. Atribut lahko zavzame celoštevilsko vrednost med 0 in 39. Opis računanja materialnega neravnovesja sledi v poglavju 2.2.2.3.

Kompensacija za materialni primanjkljaj Podatek je predstavljen z dvema atributoma. *CompensatingMoreMaterial*, ki ima vrednost *Yes*, kadar ima šahist več materiala kot nasprotnik, in *No* v nasprotnem primeru ter *CompensatingLessMaterial*, ki ima vrednost *Yes*, kadar ima šahist manj materiala kot nasprotnik, in *No* v nasprotnem primeru. Oba atributa imata avtomatično vrednost *No*, če ima atribut *BestMoveValue* (glej poglavje 2.2.1) vrednost izven intervala $(-50, 50)$, saj v tem primeru igralca kljub neravnovesju materiala nista izenačena.

Torej je materialno neravnovesje vendar ni kompenzacije z boljšo pozicijo.

Relativna ocena poteze Atribut imenujemo *RelativeMoveValue* in zavzame celoštevilsko vrednost, ki je razlika med vrednostma atributov *MovePlayedValue* in *BestMoveValue*, ki sta opisana v poglavju 2.2.1.

Razred Atribut imenujemo *Class*. Ima 3 možne vrednosti. *Good*, kadar odigrano potezo ocenimo kot dobro, *Bad*, kadar odigrano potezo ocenimo kot slabo in *Dubious*, kadar poteze ne moremo oceniti niti kot dobro niti kot slabo. Vrednost atributa določimo tako, da vzamemo absolutno razliko med vrednostma atributov *MovePlayedValue* in *BestMoveValue* in določimo številčne meje za dobre in slabe poteze. Številčne meje so različne. Odvisne so od kakovosti igranja šahistov. Pri nastavljanju mej nam je pomagal ekspert. Za svetovne šahovske prvake smo mejo za dobre poteze določili na vrednost 5 in za slabe poteze na vrednost 10. Za ostale šahiste smo mejo za dobre poteze določili na vrednost 10 in za slabe vrednost 30.

Otvoritev Atribut imenujemo *Opening* in lahko zavzame eno iz med 16 možnih vrednosti. S tem atributom razvrstimo ECO kode otvoritev v tematske skupine, kar nam olajša analizo, saj nam zmanjša število možnih različnih vrednosti. Podrobnejši opis atributa sledi v poglavju 2.2.2.5.

2.2.2.1 Blokiran center

Opis blokiranega centra je povzet po viru [5].

Center šahovnice je lahko blokiran na 4 različne načine:

- Bela kmeta sta na poljih *E4* in *D5*, črna kmeta pa na poljih *E5* in *D6*.
- Bela kmeta sta na poljih *E5* in *D4*, črna kmeta pa na poljih *E6* in *D5*.
- Bela kmeta sta na poljih *E3* in *D4*, črna kmeta pa na poljih *E4* in *D5*.

- Bela kmeta sta na poljih $E4$ in $D3$, črna kmeta pa na poljih $E5$ in $D4$.

2.2.2.2 Osamljen damin kmet

Opis osamljenega daminega kmeta je povzet po viru [5].

Šahist z belimi figurami ima osamljenega daminega kmeta, ko je beli kmet na $D4$ in sam nima drugih svojih kmetov na linijah D in E , črni kmet je na $E6$ in/ali na $C6$ in ni črnih kmetov na liniji D .

Šahist s črnimi figurami ima osamljenega daminega kmeta, ko je črni kmet na $D5$ in sam nima drugih svojih kmetov na linijah D in E , beli kmet je na $E3$ in/ali na $C3$ in ni belih kmetov na liniji D .

2.2.2.3 Materialno neravnovesje

Opis materialnega neravnovesja je povzet po viru [5].

Ko računamo materialno neravnovesje, vsaki figuri, razen kralju, pripišemo vrednost (glej tabelo 2.2). Pogledamo koliko katerih figur je belih in koliko katerih figur je črnih, nato odstranimo prekritja. Torej če ima črni 3 kmete beli pa 2, nam ostane 1 kmet. Vrednosti figur seštejemo. Če je seštevek večji ali enak 5, je na šahovnici prisotno materialno neravnovesje.

2.2.2.4 Material

Opis računanja materiala je povzet po viru [5].

Material računamo tako, da seštejemo vrednosti vseh figur (glej tabelo 2.2), ki jih ima igralec na šahovnici, razen kmetov in kralja. Vrednost, ki jo dobimo, je količina materiala. Če je količina materiala več kot 23, pravimo, da ima igralec veliko materiala.

2.2.2.5 Otvoritev

Atribut otvoritev smo v tem magistrskem delu definirali tako, da smo v skupine razvrstili tematsko podobne otvoritve, glede na njihovo oznako, ECO kodo.

Skupine je definiral šahovski ekspert.

Atribut otvoritev oz. *Opening* lahko zavzame eno iz med 16 možnih vrednosti. Možne otvoritve so opisane v tabeli 2.3.

Oznaka otvoritve	Otvoritev	ECO kode
English opening	Angleška otvoritev	A20-A39
Benoni	Benoni	A56-A79
Dutch defence	Nizozemska obramba	A80-A99
Scandinavian defence	Skandinavska obramba	B00, B01, B06
Alekhine's defence	Alekhineova obramba	B02-B05
Pirc defence	Pirceva obramba	B07-B09
Caro-Kann	Caro-Kann	B10-B19
French defence	Francoska obramba	C00-C19
Sicilian	Sicilijanske otvoritve	B20-B99
Queens gambit	Damin gambit	D00-D69
NimcoBogoIndian Defence	Nimcovičeva in Bogoljubova indijska obramba	E00-E59
Grunfeld Indian defence	Grunfeldova indijska obramba	D70-D99, E60
Kings Indian defence	Kraljeva indijska obramba	E61-E99
Open games	Odprte igre	C20-C99
Other	Ostalo	A00-A19, A40-A55
Unknown	Neznana vrednost	ECO koda otvoritve v podatkih ni prisotna

Tabela 2.3: Opis možnih vrednosti atributa otvoritev

2.3 Profil igralca

Za uspešno analizo stilov igre šahistov smo morali sestaviti profil šahista, ki dovolj dobro definira njegov stil igre. Preizkusili smo več različnih možnih profilov, ki smo jih zgradili na podlagi atributov, opisanih v poglavju 2.2.

Definirali smo 5 različnih profilov:

- Profil A: atributi.
- Profil B: atributi po otvoritvah.
- Profil C: splošni atributi po otvoritvah.
- Profil D: otvoritve.
- Profil E: Skrajšane ECO kode.

Pri iskanju podobnih šahistov smo preizkusili vse profile, pri ostalih poskusih, razvrščanje v skupine, iskanje podobnih velemojstrov, pa le profil C. Profil C je bil pri iskanju podobnih igralcev najuspešnejši izmed profilov, ki uporabljajo attribute. Za nadaljno analizo je bil uporaben le profil, sestavljen na podlagi atributov.

Profile šahistov gradimo posebej za dobre poteze, posebej za slabe poteze. Tako vsak šahist dobi 2 profila. Pri preizkušanju smo ugotovili, da imajo šahisti tako malo slabih potez, da nimamo dovolj podatkov za izračun zanesljivega profila za slabe poteze. Opisi računanja profilov so zato podani za dobre poteze. Za slabe poteze bi računali podobno.

2.3.1 Profil A: atributi

Profil A temelji na atributih, ki smo jih uporabili pri mojem diplomskem delu, ki je predhodnik tega dela (vir: [5]). To je prvi profil, ki smo ga sestavili in ga tudi uporabili za osnovo za nekatere druge profile.

Uporabljeni atributi:

- *CompensatingLessMaterial*.

- *CompensatingMoreMaterial.*
- *PawnSymmetri.*
- *Material.*
- *HasQueen.*
- *HasBishopPair.*
- *BlockedCenter.*
- *DoublePawnsPlayer.*
- *DoublePawnsOpponent.*
- *IsolatedQueensPawnPlayer.*
- *IsolatedQueensPawnOpponent.*
- *Imbalance.*

Za vsakega šahista smo za vsak atribut izračunali delež dobrih potez, pri katerih ima ta atribut pozitivno vrednost - *Yes* ali *High*. Delež potez, kjer ima atribut pozitivno vrednost, izračunamo še za vse poteze, ne le za dobre. Profil A je odstopanje vrednosti atributa za dobre poteze od vrednosti za vse poteze. Celoten profil pa je vektor teh vrednosti za vse attribute, za danega igralca.

2.3.2 Profil B: atributi po otvoritvah

Profil B izhaja iz profila A. Temelji na istih atributih, vendar pri tem profilu upoštevamo še otvoritve definirane z atributom *Opening*. Pri gradnji profila B smo izhajali iz predpostavke, da šahist pri različnih otvoritvah igra z različnimi stili. Na to smo sklepali iz dejstva, da nekatere otvoritve že po svoji naravi narekujejo bolj mirno, nekatere pa bolj napadalno igro. Za vsako izmed otvoritev opisanih v poglavju 2.2.2.5 smo izračunali profil A (poglavje 2.3.1). Profil B tako sestavlja množica profilov A.

2.3.3 Profil C: splošni atributi po otvoritvah

Profil C računamo na enak način, kot računamo profil B (poglavje 2.3.2), le da pri tem profilu uporabimo druge attribute. Delno so to enaki atributi kot tisti uporabljeni pri profilu B, del jih odstranimo, dodamo pa nekaj novih. Pri tem smo izbirali attribute, ki nam podajo čim bolj splošne informacije o igri in ne toliko specifične podatke o trenutnem dogajanju na šahovnici.

Uporabljeni atributi:

- *CompensatingLessMaterial.*
- *CompensatingMoreMaterial.*
- *PawnSymmetri.*
- *ImbalanceContinuous.*
- *MaterialContinuous.*
- *RelativeMoveValue.*
- *GoodMoves.*
- *BadMoves.*

Uporabljeni atributi so opisani v poglavju 2.2.2, razen zadnjih dveh. To sta *GoodMoves*, ki predstavlja delež dobrih potez pri dani otvoritvi za danega šahista in *BadMoves*, ki predstavlja delež slabih potez pri dani otvoritvi za danega šahista.

2.3.4 Profil D: otvoritve

Profil D je namenjen preizkušanju, koliko imajo uporabljeni atributi vpliva na analizo pri profilih B (poglavje 2.3.2) in C (poglavje 2.3.3). Profil D je sestavljen zelo preprosto. Za vsako od definiranih otvoritev preštejemo število dobrih potez. Profil D je vektor teh seštevkov. Vektorja ne normaliziramo, saj zaradi načina računanja razdalj med šahisti to ni potrebno.

Uporabljamo kosinusno podobnost (poglavje 3.1.1) in Pearsonov korelacijski koeficient (poglavje 3.1.2).

2.3.5 Profil E: Skrajšane ECO kode

Profil E računamo enako kot profil D (poglavje 2.3.4), le da ne uporabimo definiranih otvoritev, ampak prva dva znaka ECO kode (poglavje 2.2.1.3).

$$E15 \rightarrow E1$$

Ker so otvoritve, klasificirane po ECO kodah, že tematsko razdeljene, s tem načinom dobimo nekoliko bolj fino grupiranje otvoritev, kot z atributom *Opening* (poglavje 2.2.2.5).

Pri profilu E vsakega šahista predstavlja vektor števila dobrih potez za vsako skrajšano ECO kodo.

Poglavje 3

Metode

3.1 Računanje podobnosti

3.1.1 Kosinusna podobnost

Opis kosinusne podobnosti je povzet po viru [1].

S kosinusno podobnostjo merimo podobnost med dvema vektorjema tako, da gledamo kot med njima. Če oba vektorja kažeta v isto smer, je njuna podobnost 1. Če sta ortogonalna drug na drugega, je njuna podobnost 0, če pa kažeta v nasprotni smeri, je njuna podobnost -1 . Dolžina vektorja ni pomembna. Ta način merjenja podobnosti stila igre šahistov nam ustreza, saj ne iščemo šahistov, ki so enako dobri, ampak šahiste, ki bodisi dobro bodisi slabo igrajo z enakim stilom.

Pri računanju kosinusne podobnosti nam pomaga formula za skalarni produkt dveh vektorjev (glej enačbo 3.1). Iz te enačbe in enačbe za dolžino vektorja (glej enačbo 3.2) lahko izpeljemo formulo za kosinusno podobnost (glej enačbo 3.3).

$$\vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}| \cos(\phi) = \sum_{i=1}^n a_i b_i \quad (3.1)$$

$$|\vec{a}| = \sqrt{\vec{a} \cdot \vec{a}} \quad (3.2)$$

$$\text{podobnost} = \cos(\phi) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}||\vec{b}|} = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}} \quad (3.3)$$

3.1.2 Pearsonov korelacijski koeficient

Opis Pearsonovega korelacijskega koeficienta je povzet po viru [1].

Podobno kot kosinusna podobnost tudi Pearsonov korelacijski koeficient meri pozitivno in negativno korelacijo. Če sta dva vektorja pozitivno korelirana, ima Pearsonov korelacijski koeficient vrednost 1, če sta negativno korelirana, ima vrednost -1 , če pa korelacije ni možno ugotoviti, ima vrednost 0.

Vrednost Pearsonovega korelacijskega koeficienta računamo podobno kot kosinusno podobnost, le da vektorja normaliziramo tako, da vsakemu elementu vektorja odštejemo povprečno vrednost vseh elementov vektorja (glej formulo 3.4). Pearsonov korelacijski koeficient običajno označujemo z r .

$$r = \frac{\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2} \sqrt{\sum_{i=1}^n (b_i - \bar{b})^2}} \quad (3.4)$$

3.1.3 Združevanje več podobnosti

Podobnosti med šahisti smo računali z uporabo enega izmed možnih profilov. Pri profilih A, D in E je bil izračun enostaven, saj je vsak profil sestavljen iz enega vektorja. Podobnost med danima vektorjema izračunamo, z želeno mero podobnosti. Profila B in C sta bolj kompleksna, saj ju sestavlja množica vektorjev. Vsaka otvoritev dobi svoj vektor. Želeli smo upoštevati tudi število potez šahista pri dani otvoritvi. To smo naredili tako, da smo upoštevali ali je šahist sploh kdaj igral to otvoritev, ne pa tudi, kolikokrat jo je igral.

Podobnost za profila B in C smo računali po formuli 3.5. V tej formuli A predstavlja profil prvega šahista, B predstavlja profil drugega šahista, O

je množica vseh otvoritev, o je trenutna otvoritev, o_A je število potez pri dani otvoritvi za prvega šahista in o_B za drugega šahista, $sim(A_o, B_o)$ je podobnost med profiloma trenutne otvoritve, za poljubno mero podobnosti.

$$s = \frac{\sum_{o \in O} \begin{cases} sim(A_o, B_o) & ; o_A > 0 \text{ in } o_B > 0 \\ 0 & ; \text{ostali primeri} \end{cases}}{\sum_{o \in O} \begin{cases} 0 & ; o_A = 0 \text{ in } o_B = 0 \\ 1 & ; \text{ostali primeri} \end{cases}} \quad (3.5)$$

Z uporabo te formule, tistih otvoritev, kjer nobeden izmed šahistov nima potez, sploh ne upoštevamo. Če ima poteze pri neki otvoritvi le eden izmed šahistov, pa se podobnost avtomatično šteje kot 0. V tem primeru ne moremo sklepati ne na pozitivno in ne na negativno korelacijo, ker nimamo dovolj podatkov.

Prikazan način računanja podobnosti da uporabne rezultate le za mere podobnosti, pri katerih lahko vidimo tako negativno korelacijo z vrednostjo -1 , kot tudi pozitivno korelacijo z vrednostjo 1 . Obe meri podobnosti, Pearsonov korelacijski koeficient in kosinusna podobnost, to omogočata.

3.2 Hierarhično razvrščanje

Opis hierarhičnega razvrščanja je povzet po viru [2].

Hierarhično razvrščanje je metoda za razvrščanje entitet, v našem primeru šahistov, v skupine. Pri tej metodi razvrščanja v skupine, zgradimo drevesno strukturo glede na podobnost posameznih entitet z drugimi entitetami. Ker zgradimo drevesno strukturo, rezultat običajno prikažemo s pomočjo dendrograma.

Algoritem deluje po sledečem postopku:

1. **korak:** Vsako entiteto razvrstimo v svojo skupino; tako imamo na začetku toliko skupin, kot je entitet.
2. **korak:** Izračunamo razdalje med vsemi skupinami.

3. korak: Najbolj podobni skupini združimo v eno skupino.

4. korak: Koraka 2 in 3 ponavljamo dokler nam ne ostane le še ena skupina.

Da postopek poenostavimo, ne računamo razdalj med skupinami vsakič ponovno, ampak na začetku izračunamo razdalje med vsemi entitetami. Razdalje med skupinami pa računamo s pomočjo že izračunanih razdalj med entitetami v skupinah.

Nekaj pogostih načinov računanja razdalj med skupinami:

Maksimalna razdalja (*complete linkage*) Razdalja med skupinama je največja razdalja med dvema entitetama, pri čemer je prva iz prve skupine, drugi pa iz druge.

Minimalna razdalja (*single linkage*) Razdalja med skupinama je najmanjša razdalja med dvema entitetama, pri čemer je prva iz prve skupine, drugi pa iz druge.

Povprečna razdalja (*average linkage*) Razdalja med skupinama je povprečna razdalja med dvema entitetama, pri čemer je prva iz prve skupine, drugi pa iz druge.

3.3 Metoda voditeljev

Opis metode voditeljev (*k-means*) je povzet po virih [9, 10].

Metoda voditeljev je metoda za razvrščanje entitet, šahistov, v skupine, kjer število skupin eksplicitno določimo. Število skupin označujemo s k .

Metoda deluje po sledečem postopku:

1. korak: Izberemo k voditeljev. To so lahko naključne entitete iz množice entitet, ki jih razvrščamo v skupine. Lahko jih izberemo z nekim določenim postopkom, recimo tako da poiščemo najbolj oddaljene eden od drugega. Lahko jih izračunamo iz ostalih entitet.

2. **korak:** Vse entitete razvrstimo v k skupin, tako da jih razdelimo k najbližjim voditeljem.
3. **korak:** Nov voditelj je centroid dobljene skupine.
4. **korak:** Ponavljamo koraka 2 in 3 dokler ni število primerov, katerih skupina se spremeni, dovolj majhno.

Z metodo voditeljev ne najdemo vedno optimalne rešitve, saj se lahko ujame v lokalni minimum. Ta problem rešujemo tako, da algoritem poženemo večkrat, in vsakič naključno določimo voditelje. Za rezultat vzamemo najboljšo razvrstitev v skupine. Najboljša je tista, kjer je razdalja med primeri znotraj skupine najmanjša.

3.4 Povprečna medsebojna informacija

Opis povprečne medsebojne informacije (ang. *Information Gain*) je povzet po viru [3].

Povprečna medsebojna informacija nam pove koliko informacije nam da določen atribut. Tako ugotavljamo kako pomemben je kateri od atributov za našo analizo. Povprečno medsebojno informacijo, ki jo označujemo $IG(T, a)$ izračunamo po formuli 3.6. V formuli je T množica vseh primerov za katere računamo povprečno medsebojno informacijo, a je atribut katerega povprečno medsebojno informacijo računamo in $H(X)$ je entropija spremenljivke X . Entropija je opisana v poglavju 3.5.

$$IG(T, a) = H(T) - H(T|a) \quad (3.6)$$

3.5 Entropija

Opis entropije je povzet po viru [3]

Entropija je mera nedoločenosti sistema. Če imamo sistem z večimi možnimi stanji, ki so vsa enako verjetna, potem je sistem zelo nedoločen. Zelo težko napovemo stanje tega sistema. Če pa imajo vsa stanja tega sistema

verjetnost 0, razen enega, ki ima verjetnost 1, potem sistem ni nedoločen, entropija takšnega sistema je 0. Entropijo označujemo $H(X)$ pri čemer je X spremenljivka, katere entropijo merimo. Entropijo izračunamo po formuli 3.7, v kateri je $P(X = x)$ verjetnost, da ima spremenljivka X vrednost x .

$$H(X) = - \sum_{x \in X} P(X = x) \log_2 P(X = x) \quad (3.7)$$

Poznamo tudi pogojno entropijo, ki jo označujemo s $H(X|Y)$, pri čemer računamo nedoločenost spremenljivke X , če nam je vrednost spremenljivke Y poznana. Pogojno entropijo računamo s pomočjo formule 3.8, v kateri je $P(X = x|Y = y)$ verjetnost, da ima spremenljivka X vrednost x , če vemo, da ima spremenljivka Y vrednost y .

$$\begin{aligned} H(X|Y) &= \sum_{y \in Y} P(Y = y) H(X|Y = y) \\ &= - \sum_{y \in Y} P(Y = y) \sum_{x \in X} P(X = x|Y = y) \log_2 P(X = x|Y = y) \end{aligned} \quad (3.8)$$

3.6 Standardni odklon

Opis standardnega odklona je povzet po viru [4].

S pomočjo standardnega odklona lahko ugotovimo kakšen razpon imajo vrednosti podatkov. Standardni odklon računamo s pomočjo ugotavljanja odstopanja vrednosti od \bar{x} , ki je lahko povprečje vrednosti, ali mediana. Povprečje izračunamo po formuli 3.9, v kateri je N število podatkov, x_i pa je i -ti podatek. Mediano izračunamo tako, da vse podatke razporedimo po velikosti in vzamemo srednjo vrednost.

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} \quad (3.9)$$

Standardni odklon, ki ga običajno označujemo z črko σ , izračunamo s pomočjo formule 3.10.

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}} \quad (3.10)$$

Poglavje 4

Rezultati

4.1 Iskanje podobnih šahistov

Podobne šahiste smo iskali tako, da smo računali podobnosti med vsemi pari šahistov. Poskušali smo tako z uporabo kosinusne razdalje kot s Pearsonovim korelacijskim koeficientom. Metodi računanja razdalj sta dali pričakovane rezultate. Pri profilih A, B in C je dala boljše rezultate kosinusna razdalja, saj vrednosti med seboj niso povezane. Pri profilih D in E je dal boljše rezultate Pearsonov korelacijski koeficient, saj takrat med vrednostmi v vektorju, ki predstavlja profil, obstaja korelacija.

4.1.1 Evalvacija metode

Metodo smo evalvirali tako, da smo poteze nekaj šahistov razdelili v dve skupini in poskušali ugotoviti katere poteze pripadajo istim šahistom.

Za analizo smo izbrali tiste šahiste, pri katerih smo imeli na voljo zadostno količino potez, ki so bile poprej predmet računalniške analize s šahovskim programom *Houdini*. Mejo števila potez za analizo enega šahista smo nastavili na 252, saj ima toliko potez svetovni šahovski prvak, ki ima najmanj potez - Fischer. Če želimo tega svetovnega šahovskega prvaka primerjati s komerkoli drugim, moramo za to število potez ugotoviti, ali je metoda iskanja podobnih šahistov učinkovita.

Ker smo poteze vsakega šahista razdelili na dve skupini, je moral vsak izmed analiziranih šahistov imeti v bazi vsaj 504 poteze. Tako smo analizirali 70 igralcev.

Poteze vsakega šahista smo razdelili na dve skupini. To smo naredili naključno, vendar tako, da smo ohranili poteze ene partije v isti skupini.

Metodo smo evalvirali tako, da smo za vsakega šahista v prvi skupini izračunali podobnost z vsemi šahisti v drugi skupini. Podobnost smo računali po metodah opisanih v poglavju 3.1. Šahistu iz prve skupine smo poiskali šahista iz druge skupine, ki mu je bil najbolj podoben. Če sta bila ta dva šahista enaka, torej če je bil igralec s partijami iz prve skupine najbolj podoben samemu sebi s partijami iz druge skupine, smo to vzeli za pozitiven rezultat, saj smo našli ujemanje. Pri tem predvidevamo, da ima isti šahist enak stil pri igranju različnih partij. Slednje seveda velja le v določeni meri, zato ni bilo moč pričakovati popolnega ujemanja.

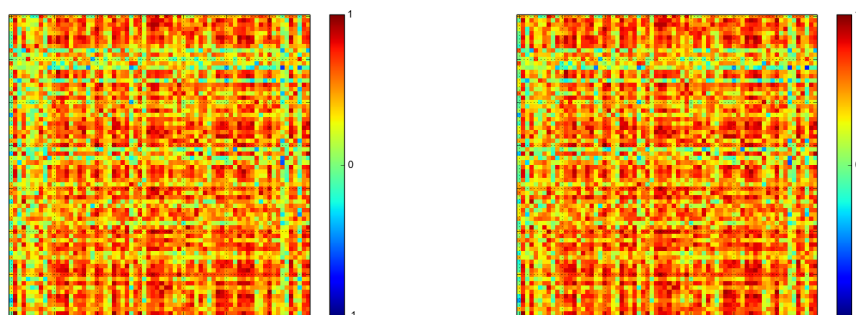
Analizo smo ponovili za vseh 5 profilov opisanih v poglavju 2.3. Povzetek rezultatov je viden v tabeli 4.1.

Profil	Kosinusna podobnost	Pearsonov korelacijski koeficient
Profil A	2	1
Profil B	3	2
Profil C	19	18
Profil D	15	17
Profil E	20	20

Tabela 4.1: Rezultati testiranja iskanja podobnih Šahistov. Število ujemanj za vsak profil pri dani meri podobnosti. Analizirane so bile poteze 70 šahistov.

Sledijo grafi podobnosti šahistov iz ene skupine s šahisti iz druge skupine. Na osi x so razporejeni šahisti iz prve skupine, na osi y pa šahisti iz druge skupine, vendar v enakem vrstnem redu. Modra barva pomeni negativno korelacijo - šahista se ne ujemata. Rdeča barva pomeni pozitivno korelacijo

- šahista se ujemata. Da je bila za šahista ugotovljena podobnost, za partije iz obeh skupin kaže kvadrataček na diagonali, ki je bolj rdeč kot tisti drugje na liniji, ki predstavlja tega šahista.



(a) Kosinusna razdalja

(b) Pearsonov korelacijski koeficient

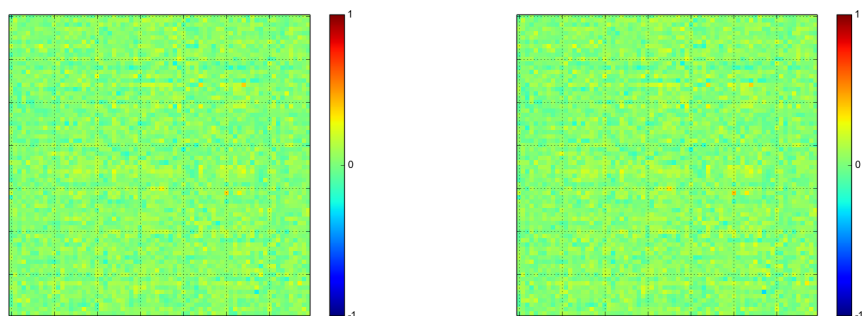
Slika 4.1: Rezultati testiranja iskanja podobnih šahistov za profil A.

Na sliki 4.1 so prikazani rezultati testiranja iskanja podobnih šahistov glede na profil A. Z uporabo kosinusne razdalje se ujemata 2 šahista, na osnovi Pearsonovega korelacijskega koeficienta pa le 1. Testiranje glede na profil A da slabe rezultate, saj bi takšne rezultate zlahka dobili tudi po naključju.

Profil A smo poskušali izboljšati tako, da smo ga razdelili na otvoritve, kot je opisano v poglavju 2.3.2. S tem smo dobili profil B.

Na sliki 4.2 so prikazani rezultati testiranja iskanja podobnih šahistov glede na profil B. Z uporabo kosinusne razdalje se ujemajo 3 igralci, z uporabo Pearsonovega korelacijskega koeficienta pa 2. S testiranjem glede na profil B nam ni uspelo pomembno izboljšati rezultatov glede na profil A. Glavna razlika pri rezultatih, glede na profil A, je da so vrednosti kosinusne podobnosti in Pearsonovega korelacijskega koeficienta bližje 0. Torej ne moremo sklepati ne na pozitivno in ne na negativno korelacijo.

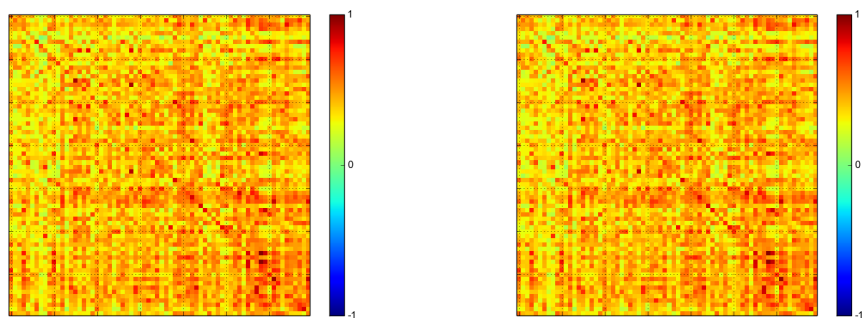
Profil B smo izboljševali z izbiro drugih atributov. Pri tem smo si pomagali s povprečno medsebojno informacijo za vsak atribut. Rezultati računanja povprečne medsebojne informacije so prikazani v poglavju 4.2.



(a) Kosinusna razdalja

(b) Pearsonov korelacijski koeficient

Slika 4.2: Rezultati testiranja iskanja podobnih šahistov za profil B.



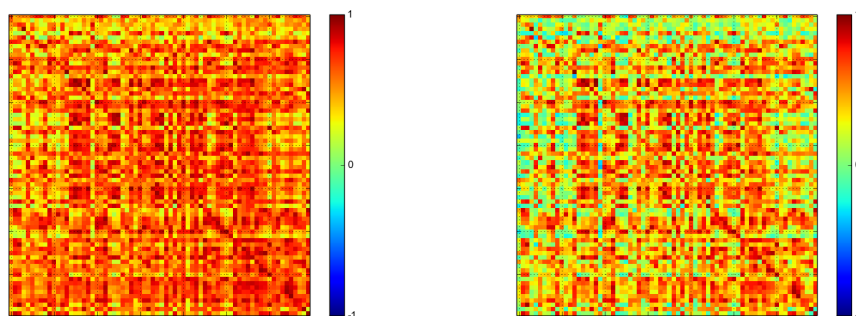
(a) Kosinusna razdalja

(b) Pearsonov korelacijski koeficient

Slika 4.3: Rezultati testiranja iskanja podobnih šahistov za profil C.

Na sliki 4.3 so prikazani rezultati testiranja iskanja podobnih šahistov glede na profil C. Pri profilu C lahko opazimo veliko izboljšanje rezultatov testiranja v primerjavi s profilom B. Z uporabo kosinusne razdalje se ujema kar 19 šahistov, z uporabo Pearsonovega korelacijskega koeficienta pa 18. Na sliki je vidna rdeča diagonalna linija, ki nakazuje ujemanje.

Na sliki 4.4 so prikazani rezultati iskanja podobnih šahistov glede na profil D. Profil D je bil namenjen preizkušanju profila C, in sicer smo želeli ugotovi, koliko k rezultatom primorejo atributi, in koliko vpliva le dejstvo, da je šahist igral neko otvoritev. Z uporabo kosinusne razdalje se ujema 15 igralcev, z



(a) Kosinusna razdalja

(b) Pearsonov korelacijski koeficient

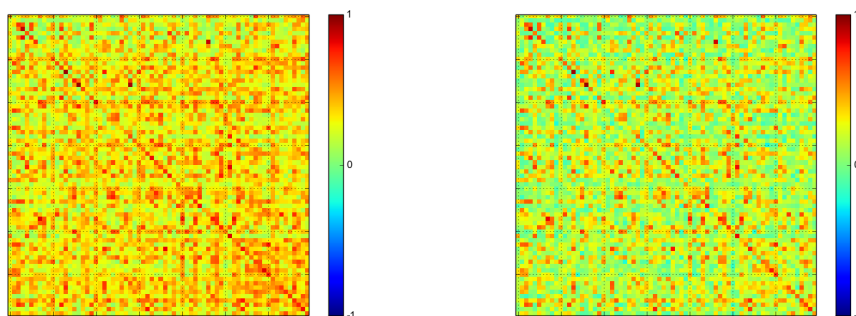
Slika 4.4: Rezultati testiranja iskanja podobnih šahistov za profil D.

uporabo Pearsonovega korelacijskega koeficienta pa 17. Ti rezultati so slabši kot pri profilu C.

Lahko sklepamo, da atributi, ki jih uporabimo za profil C, dodajo dovolj informacije, da se rezultati pri analizi izboljšajo. Ugotovimo tudi, da atributi, uporabljeni pri profilu B, nimajo enakega učinka, saj so tam rezultati veliko slabši. Sklepamo lahko, da imajo atributi velik pomen za iskanje podobnih šahistov, saj lahko rezultate tako izboljšajo, kot tudi poslabšajo. Stil igranja šahista ne moremo preprosto enačiti z otvoritvenim repertoarjem šahista.

Pri profilu D se kot boljši izkaže Pearsonov korelacijski koeficient. To je pričakovan rezultat, saj pri računanju Pearsonovega korelacijskega koeficienta podatke normaliziramo. Normalizacija je smiselna le, če so podatki v vsakem vektorju, ki predstavlja šahista na enaki skali. Pri profilih A, B in C normalizacija ni smiselna, saj so elementi v vektorjih pri teh profilih na različnih skalah.

Na sliki 4.5 so prikazani rezultati iskanja podobnih šahistov glede na profil E. Tako pri uporabi kosinusne razdalje, kot Pearsonovega korelacijskega koeficienta se ujema 20 parov. Profil E smo sestavili, ker nas je zanimalo, ali bo večja segmentacija otvoritev, v primerjavi s profilom D, izboljšala rezultate. Rezultati so se res izboljšali. S profilom E smo dosegli najboljše rezultate pri tej analizi. Dosegli smo kar 29% natančnost. Profila E kljub



(a) Kosinusna razdalja

(b) Pearsonov korelacijski koeficient

Slika 4.5: Rezultati testiranja iskanja podobnih šahistov za profil E.

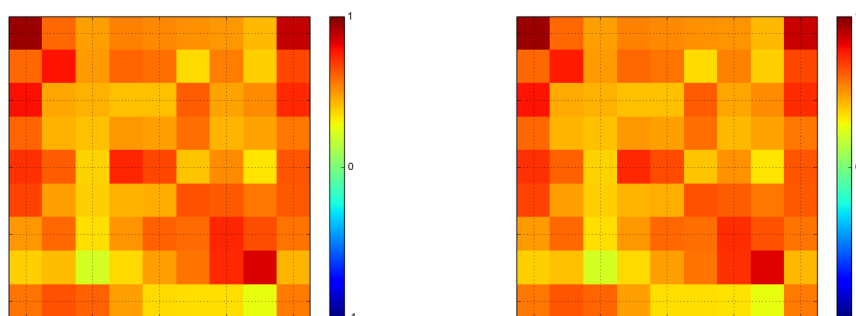
temu nismo uporabili za nadaljnjo analizo, saj nam ne pove nič o stilu igranja šahista, ampak le katere otvoritve rad igra. Če bi bil naš cilj prepoznavanje igralcev, bi bil ta profil pomemben del analize.

Za nadaljnjo analizo smo izbrali profil C, saj nas ne zanima le katere otvoritve igra šahist, ampak tudi kako jih igra. Pri tem profilu smo dosegli 27% natančnost, pri iskanju istih šahistov z uporabo kosinusne razdalje.

4.1.2 Vpliv števila analiziranih potez na natančnost ugotavljanja šahista

Zanimalo nas je, kakšen vpliv ima število analiziranih potez na rezultat. Zato smo analizo s profilom C ponovili tako, da smo zvišali mejo za število potez. Igralec je moral sedaj imeti vsaj 400 potez v vsaki skupini, torej vsaj 800 potez v bazi. Rezultati so prikazani na sliki 4.6.

Tako z uporabo kosinusne razdalje, kot z uporabo Pearsonovega korelacijskega koeficienta zadanemo 4 igralce od 9. Pri tem dosežemo kar 44% natančnost. Vendar pa je igralcev s toliko potezami le 9, kar ni dovolj, da bi lahko z gotovostjo trdili, da povečano število potez res toliko pripomore k izboljšanju rezultatov.



(a) Kosinusna razdalja

(b) Pearsonov korelacijski koeficient

Slika 4.6: Rezultati testiranja iskanja podobnih šahistov za profil C s 400 potezami na skupino.

4.1.3 Primerjava igralcev s svetovnimi šahovskimi prvaki

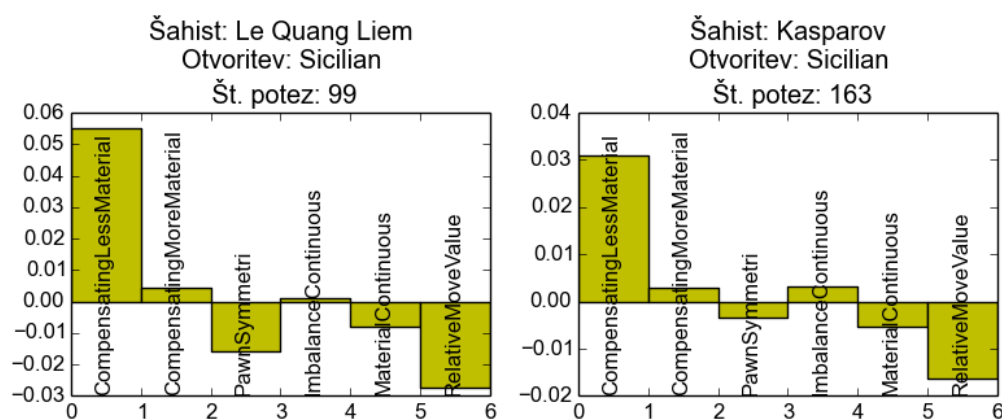
Pri primerjavi šahistov s svetovnimi šahovskimi prvaki šahistom nismo iskali splošno najpodobnejših svetovnih šahovskih prvakov. Analizo smo izvajali ob predpostavki, da otvoritev do neke mere določa stil šahista. Za to smo za vsakega šahista izbrali otvoritev, za katero smo imeli v bazi podatkov največ potez, ker je analiza zanjo najbolj zanesljiva. Šahistom smo nato poiškali svetovnega šahovskega prvaka, ki isto otvoritev igra z najbolj podobnim stilom.

Zaradi zanesljivosti rezultatov smo analizirali le šahiste, ki imajo več kot 400 potez v bazi. Takšnih šahistov je 101. Primerjali smo jih z vsemi svetovnimi šahovskimi prvaki, ki so bili v uporabljeni bazi.

Stil otvoritve smo določili na podlagi profila C, pri čemer dveh atributov nismo upoštevali: *GoodMoves* in *BadMoves*. Teh dveh atributov nismo upoštevali, ker kljub visoko nastavljenim mejam za dobre in slabe poteze, pri svetovnih šahovskih prvakih ne moremo primerjati njihovega deleža dobrih in slabih potez z deleži ostalih šahistov. Svetovni šahovski prvaki imajo še vedno premalo slabih potez.

Rezultat za vsakega šahista je prikazan z 2 grafoma. Prvi graf prikazuje profil danega šahista za izbrano otvoritev. Drugi graf pa profil najpodobnejšega svetovnega šahovskega prvaka za izbrano otvoritev.

Rezultati za vse analizirane šahiste so v prilogi A. V tem poglavju so komentirani le izbrani primeri.



Slika 4.7: Primerjava igralcev in svetovnih šahovskih prvakov: par z največjo podobnostjo.

Prvi primer, ki smo ga izbrali, je tisti, pri katerem je kosinusna podobnost največja, kar 0.985. Tudi z grafa (glej sliko 4.7) lahko vidimo, da sta si šahista zelo podobna.

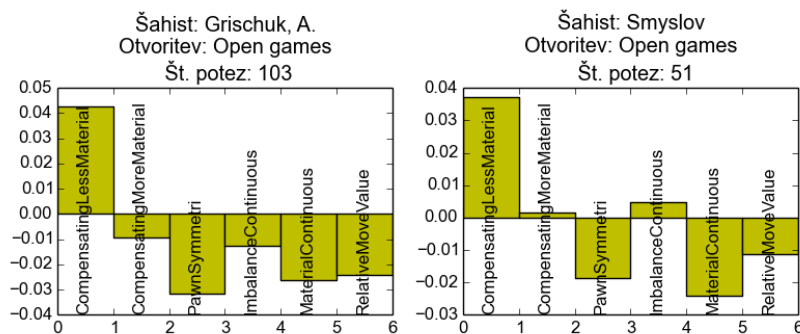
Primerjali smo šahista Le Quang Liema in svetovnega šahovskega prvaka Kasparova. Otvoritev, na osnovi katere ju primerjamo, je sicilijanska otvoritev. Le Quang Liem ima za to otvoritev 99 potez, Kasparov pa 163. Oba imate torej dovolj potez za gradnjo profila.

Značilnosti profila te otvoritve, za ta dva šahista sta visoka vrednost atributa *CompensatingLessMaterial* in nizka vrednost atributa *RelativeMoveValue*. Ostali atributi imajo vrednost bližje 0.

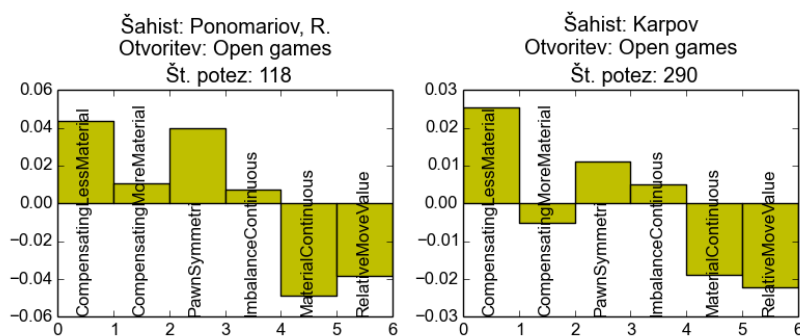
Glavna razlika med svetovnim šahovskim prvakom in navadnim šahistom je ta, da ima svetovni šahovski prvak manjše vrednosti vseh atributov.

Naslednja dva primera, ki smo ju izbrali, prikazujeta dva šahista z isto najpogostejšo otvoritvijo, ki jo igrata z različnima stiloma. Temu primerna

sta tudi podobna svetovna prvaka. Rezultat analize je prikazan na sliki 4.8.



(a) Primerjva najpogostejše otvoritve za šahista: Grischuk, A..
Podobnost: 0.919



(b) Primerjva najpogostejše otvoritve za šahista: Ponomarev, R..
Podobnost: 0.933

Slika 4.8: Ista otvoritev z dvema različnima profiloma.

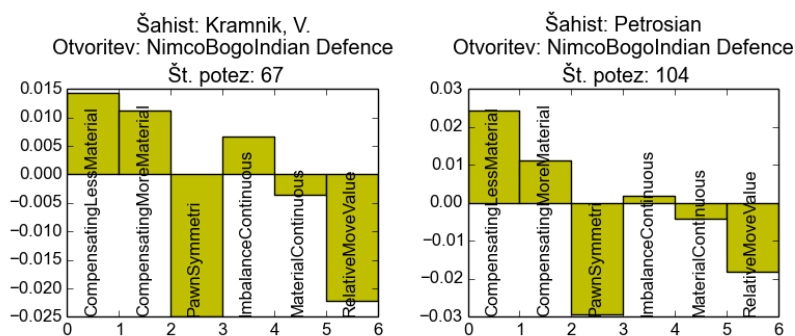
Ta rezultat je pomemben, ker nam pokaže, da ne igrajo vsi šahisti iste otvoritve enako in je edini podatek na osnovi katerega jih lahko ločimo število ponovitev neke otvoritve. Na grafih je vidno, da se profila otvoritve za šahista Grischuka in Ponomariova razlikujeta. Največja razlika je v vrednosti atributa *PawnSymmetry*. Šahist Grischuk in njemu podoben svetovni šahovski prvak Smyslov imata v profilih negativno vrednost tega atributa. Pri Ponomariovu in Karpovu pa je vrednost tega atributa pozitivna.

Za šahista Vladimirja Kramnika smo imeli podatke v obeh podatkovnih bazah. V prvi bazi so bili podatki za partije, ki jih je odigral na prvenstvih

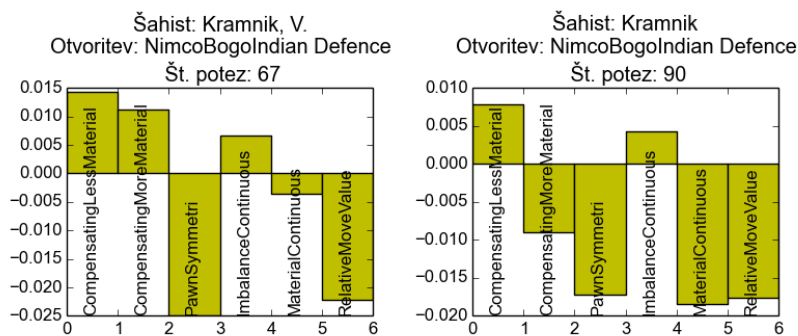
za svetovnega šahovskega prvaka, v drugi pa za partije, ki jih je odigral proti različno rangiranim igralcem.

Na sliki 4.9 so prikazani rezultati analize stila igre Vladimirja Kramnika. Na sliki 4.9a so prikazani rezultati primerjave z najpodobnejšim svetovnim šahovskim prvakom, pri čemer so uporabljeni podatki za Kramnika iz baze ostalih igralcev. Na sliki 4.9b je prikazana primerjava Kramnika samega s seboj za isto otvoritev, pri čemer je levi profil stila iz baze ostalih podatkov in desni iz baze svetovnih šahovskih praks.

Analizirana otvoritev je Nimcovičeva in Bogoljubova indijska obramba.



(a) Kramnik in najpodobnejši svetovni šahovski prvak Petrosian.
Podobnost: 0.961



(b) Primerjava Kramnika profila iz podatkov za ostale šahiste in njegovega profila iz partij za svetovnega šahovskega prvaka. Podobnost: 0.717

Slika 4.9: Analiza Vladimirja Kramnika

Zanimivo je, da Kramniku, kot najpodobnejšega svetovnega šahovskega prvaka za dano otvoritev ne najdemo njega samega, ampak nekoga drugega, in sicer svetovnega šahovskega prvaka Petrosiana. Dobljeni rezultat lahko utemeljimo z dejstvom, da je Kramnik v nedavnem intervjuju Petrosiana izpostavil kot šahista, ki je imel velik vpliv na njegov stil (vir: [18]).

Za nenavaden rezultat je lahko več razlogov. Možno je, da se stil šahista skozi čas preveč spreminja, da bi lahko uspešno ugotovili ujemanje z njim samim. Možno je tudi, da Kramnik drugače igra partije za naslov svetovnega šahovskega prvaka in ostale partije. V bolj sproščenem okolju, proti pogostokrat slabšim nasprotnikom, šahistu ni vedno potrebno igrati tako previdno kot na prvenstvih za svetovnega prvaka.

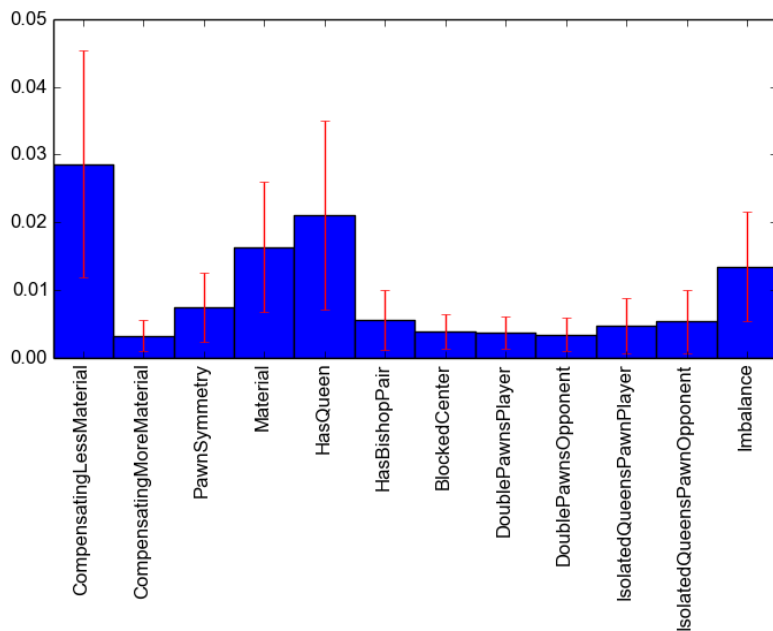
4.2 Povprečna medsebojna informacija atributov

Na sliki 4.10 je prikazano povprečje povprečne medsebojne informacije atributov za vse šahiste. Z rdečo črto je označen standardni odklon vrednosti povprečne medsebojne informacije.

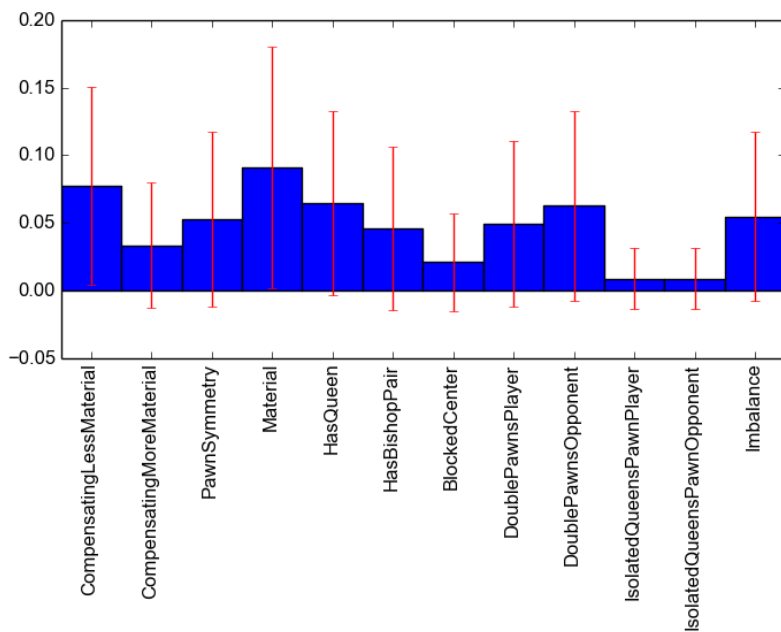
Na slikah vidimo, da so nekateri atributi bolj pomembni kot drugi. Posebej bi izpostavili attribute *CompensatingLessMaterial* - kompenzacija za pomanjkanje materiala, *Material* - količina materiala in *Imbalance* - kako neravnovesna je pozicija. Ti atributi se izkažejo za pomembne tako pri svetovnih šahovskih prvakih, kot tudi pri ostalih šahistih. Ti atributi so bili, zaradi svojega vpliva na kvaliteto odigrane partije, izbrani za osnovo za gradnjo profila C.

Na grafu za svetovne šahovske prvake lahko vidimo, da je pomemben tudi atribut *HasQueen* - prisotnost kraljice na šahovnici.

Pri ostalih igralcih lahko iz grafa kot pomembna atributa razberemo *DoublePawnsPlayer* - dvojni kmetje na šahistovi strani in *DoublePawnsOpponent* - dvojni kmetje na strani nasprotnika.



(a) Svetovni šahovski prvaki



(b) Ostali šahisti

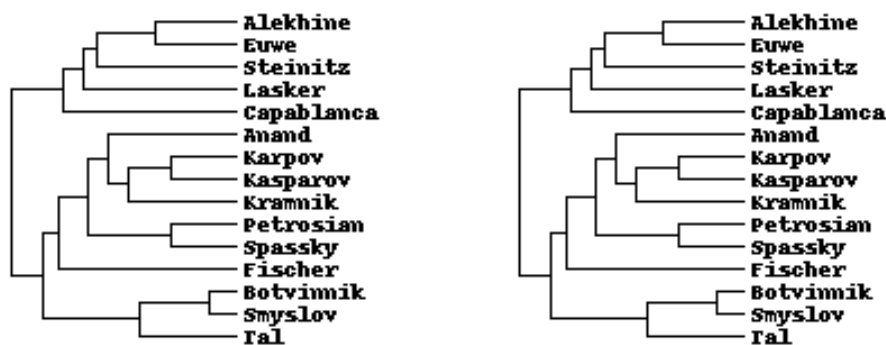
Slika 4.10: Povprečje povprečne medsebojne informacije za attribute in njen standardni odklon

4.3 Razvrščanje svetovnih šahovskih prvakov v skupine

Svetovne šahovske prvake smo razvrščali v skupine z namenom preizkusiti, ali se naš način ugotavljanja podobnosti med šahisti ujema z analizo eksperta, ali so naši profili relevantni za šahistovo analizo.

Zaradi celostne obravnave podatkov smo analizo naredili tako s pomočjo kosinusne podobnosti kot s pomočjo Pearsonovega korelacijskega koeficienta.

Pri hierarhičnem razvrščanju svetovnih šahovskih prvakov, glede na njihov stil igranja, definiran s profilom C (slika 4.11), je razlika med merama podobnosti tako majhna, da razlika na grafih ni opazna. Za računanje razdalj med skupinami smo uporabili maksimalno razdaljo.



(a) Kosinusna razdalja

(b) Pearsonov korelacijski koeficient

Slika 4.11: Hierarhično razvrščanje svetovnih šahovskih prvakov.

Pričakovali smo, da bo hierarhično razvrščanje v iste skupine razvrstilo šahiste, ki igrajo podobno, vendar se to ni zgodilo. Najpodobnejši pari igralcev, to so Alekhine in Euwe, Karpov in Kasparov, Petrosian in Spassky ter Botvinnik in Smyslov, v splošnem igrajo zelo različno (vir [15]). Imajo pa skupno značilnost. Med podatki, ki smo jih analizirali, so partije, ki so jih ti šahisti igrali drug proti drugemu. Te partije predstavljajo velik delež analiziranih potez šahista (tabela 4.2). Dobljeni rezultati sicer govori v prid metodi, je pa lahko tudi posledica izbora atributov, ki naj bi določali stil igranja.

Šahist	Nasprotnik	Delež analiziranih potez, ki jih je šahist odigral proti nasprotniku
Alekhine	Euwe	36%
Euwe	Alekhine	72%
Kasparov	Karpov	74%
Karpov	Kasparov	72%
Petrosian	Spassky	69%
Spassky	Petrosian	70%
Botvinnik	Smyslov	49%
Smyslov	Botvinnik	100%

Tabela 4.2: Delež analiziranih potez, ki so jih šahisti odigrali drug proti drugemu.

Alekhine je znan kot univerzalni igralec, ki je slovel po svojih kombinacijah, Euwe pa kot miren in taktičen šahist. Dvoboj sta odigrala leta 1935 in ponovno leta 1937 (vir [15]).

Kasparov in Karpov sta se za naslov svetovnega velemejstra pomerila leta 1984, 1985, 1986, 1987 in 1990. Kasparova se pogosto primerja z Alekhineom, znan je po energični igri, medtem ko je za Karpova značilna učinkovita pozicijska igra (vir [15]).

Petrosian je znan po pasivnem stilu, raje brani, kot napada. Spassky pa je univerzalni igralec, ki je znan po svojih direktnih napadih. Pomerila sta se leta 1966 in 1969 (vir [15]).

Botvinnik in Smyslov sta dvoboj za svetovnega šahovskega prvaka odigrala leta 1954, 1957 in leta 1958. Botvinnik je predvsem strateški igralec. Smyslov pa je znan po harmonični pozicijski igri in se, nasprotno od Botvinnika, ne posveča preveč preračunavanju (vir [15]).

Šahisti lahko med igranjem partije drug drugemu vsiljujejo svoj stil igranja. Še posebej je to značilno za svetovna šahovska prvenstva, saj imajo tam šahisti toliko več časa za pripravo na točno določenega nasprotnika.

Znan je dvboj iz leta 1961, ko je Botvinnik, bolj miren, strateški šahist, premagal Talja, ki ima izrazito divji stil. To je storil tako, da je Talju preprečil dostop do pozicij, v katerih, bi le-ta lahko pokazal in uporabil svoj stil. Tal je bil prisiljen igrati bolj mirno igro, kot jo je bil navajen (vir [15]).

Poleg tega imajo atributi, ki smo jih uporabili pri analizi, vrednost pogosto enako za oba igralca. Takšen je npr. atribut simetrična kmečka struktura.

V primeru da je vrednost nekega atributa pri neki poziciji enaka za oba šahista, je tudi vrednost tega atributa v profilu teh dveh šahistov zelo podobna. Pri svetovnih šahovskih prvakah se vrednost atributa v profilu, delež dobrih potez, ki so imele pozitivno vrednost tega atributa, zelo približa kar deležu vseh potez, ki so imele pozitivno vrednost tega atributa, saj so skoraj vse poteze dobre.

Rezultate hierarhičnega razvrščanja bi lahko izboljšali tako, da bi uporabili večjo bazo podatkov, kjer bi svetovni šahovski prvak igral z večjim številom različnih šahistov. Delež slabih potez za svetovnega šahovskega prvaka bi bil še vedno zelo majhen, bi pa zmanjšali vpliv enega šahista na stil svetovnega šahovskega prvaka.

Višanje meje za slabo in dobro potezo ni primerna rešitev, saj sta meji že sedaj nastavljeni zelo visoko. Meja za dobre poteze je 5 in za slabe 10 stotink kmeta (ang. centipawn). Pri običajnem igralcu so poteze s takšnimi napakami še vedno zelo dobre.

Rezultati razvrščanja svetovnih šahovskih prvakov v skupine z metodo voditeljev so dali podobne rezultate. Tudi ta metoda pogosto v isto skupino uvrsti šahiste, ki so partije odigrali drug proti drugemu. Rezultati razvrščanja v skupine z metodo voditeljev so prikazani v tabeli 4.3. Pri tem so skupine za različni meri podobnosti razporejene tako, da je razvidna podobnost med rezultatoma.

Pri razvrščanju z metodo voditeljev smo igralce razvrščali v 5 skupin. Da se ne bi ujeli v lokalni minimum, smo razvrščanje ponovili stokrat, pri čemer smo vsakič naključno določili začetne voditelje.

Skupina	Kosinusna razdalja	Pearsonov korelacijski koeficient
Skupina 1	Capablanca Steinitz	Lasker Steinitz
Skupina 2	Alekhine Euwe Lasker	Alekhine Euwe
Skupina 3	Fischer Spassky	Capablanca
Skupina 4	Anand Karpov Kasparov Kramnik Petrosian	Fischer Spassky Anand Karpov Kasparov Kramnik Petrosian
Skupina 5	Botvinnik Smyslov Tal	Botvinnik Smyslov Tal

Tabela 4.3: Rezultati razvrščanja svetovnih šahovskih prvakov v skupine z metodo voditeljev

Poglavje 5

Zaključek

Šah je igra, ki jo ljudje igrajo že stoletja. Veliko šahistov je posvetilo mnogo ur izboljšavi svoje igre. Trudijo se izboljšati tako kvaliteto kot tudi stil svoje igre. Namen tega dela je olajšanje analize, potrebne za izboljšavo igre.

Cilj dela je bil ugotoviti stil igralca. Stil je bilo potrebno definirati na način, ki ga je možno analizirati z računalnikom in hkrati ohraniti neko mero razumljivosti za človeka. Stile smo želeli analizirati in z njihovo pomočjo ugotavljati značilnosti igre šahistov. Iskali smo tudi podobne šahiste.

Na osnovi rezultatov bi lahko šahistu svetovali dvoje. V kakšnih pozicijah dela več ali manj napak in kdo so šahisti, ki igrajo podobno kot on. Torej šahiste s primerljivim stilom igre. Najdeni po stilu podobni šahisti pa bi npr. igrali bolje od izbranega šahista. S pregledom partij boljših, a po stilu igranja sorodnih šahistov bi šahist lažje izboljšal svojo igro. Šahistom bi lahko s tovrstno analizo olajšali delo, saj jim ne bi bilo več potrebno "ročno" iskati po stilu igranja podobnih šahistov, katerih partije bi bilo koristno analizirati za izboljšavo lastne igre.

Osnovo analize so nam predstavljali atributi, katerih vrednosti smo določali za vsako potezo posebej. Uporabili smo nekaj preprostih atributov, kot sta prisotnost kraljice in prisotnost lovskega para, ter nekaj bolj kompleksnih atributov, ki opisujejo kmečke strukture, prisotne na šahovnici. Takšni atributi so blokiran center, dvojni kmetje in osamljen damin kmet. Poleg

tega smo uporabili tudi nekaj atributov, ki so zajemali dogajanje na celotni šahovnici in ne le del figur. Takšni atributi so kompenzacija za pomanjkanje materiala, neravnovesje in količina materiala.

Z uporabo atributov in tudi brez nje smo sestavili 5 različnih profilov, ki definirajo stil igranja šahista. Te profile smo preizkusili z iskanjem istih šahistov v dveh različnih bazah potez, ki smo ju sestavili iz podatkovne baze, ki je vsebovala podatke o partijah različno rangiranih šahistov. Pri tem se je za najuspešnejšega izkazal profil E, ki je dosegel 29% točnost. Ker ta profil ne uporablja atributov, pa smo za nadaljno analizo uporabili profil C, ki je osnovan na atributih, ki zajemajo dogajanje na celotni šahovnici. Tudi ta profil je bil podobno uspešen kot profil E, z doseženo točnostjo 27% pri uporabi kosinusne podobnosti. Ugotovili smo, da se točnost izboljša z večjim številom potez za šahista. Pri sestavljanju profilov smo si pomagali z računanjem povprečne medsebojne informacije atributov in njene standardne deviacije.

Na podlagi izbranega profila, tj. profila C, ki predvideva drugačen stil igre za drugo otvoritev, sam stil pa definira z atributi, smo šahistom iskali podobne svetovne šahovske prvake. Ker smo predpostavljali, da igralec pri različnih otvoritvah igra z različnimi stili, smo mu iskali svetovne prvake, ki pri določeni otvoritvi igrajo podobno kot on. Temu primerno smo prilagodili iskanje podobnih šahistov in profil. Pri tej analizi smo prišli do nekaterih zanimivih rezultatov, kot je npr. ugotovljena podobnost svetovnega prvaka Kramnika s svetovnim prvacom Petrosianom, ki jo tudi Kramnik sam potrjuje.

Iskali smo tudi podobne svetovne šahovske prvake. To smo delali z metodama za razvrščanje v skupine. Preizkusili smo hierarhično razvrščanje in metodo voditeljev. Pri tej analizi smo dobili zanimiv rezultat, saj nismo našli podobnih šahistov, ampak šahiste, ki so partije odigrali drug proti drugemu. Iz tega smo sklepali, da bi bili za takšno analizo bolj primerni bolj raznoliki podatki. Pri tej slednji raziskavi smo namreč uporabili samo podatke iz dvobojev za svetovnega šahovskega prvaka. V tej bazi ima vsak šahist

malo različnih nasprotnikov, saj so v omenjenih dvobojih igrali le svetovni prvaki in njihovi izzivalci. Bolj primerna bi bila baza s partijami, ki jih je svetovni šahovski prvak odigral proti več različnim nasprotnikom, saj lahko šahista med igranjem partije drug drugemu vsiljujeta svoj stil (vir [15]). Ker se to ne zgodi vedno, bi bolj raznolika baza omilila vpliv vsiljevanja stila. Za uporabo podatkovne baze, ki je vsebovala samo partije s svetovnih šahovskih prvenstev, smo se odločili, ker se partije odigrane na svetovnih šahovskih prvenstvih preveč razlikujejo od večine ostalih partij. Pred svetovnim šahovskim prvenstvom se vsak šahist dalj časa pripravlja na znanega nasprotnika in napake, ki jih šahisti delajo so manjše. Če bi uporabili bolj raznoliko bazo, bi verjetno morali podatke o partijah odigranih na svetovnih šahovskih prvenstvih obravnavati ločeno od ostalih partij.

Rezultate dela bi lahko izboljšali tudi tako, da bi zasnovali način preverjanja točnosti rezultatov. Tako bi lahko ugotovili, ali sta si dva šahista, za katera na podlagi rezultatov trdimo, da sta si podobna, tudi v resnici podobna, oziroma ali sta si podobna na enak način kot ga vidi šahist. Trenutno baza znanih podobnih šahistov ne obstaja. Sestavljanje takšne baze pa je težavno, saj jo lahko sestavljajo le šahovski eksperti, ki za analizo enega šahista običajno potrebujejo veliko časa. Problem je tudi, da je njihovo mnenje subjektivno, in zato bi do zanesljivih podatkov lahko prišli le s konsenzom več šahovskih ekspertov.

Rezultate bi bilo lahko izboljšali tudi z uporabo več atributov. Pri tem moramo biti previdni, če želimo, da je profil igralca dokaj preprosto razumljiv tudi človeku. Zato ne smemo dodati preveč atributov, dodani atributi pa morajo biti razumljivi človeku.

Poleg analize stilov na podlagi potez bi se lahko lotili tudi analize stilov na podlagi partij. Pri tem bi lahko poiskali značilnosti partij in šahiste analizirali na podlagi atributov partij ali pa bi iskali podobne partije. Pri iskanju podobnih partij je nekoliko olajšan problem ugotavljanja pravilnosti naših rezultatov, saj vemo, da med partijami z isto otvoritvijo obstaja podobnost.

Kot glavni prispevek dela bi izpostavili predvsem način sestavljanja profi-

lov za analizo šahistov, saj je takšen profil možno analizirati z računalnikom, obenem pa tudi šahist sam s pregledom profila dobi veliko informacij o stilu igre. Prav tako smo uspešno iskali podobne šahiste. Menimo tudi, da bi podoben način analize, z gradnjo atributnega profila, lahko aplicirali tudi na druge domene.

Literatura

- [1] D. Jannach, M. Zanker, A. Felfernig, G. Friedrich. *Recommender Systems: An Introduction*. Cambridge University Press, 2011.
- [2] F. Murtagh, “A survey of recent advances in hierarchical clustering algorithms.”, *The Computer Journal*, št. 26, zv. 4, str. 354–359, 1983.
- [3] A. Dobnikar. *Teorija informacij in sistemov*. Založba FE in FRI, 2009.
- [4] D. Kavka. *Matematika v srednji šoli: zbirka nalog*. Modrijan, 2007.
- [5] U. Krevs. *Računalniška analiza človekovih napak pri igranju šaha*. Diplomsko delo, univerzitetni študijski program prve stopnje računalništvo in informatika, 2012.
- [6] S. J. Edwards. *Portable Game Notation Specification and Implementation Guide*. 1994.
- [7] A. Matanović. *Šahovski informator: Classification of Chess Openings*. Centar za unapredivanje šaha, 1971.
- [8] A. E. Elo. *The Rating of Chess Players, Past and Present*. Ishi Press, 2008.
- [9] J. MacQueen, “Some methods for classification and analysis of multivariate observations.”, *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, št. 14, zv. 1, str. 281–297, 1967.

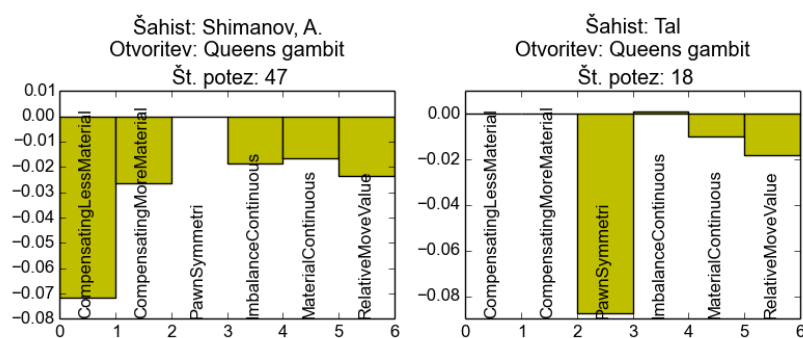
- [10] K. Wagstaff, C. Cardie, S. Rogers, S. Schrödl, “Constrained k-means clustering with background knowledge.”, *Proceedings of the Eighteenth International Conference on Machine Learning*, št. 18 zv. 1, str. 577–584, 2001.
- [11] M. Levene, T. Fenner, “A Methodology for Learning Players’ Styles from Game Records”, *IJAISC*, zv. 2, št. 4, str. 272-286, 2011.
- [12] M. Guid, I. Bratko, “Computer Analysis of World Chess Champions”, *ICGA Journal*, zv. 29, št. 2, str. 65-73, 2006.
- [13] M. Guid, M. Možina, J. Krivec, A. Sadikov, I. Bratko, “Learning Positional Features for Annotating Chess Games: A Case Study.”, *CG 2008, Beijing, China, September/October, 2008. Computers and Games. Lecture Notes in Computer Science*, št. 5131, str. 192-204, 2008. Springer.
- [14] M. Guid, I. Bratko, “Using Heuristic-Search Based Engines for Estimating Human Skill at Chess.” *ICGA Journal*, zv. 34, št. 2, str. 71-81, 2011.
- [15] G. Kasparov, *Moji veliki predhodniki*, knj. 1–5. Šahohlačnik - zavod za napredek šaha, 2004.
- [16] J. Demšar, T. Curk, A. Erjavec, Č. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik, B. Zupan, “Orange: Data Mining Toolbox in Python”, *Journal of Machine Learning Research*, št. 14, str. 2349–2353, 2013.
- [17] R. Houdart, “Houdini Chess Engine”. Dostopno na: <http://www.cruxis.com/chess/houdini.htm>. Dostopano: 15. 2. 2015.
- [18] E. Surov, “Vladimir Kramnik: ”Petrosian’s chess legacy helped me a lot in my match vs Kasparov.“”, *Chess-news*, 6. 11. 2014. Dostopno na: <http://chess-news.ru/en/node/17187>. Dostopano: 1. 3. 2015.

- [19] "Computer Chess Rating List". Dostopno na: <http://www.computerchess.org.uk/ccr1/>. Dostopano: 10. 3. 2015.

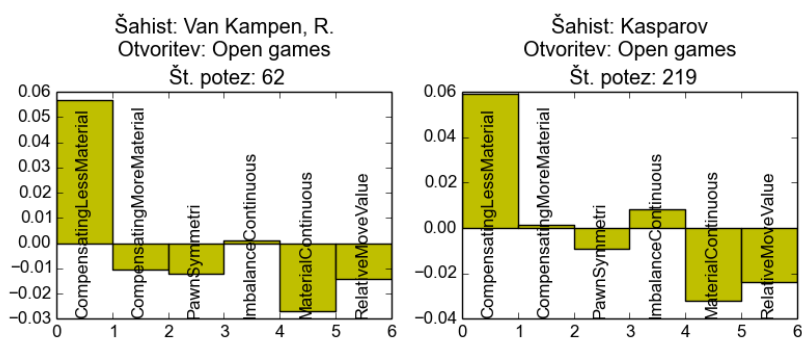
Dodatek A

Primerjava igralcev s svetovnimi šahovskimi prvaki

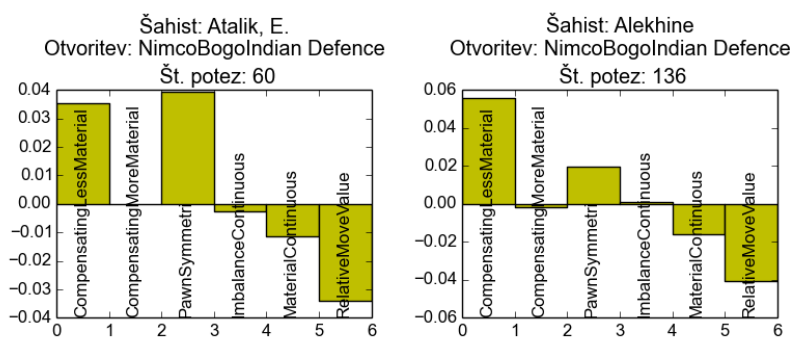
V tem poglavju so predstavljeni celotni rezultati iskanja podobnih svetovnih šahovskih prvakov igralcem, za določeno otvoritev. Analizo smo izvajali za vse šahiste, za katere smo imeli v podatkovni bazi na voljo več kot 400 potez, tj. 101 šahist.



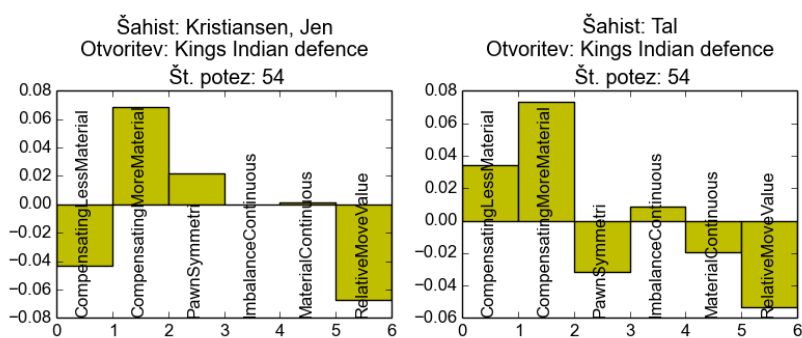
Slika A.1: Shimanov, A., Tal. Podobnost: 0.077



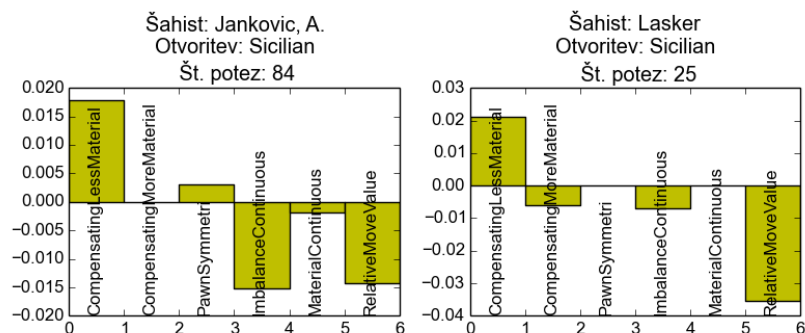
Slika A.2: Van Kampen, R., Kasparov. Podobnost: 0.971



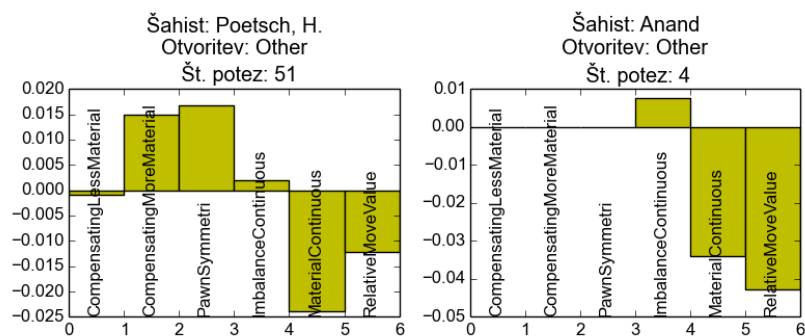
Slika A.3: Atalik, E., Alekhine. Podobnost: 0.914



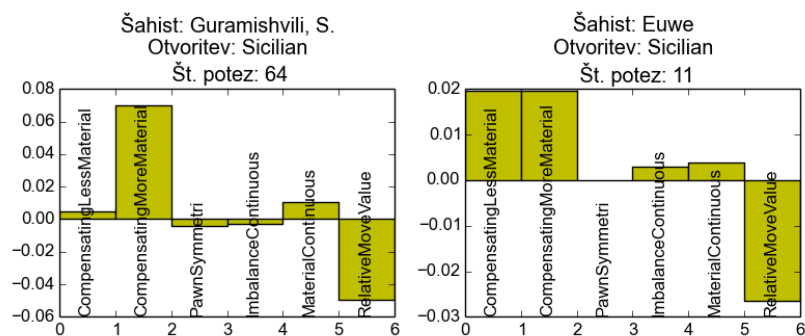
Slika A.4: Kristiansen, Jen, Tal. Podobnost: 0.572



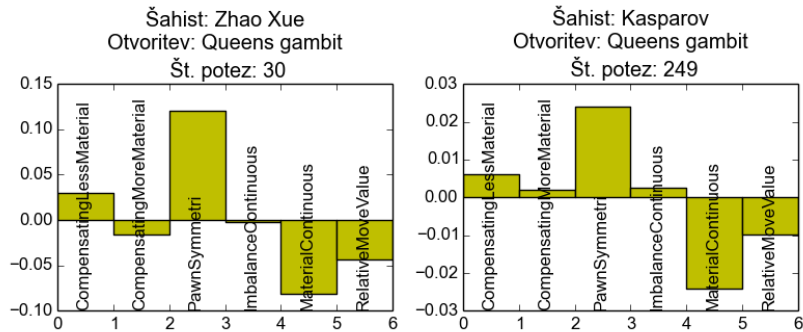
Slika A.5: Jankovic, A., Lasker. Podobnost: 0.848



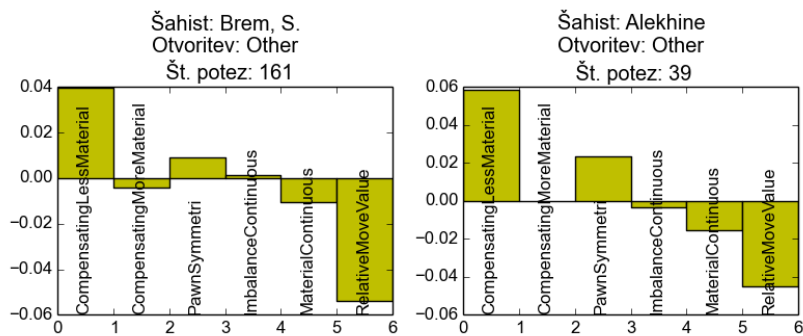
Slika A.6: Poetsch, H., Anand. Podobnost: 0.696



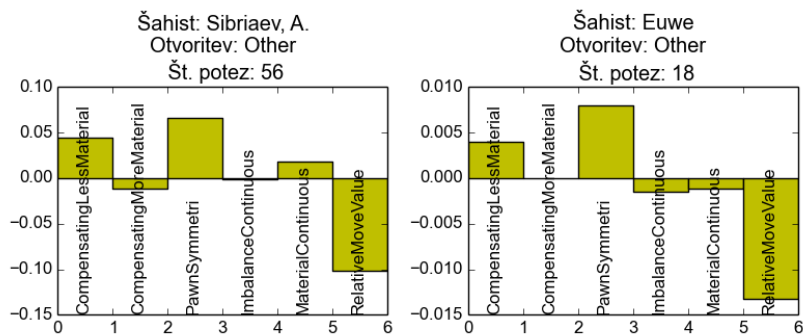
Slika A.7: Guramishvili, S., Euwe. Podobnost: 0.838



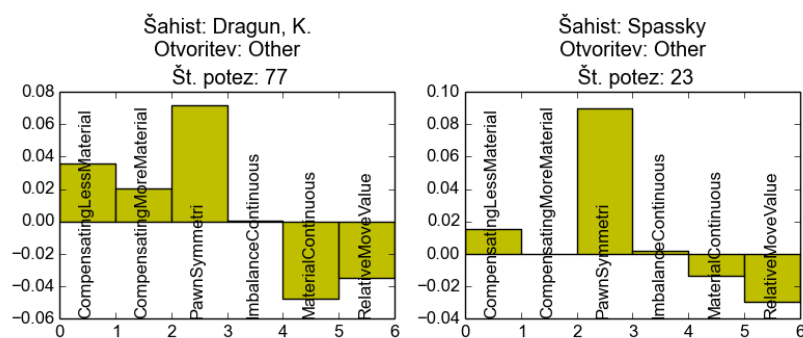
Slika A.8: Zhao Xue, Kasparov. Podobnost: 0.967



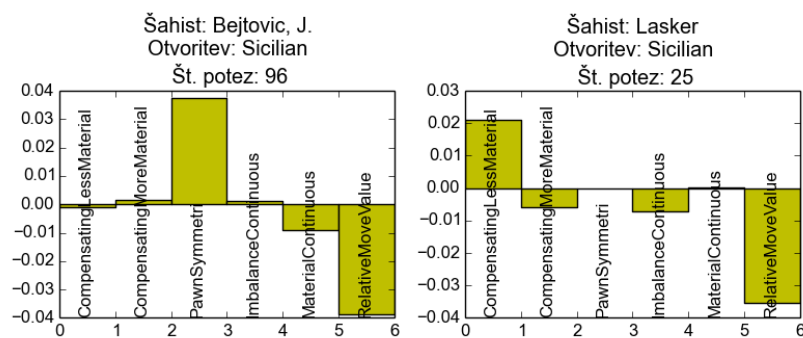
Slika A.9: Brem, S., Alekhine. Podobnost: 0.945



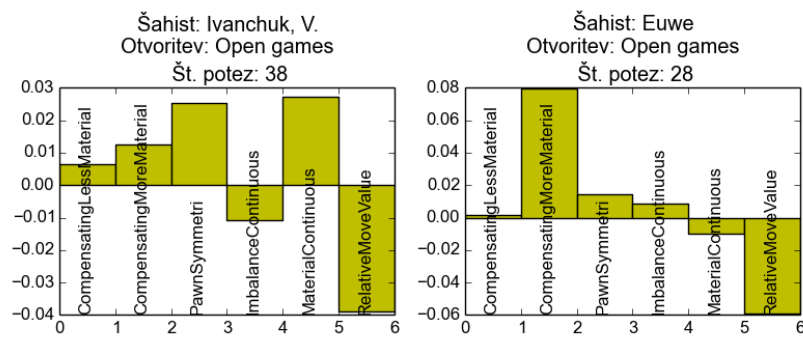
Slika A.10: Sibiriaev, A., Euwe. Podobnost: 0.965



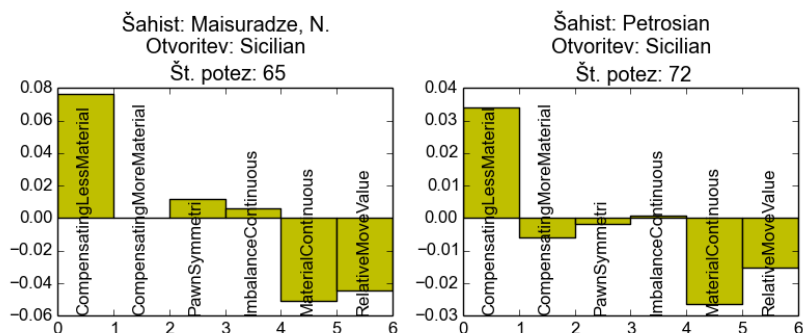
Slika A.11: Dragun, K., Spassky. Podobnost: 0.881



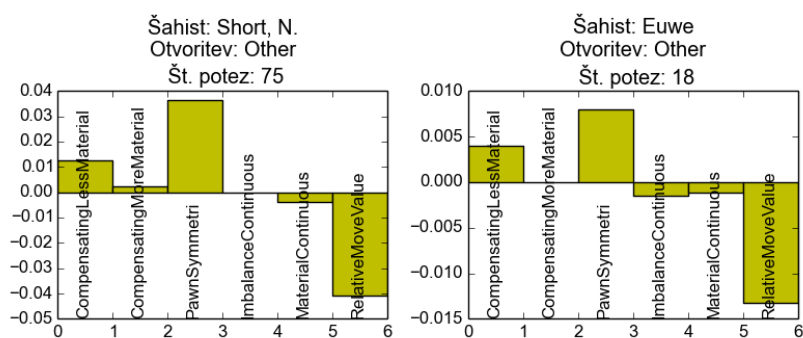
Slika A.12: Bejtovic, J., Lasker. Podobnost: 0.577



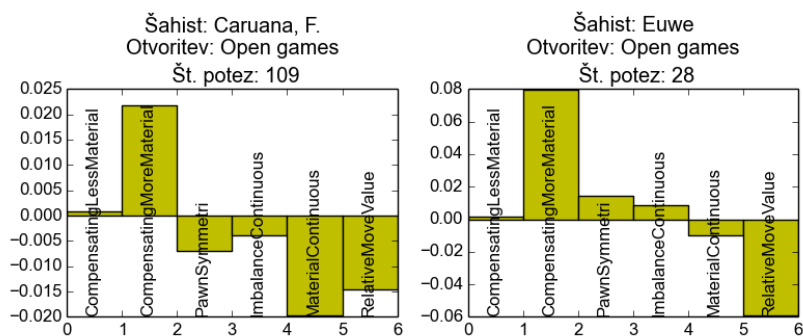
Slika A.13: Ivanchuk, V., Euwe. Podobnost: 0.580



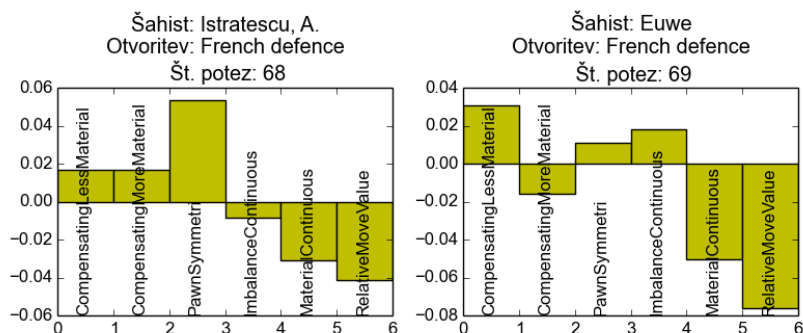
Slika A.14: Maisuradze, N., Petrosian. Podobnost: 0.971



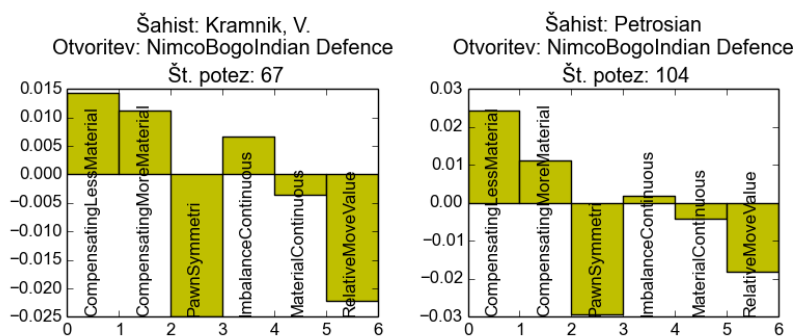
Slika A.15: Short, N., Euwe. Podobnost: 0.978



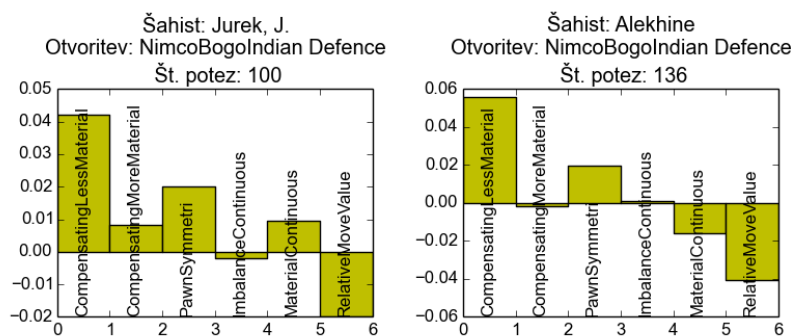
Slika A.16: Caruana, F., Euwe. Podobnost: 0.777



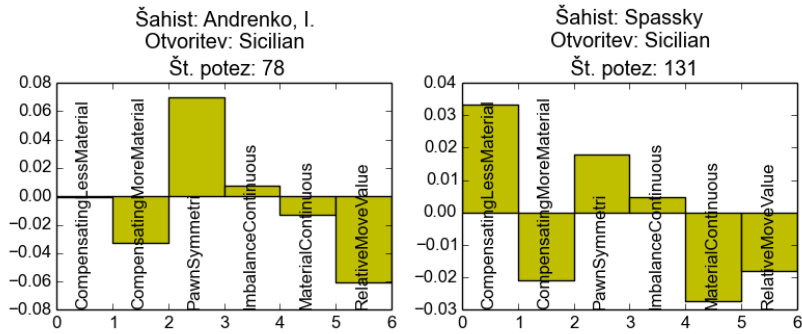
Slika A.17: Istratescu, A., Euwe. Podobnost: 0.685



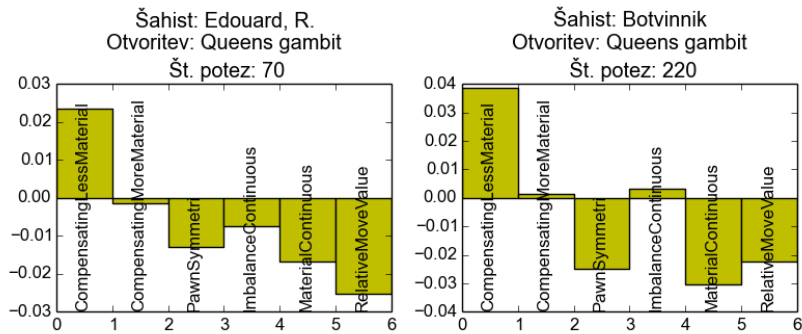
Slika A.18: Kramnik, V., Petrosian. Podobnost: 0.961



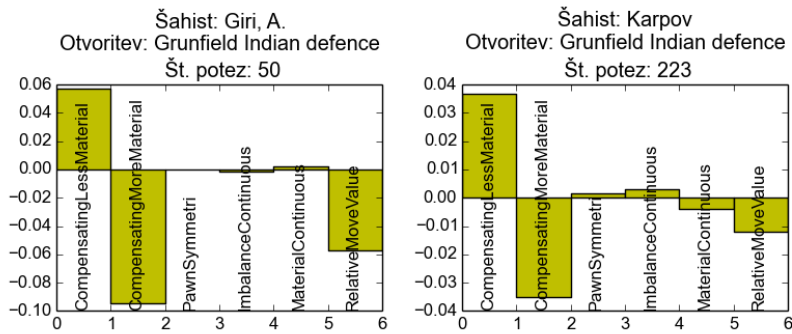
Slika A.19: Jurek, J., Alekhine. Podobnost: 0.878



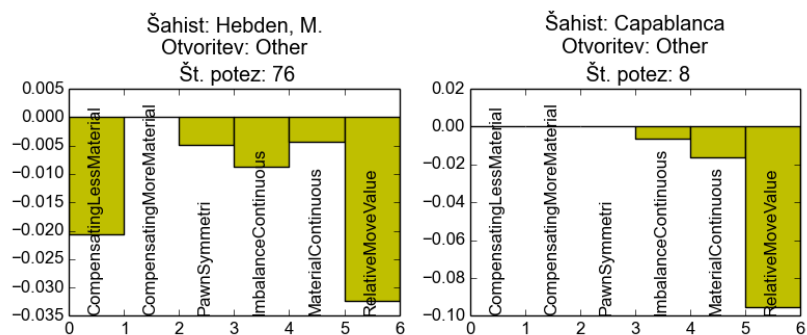
Slika A.20: Andrejko, I., Spassky. Podobnost: 0.631



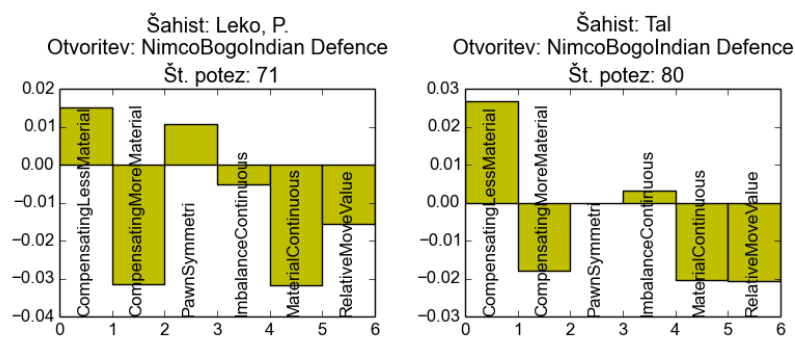
Slika A.21: Edouard, R., Botvinnik. Podobnost: 0.927



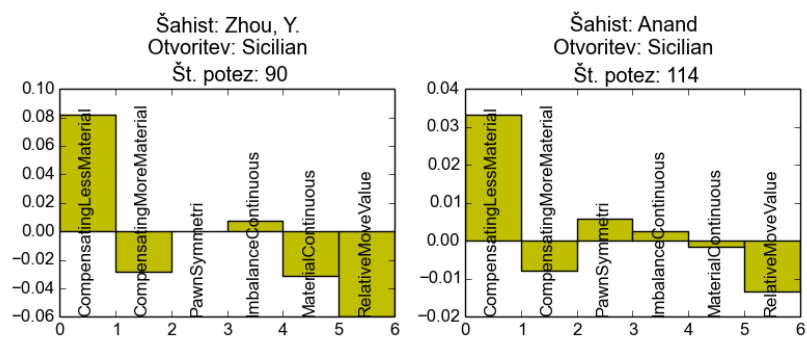
Slika A.22: Giri, A., Karpov. Podobnost: 0.934



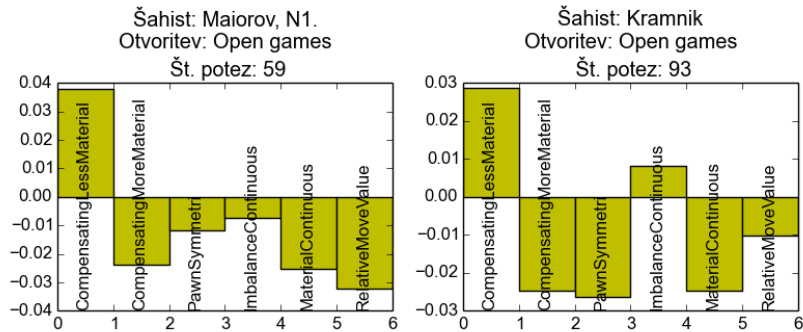
Slika A.23: Hebden, M., Capablanca. Podobnost: 0.830



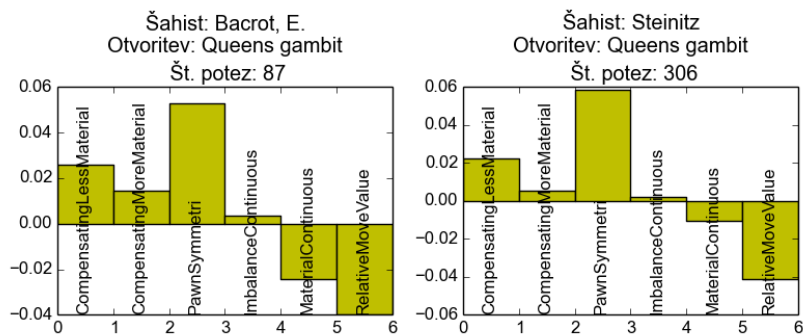
Slika A.24: Leko, P., Tal. Podobnost: 0.864



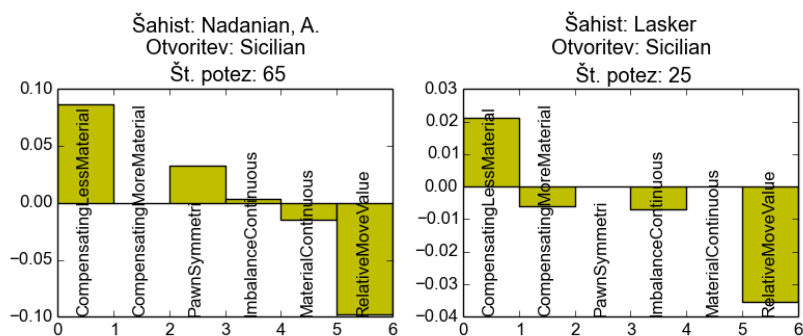
Slika A.25: Zhou, Y., Anand. Podobnost: 0.928



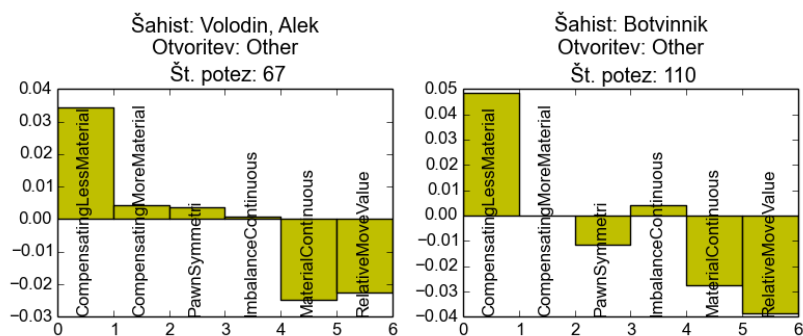
Slika A.26: Maiorov, N1., Kramnik. Podobnost: 0.857



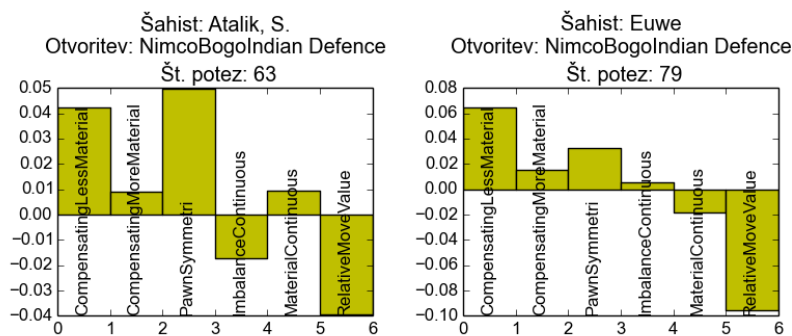
Slika A.27: Bacrot, E., Steinitz. Podobnost: 0.971



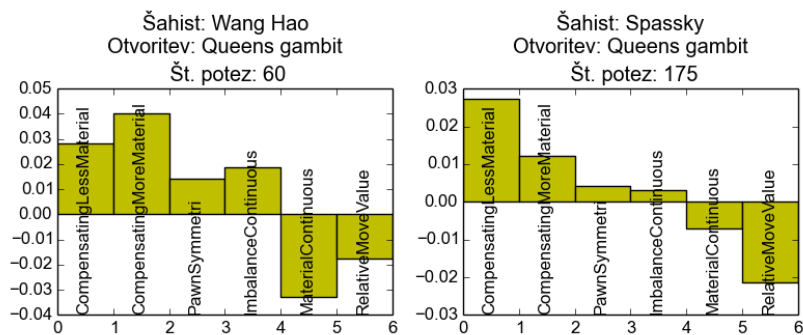
Slika A.28: Nadeian, A., Lasker. Podobnost: 0.921



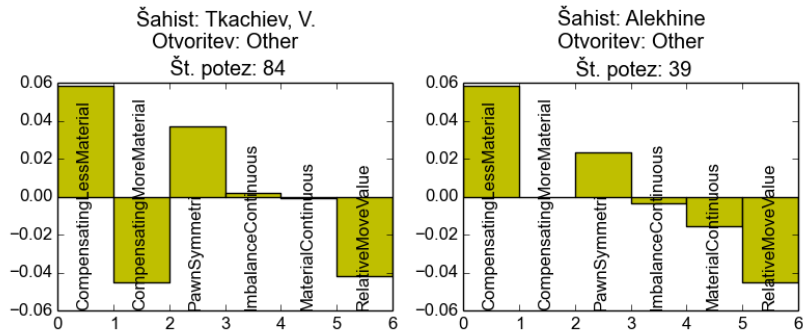
Slika A.29: Volodin, Alek, Botvinnik. Podobnost: 0.953



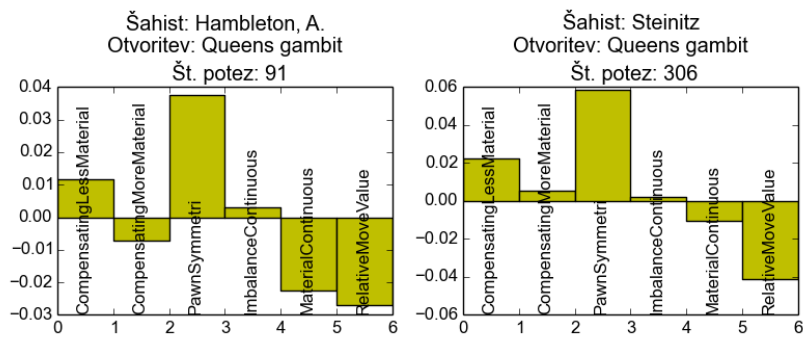
Slika A.30: Atalik, S., Euwe. Podobnost: 0.824



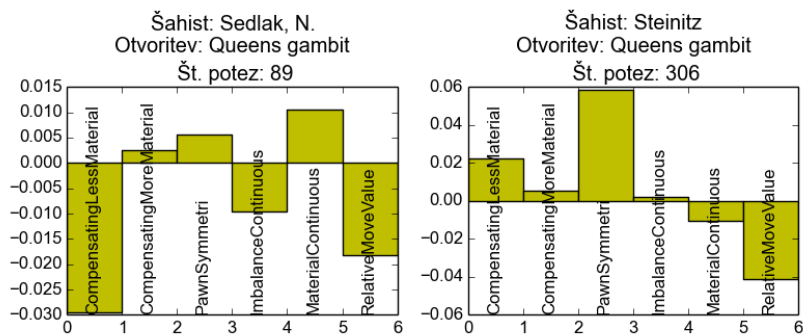
Slika A.31: Wang Hao, Spassky. Podobnost: 0.794



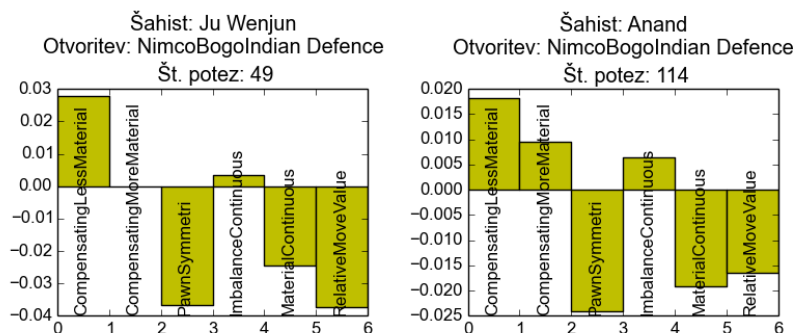
Slika A.32: Tkachiev, V., Alekhine. Podobnost: 0.842



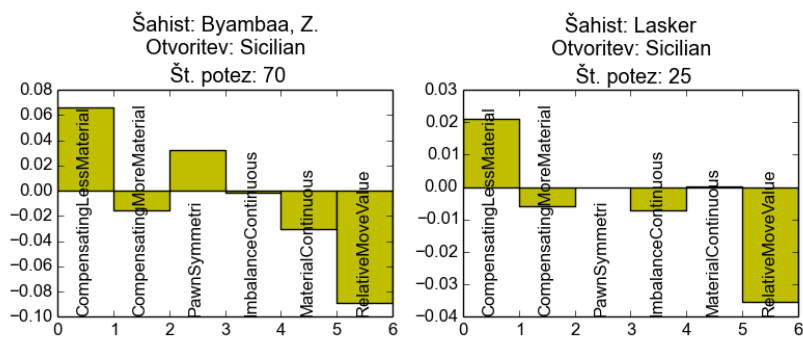
Slika A.33: Hambleton, A., Steinitz. Podobnost: 0.930



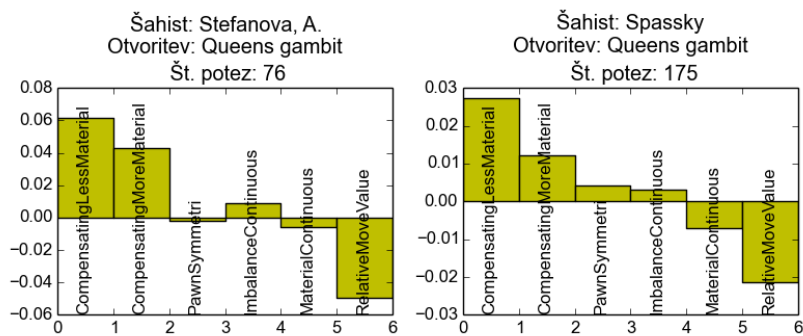
Slika A.34: Sedlak, N., Steinitz. Podobnost: 0.106



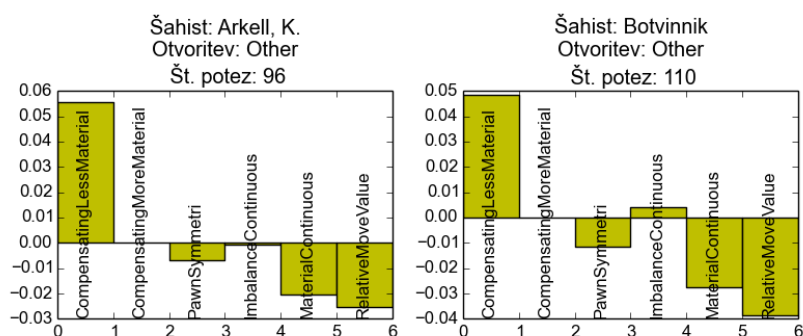
Slika A.35: Ju Wenjun, Anand. Podobnost: 0.948



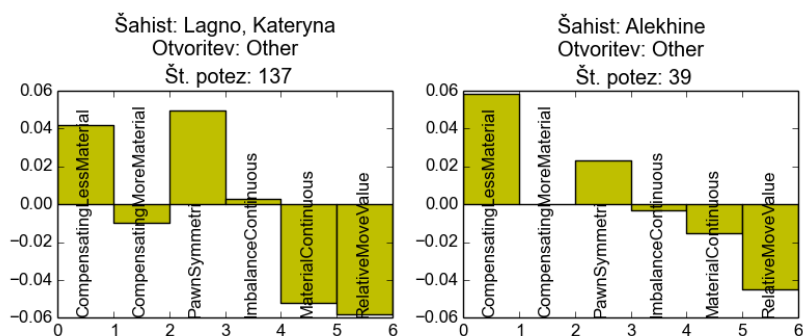
Slika A.36: Byambaa, Z., Lasker. Podobnost: 0.913



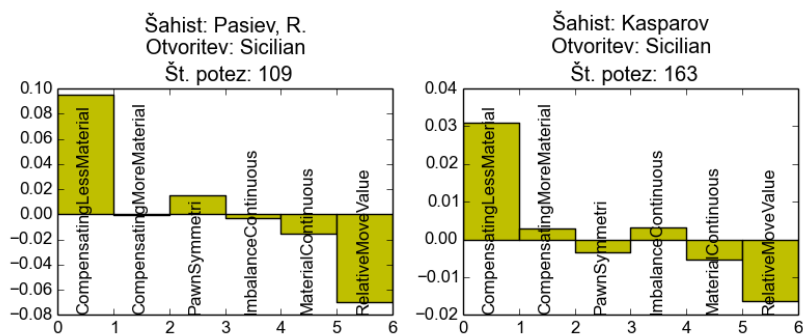
Slika A.37: Stefanova, A., Spassky. Podobnost: 0.970



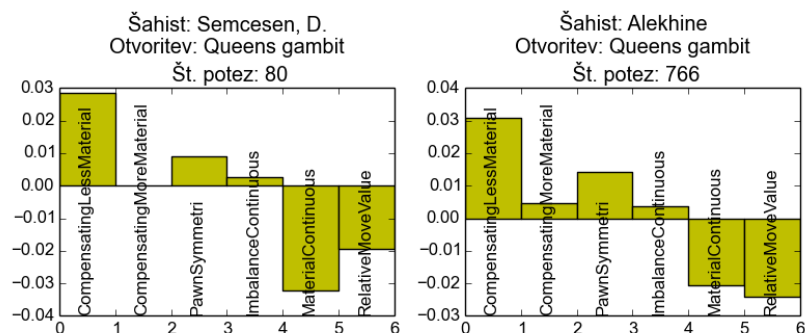
Slika A.38: Arkell, K., Botvinnik. Podobnost: 0.966



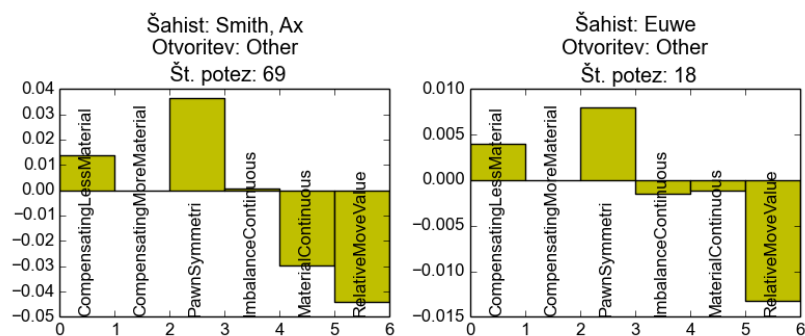
Slika A.39: Lagno, Kateryna, Alekhine. Podobnost: 0.870



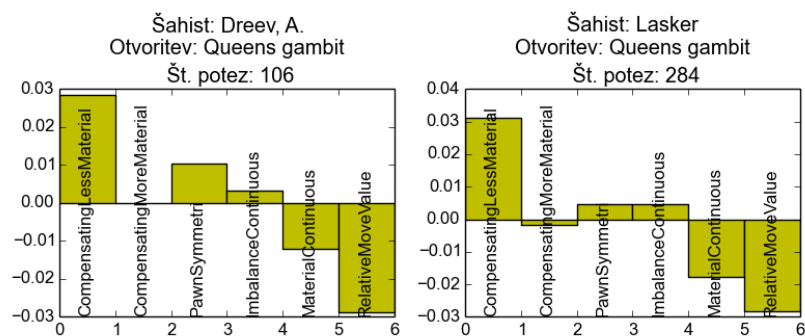
Slika A.40: Pasiev, R., Kasparov. Podobnost: 0.955



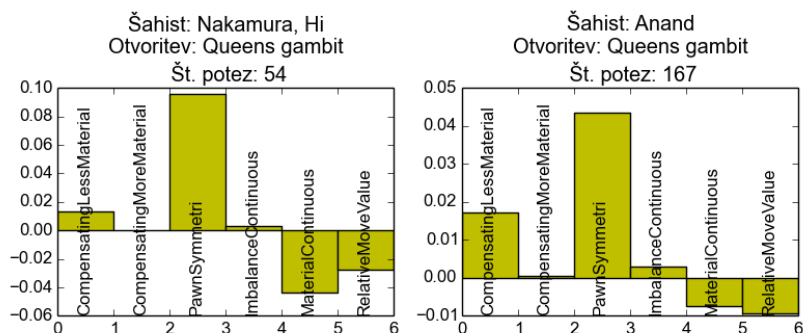
Slika A.41: Semcesen, D., Alekhine. Podobnost: 0.954



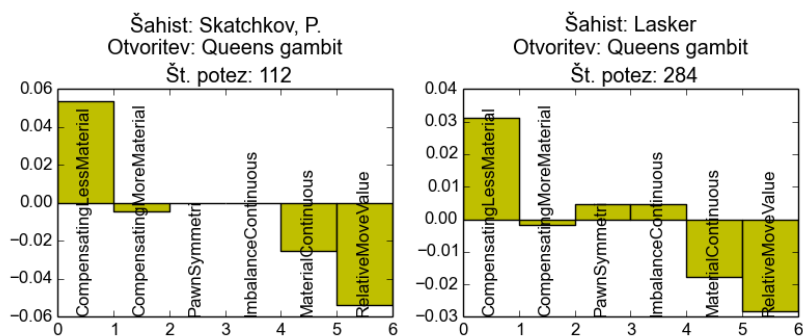
Slika A.42: Smith, Ax, Euwe. Podobnost: 0.908



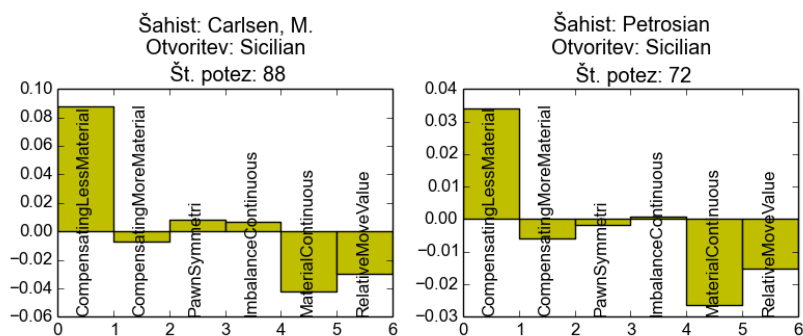
Slika A.43: Dreev, A., Lasker. Podobnost: 0.982



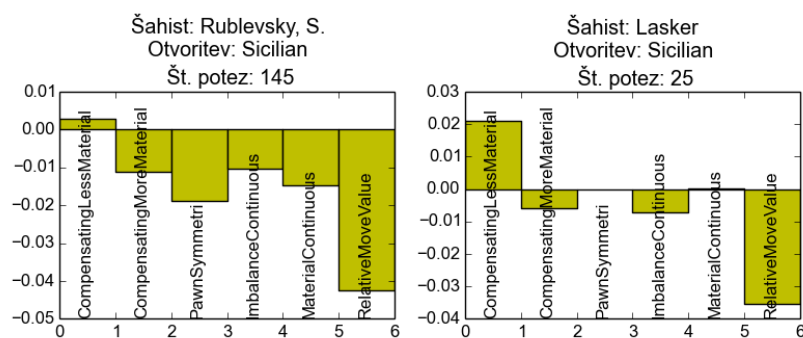
Slika A.44: Nakamura, Hi, Anand. Podobnost: 0.939



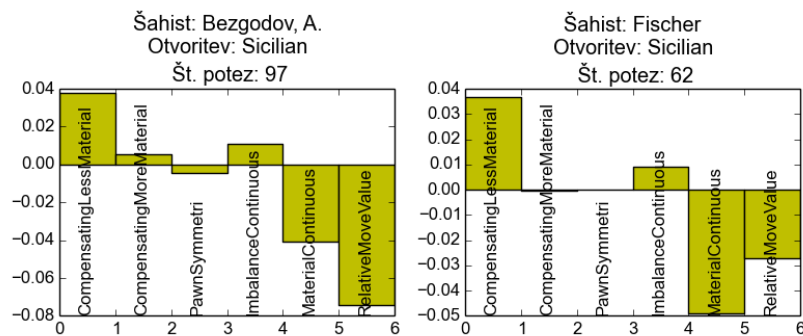
Slika A.45: Skatchkov, P., Lasker. Podobnost: 0.984



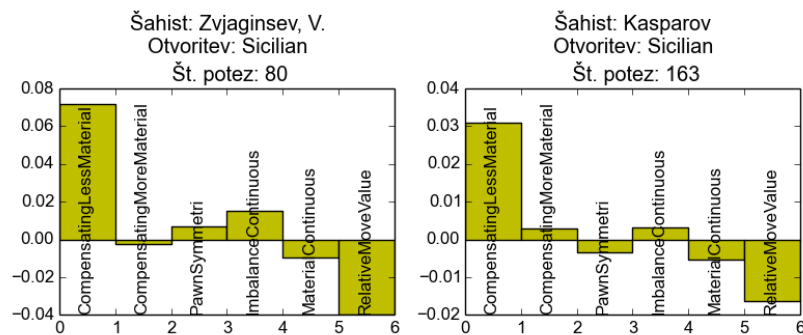
Slika A.46: Carlsen, M., Petrosian. Podobnost: 0.970



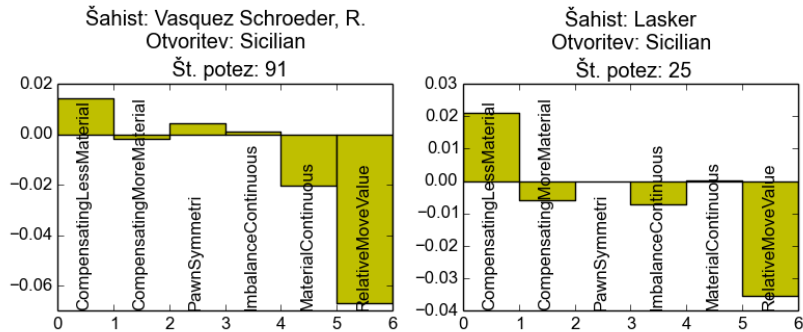
Slika A.47: Rublevsky, S., Lasker. Podobnost: 0.785



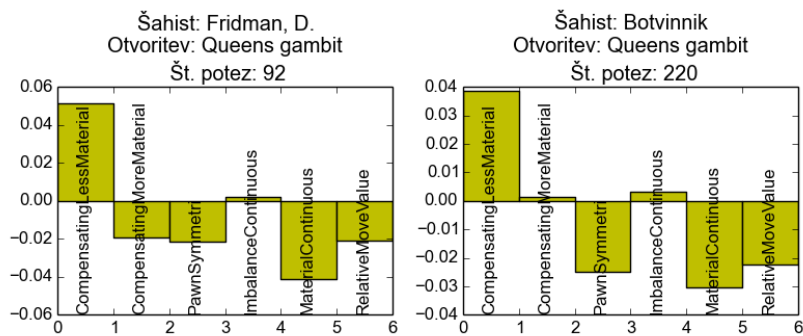
Slika A.48: Bezgodov, A., Fischer. Podobnost: 0.868



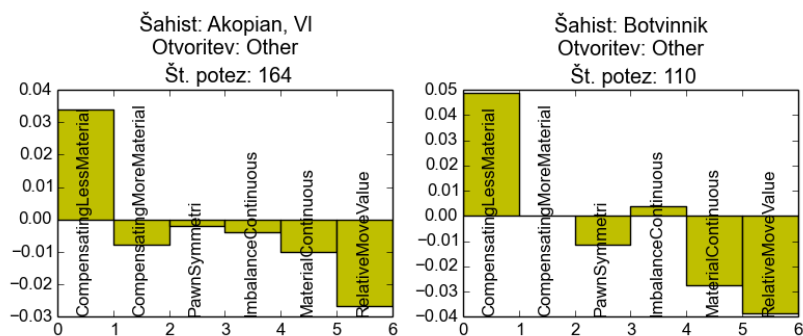
Slika A.49: Zvjaginsev, V., Kasparov. Podobnost: 0.973



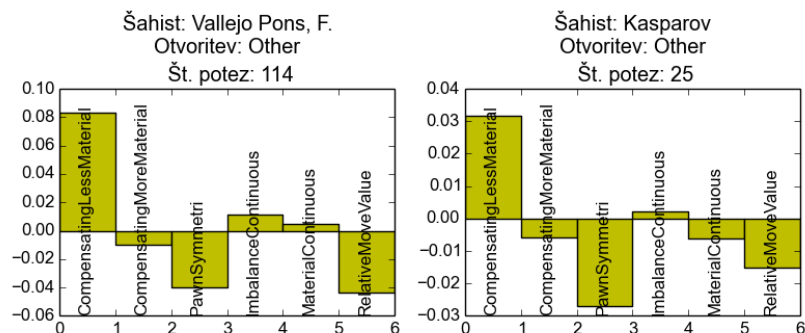
Slika A.50: Vasquez Schroeder, R., Lasker. Podobnost: 0.882



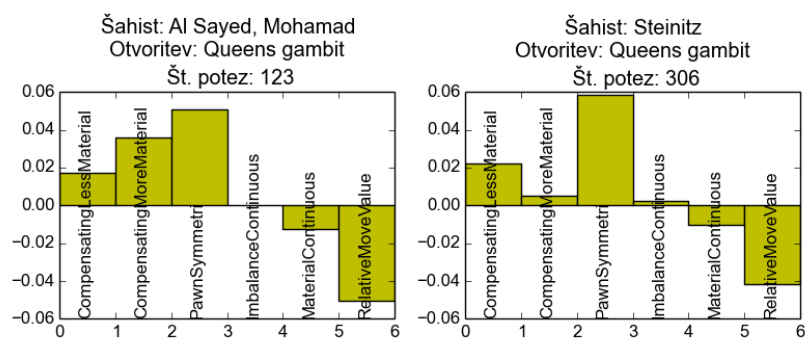
Slika A.51: Fridman, D., Botvinnik. Podobnost: 0.944



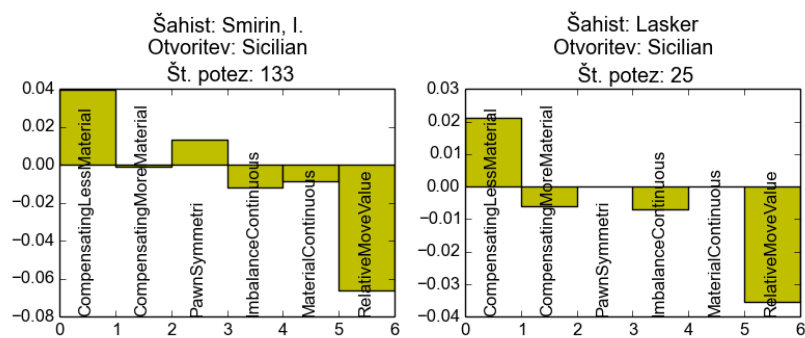
Slika A.52: Akopian, VI, Botvinnik. Podobnost: 0.948



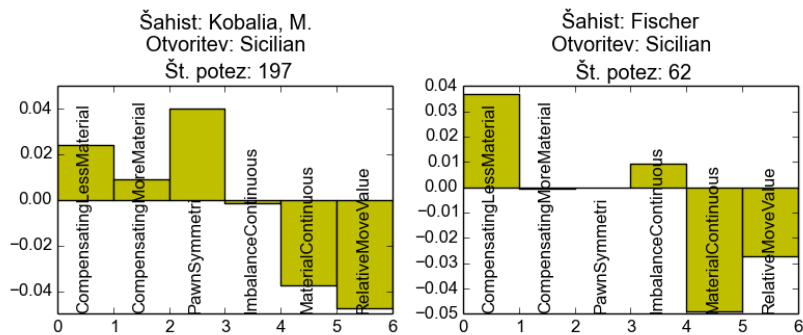
Slika A.53: Vallejo Pons, F., Kasparov. Podobnost: 0.950



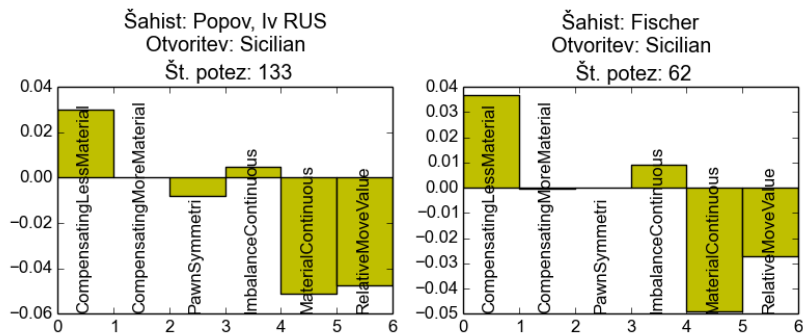
Slika A.54: Al Sayed, Mohamad, Steinitz. Podobnost: 0.915



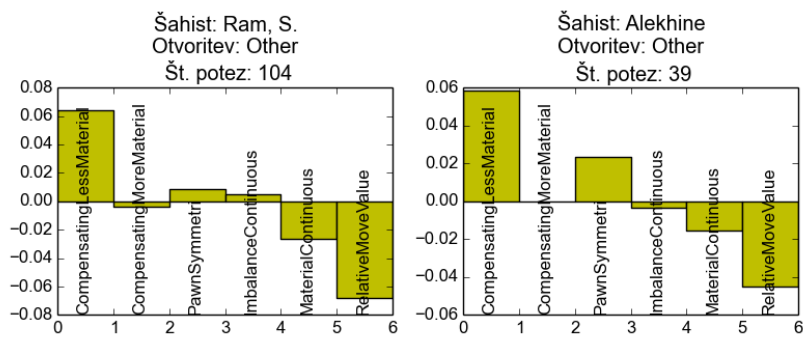
Slika A.55: Smirin, I., Lasker. Podobnost: 0.971



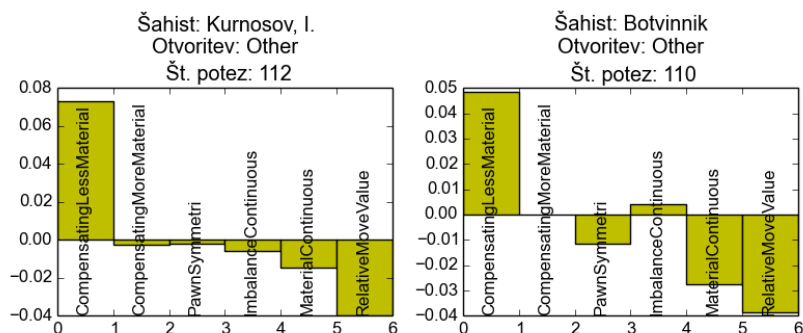
Slika A.56: Kobalia, M., Fischer. Podobnost: 0.768



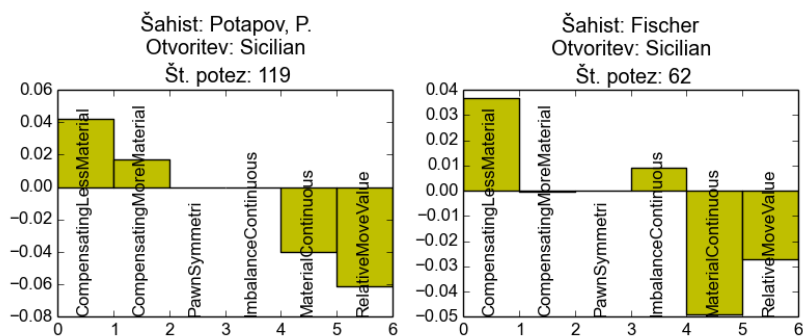
Slika A.57: Popov, Iv RUS, Fischer. Podobnost: 0.954



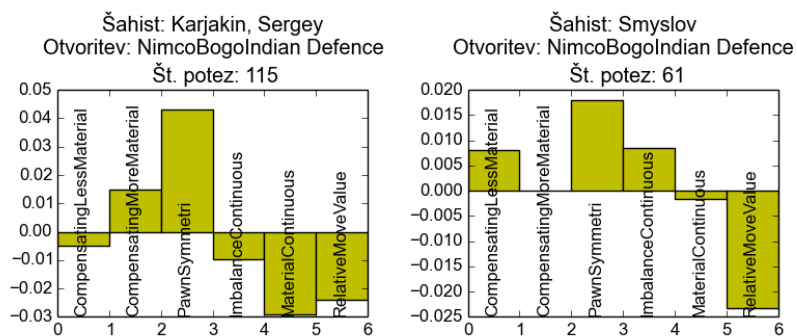
Slika A.58: Ram, S., Alekhine. Podobnost: 0.958



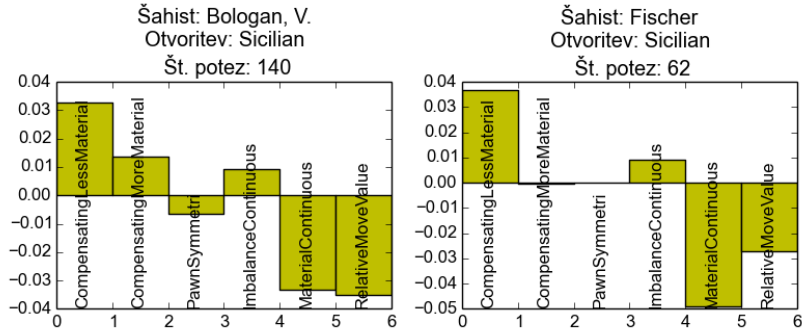
Slika A.59: Kursov, I., Botvinnik. Podobnost: 0.939



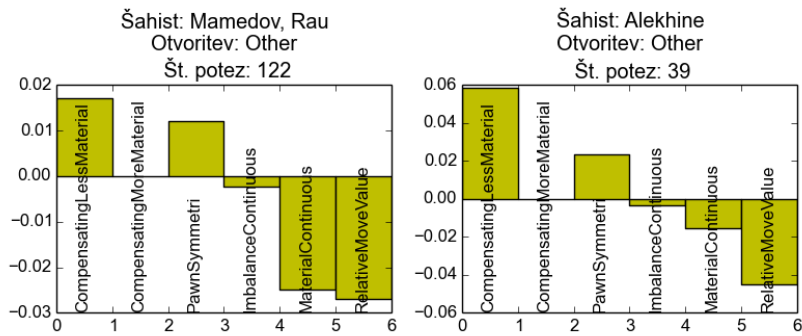
Slika A.60: Potapov, P., Fischer. Podobnost: 0.888



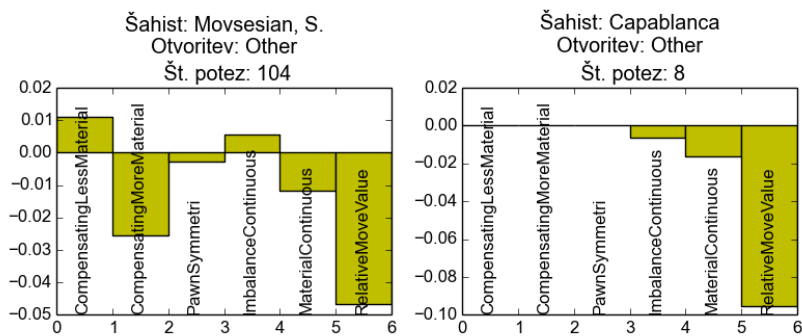
Slika A.61: Karjakin, Sergey, Smyslov. Podobnost: 0.657



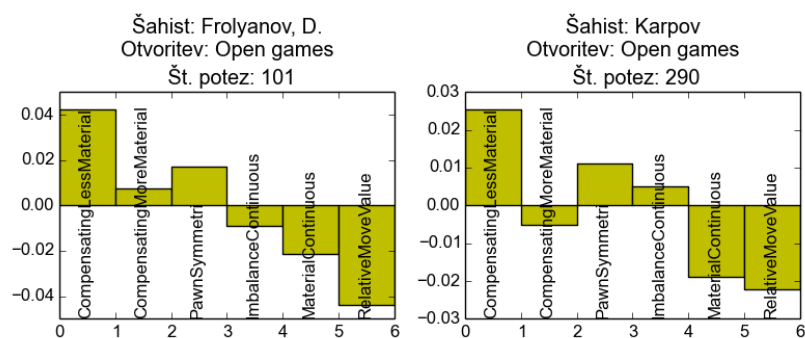
Slika A.62: Bologan, V., Fischer. Podobnost: 0.937



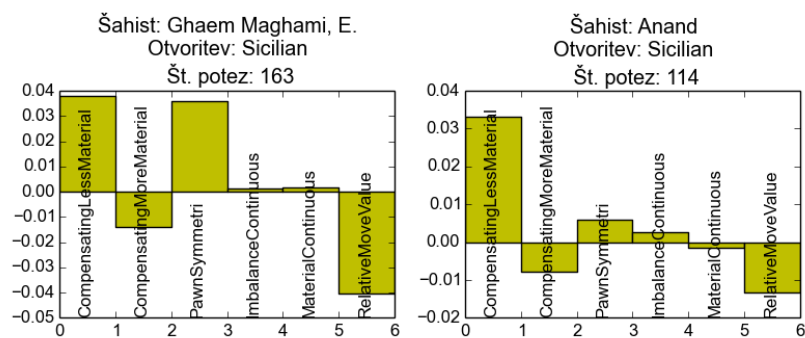
Slika A.63: Mamedov, Rau, Alekhine. Podobnost: 0.861



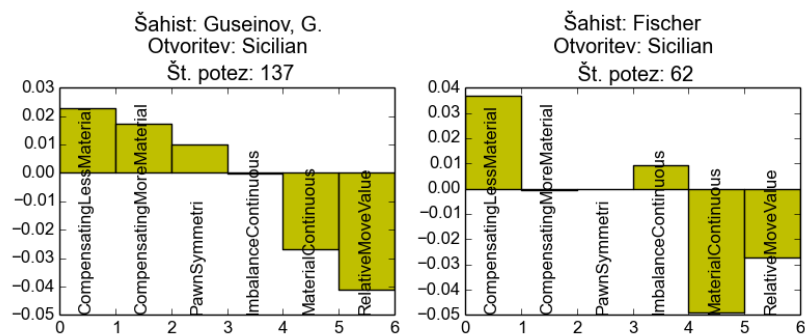
Slika A.64: Movsesian, S., Capablanca. Podobnost: 0.850



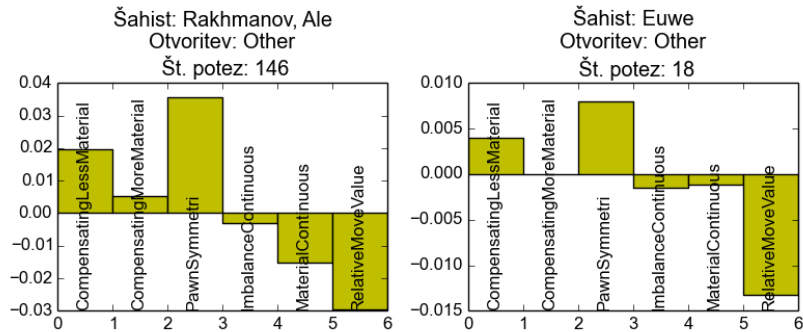
Slika A.65: Frolyanov, D., Karpov. Podobnost: 0.923



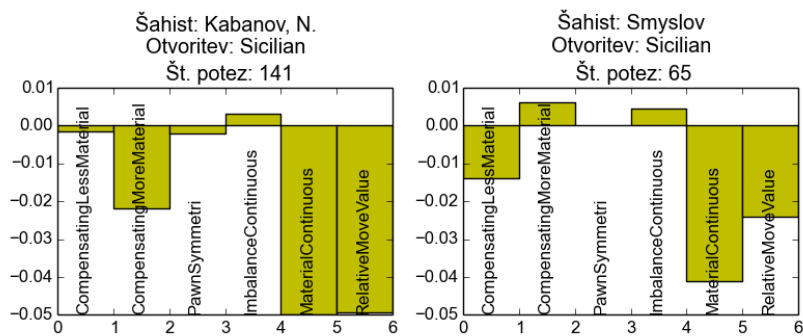
Slika A.66: Ghaem Maghami, E., Anand. Podobnost: 0.844



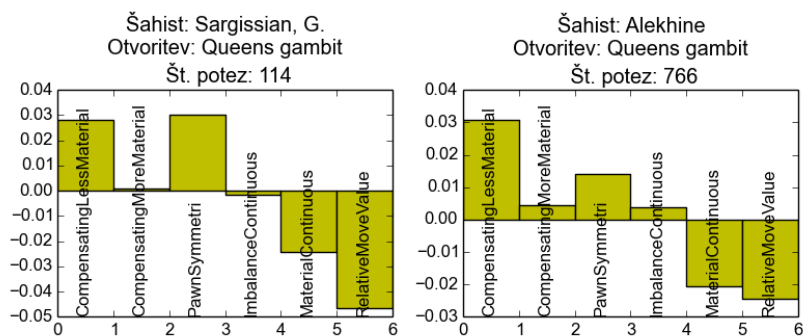
Slika A.67: Guseinov, G., Fischer. Podobnost: 0.838



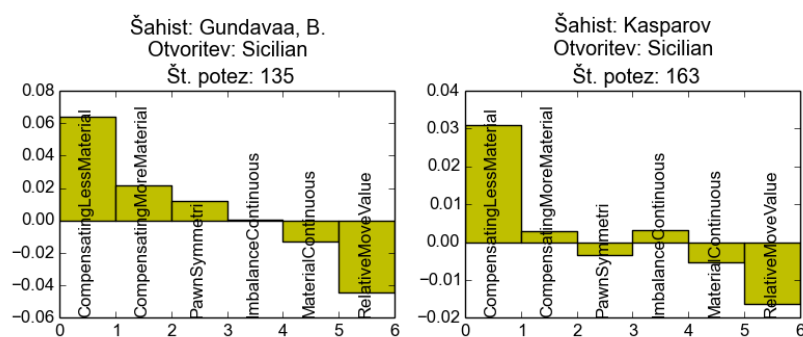
Slika A.68: Rakhmanov, Ale, Euwe. Podobnost: 0.911



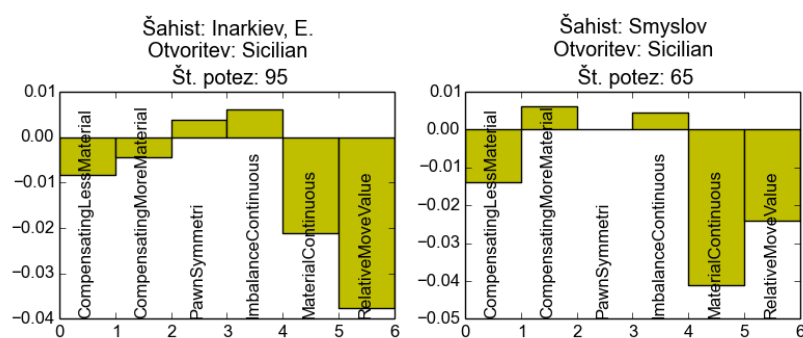
Slika A.69: Kabanov, N., Smyslov. Podobnost: 0.851



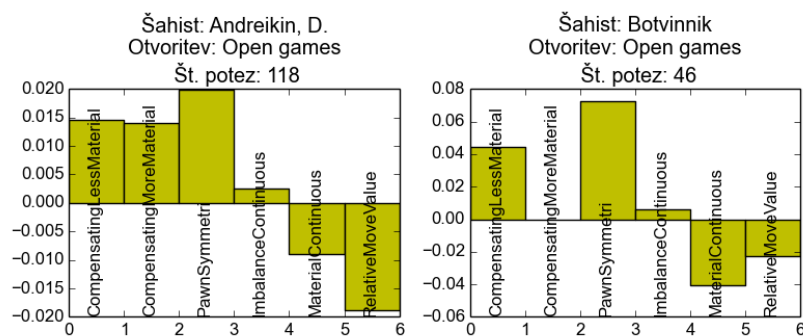
Slika A.70: Sargissian, G., Alekhine. Podobnost: 0.933



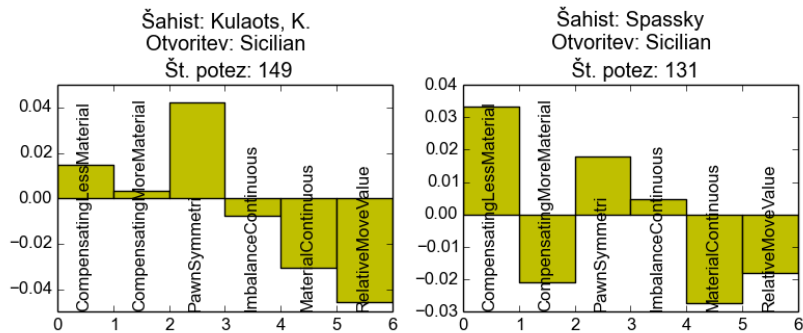
Slika A.71: Gundavaa, B., Kasparov. Podobnost: 0.946



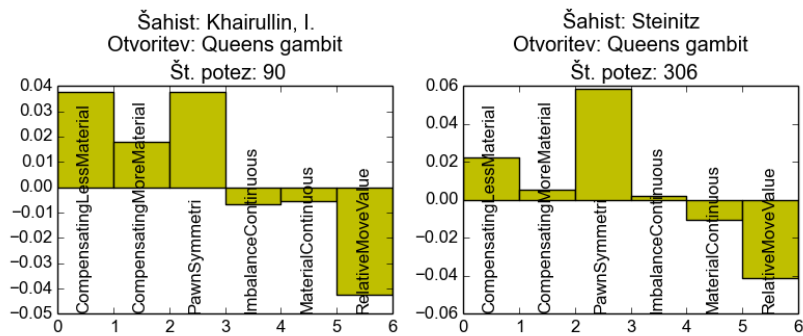
Slika A.72: Inarkiev, E., Smyslov. Podobnost: 0.841



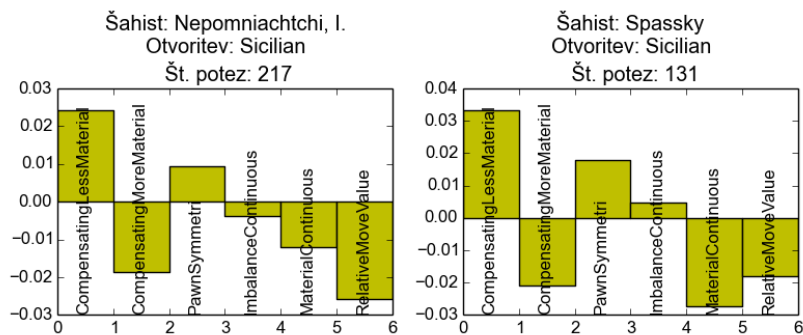
Slika A.73: Andreikin, D., Botvinnik. Podobnost: 0.844



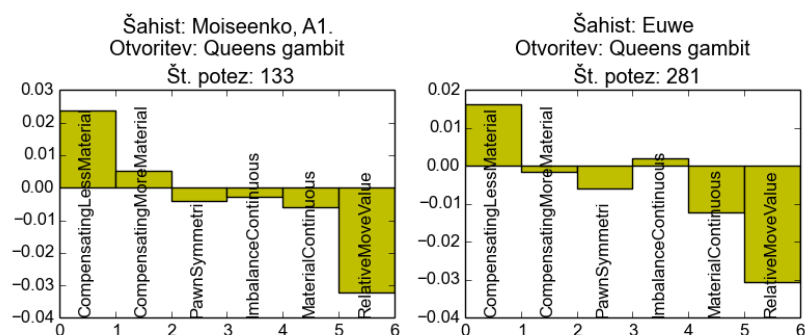
Slika A.74: Kulaots, K., Spassky. Podobnost: 0.722



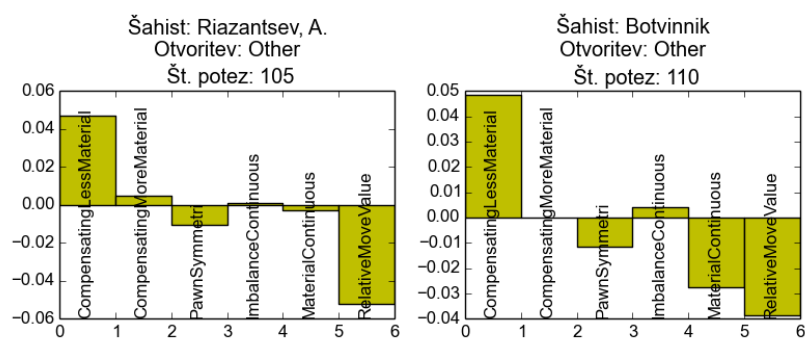
Slika A.75: Khairullin, I., Steinitz. Podobnost: 0.916



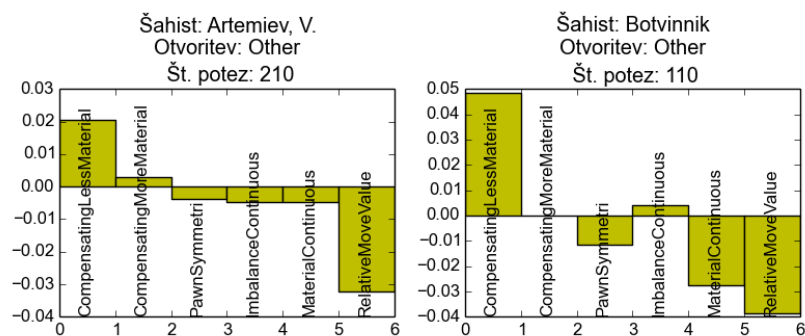
Slika A.76: Nepomniachtchi, I., Spassky. Podobnost: 0.915



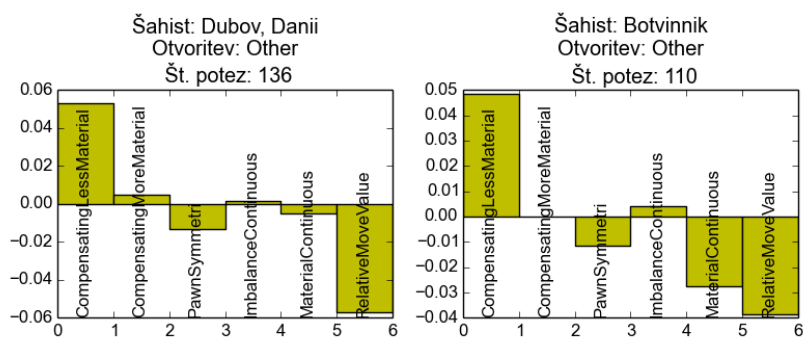
Slika A.77: Moiseenko, A1., Euwe. Podobnost: 0.950



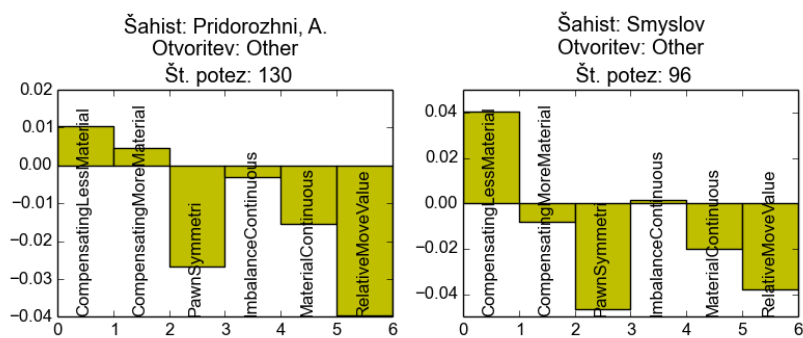
Slika A.78: Riazantsev, A., Botvinnik. Podobnost: 0.915



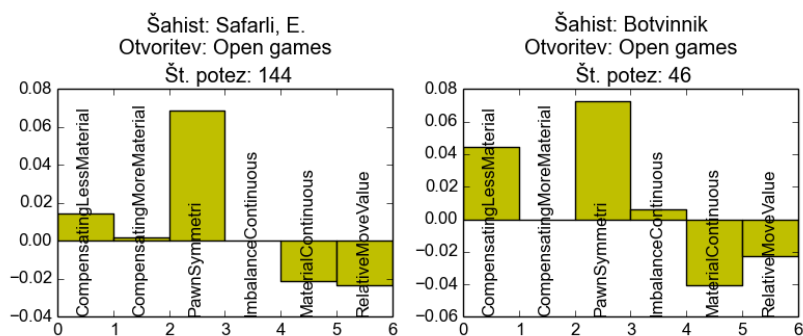
Slika A.79: Artemiev, V., Botvinnik. Podobnost: 0.887



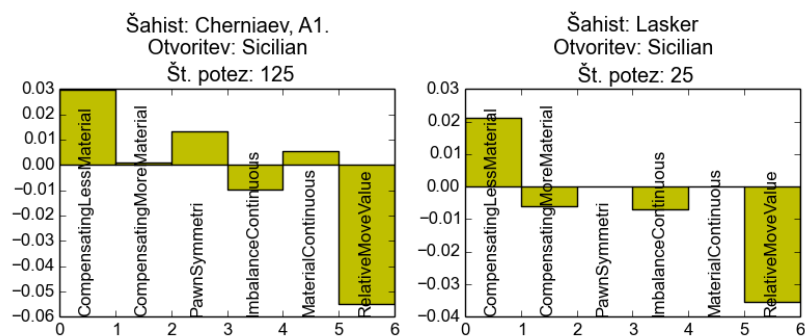
Slika A.80: Dubov, Danii, Botvinnik. Podobnost: 0.926



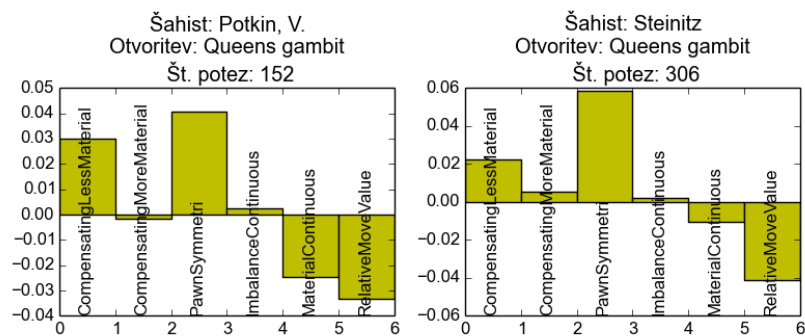
Slika A.81: Pridorozhni, A., Smyslov. Podobnost: 0.879



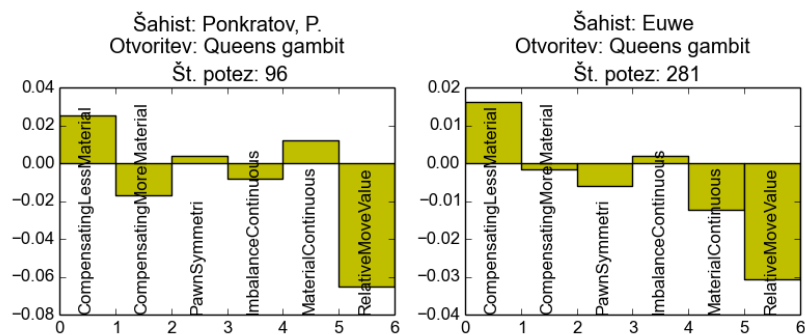
Slika A.82: Safarli, E., Botvinnik. Podobnost: 0.936



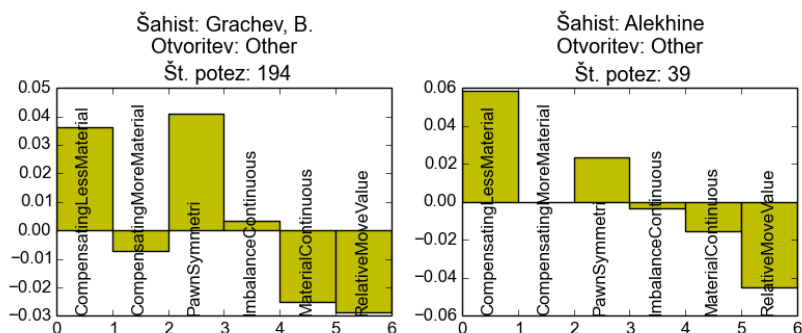
Slika A.83: Cherniaev, A1., Lasker. Podobnost: 0.962



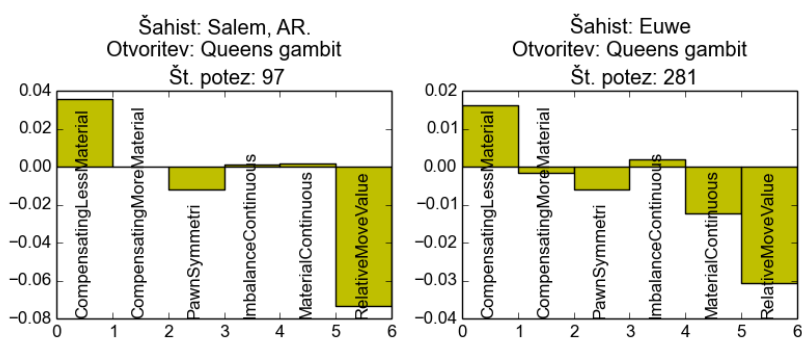
Slika A.84: Potkin, V., Steinitz. Podobnost: 0.941



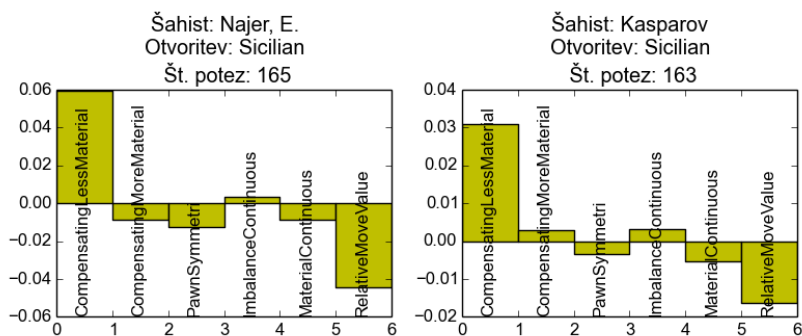
Slika A.85: Ponkratov, P., Euwe. Podobnost: 0.816



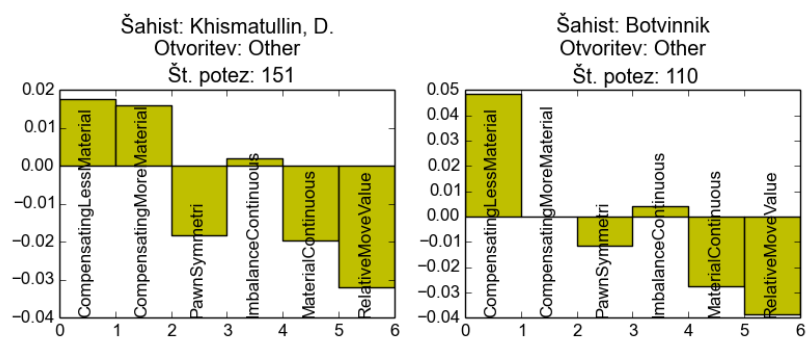
Slika A.86: Grachev, B., Alekhine. Podobnost: 0.893



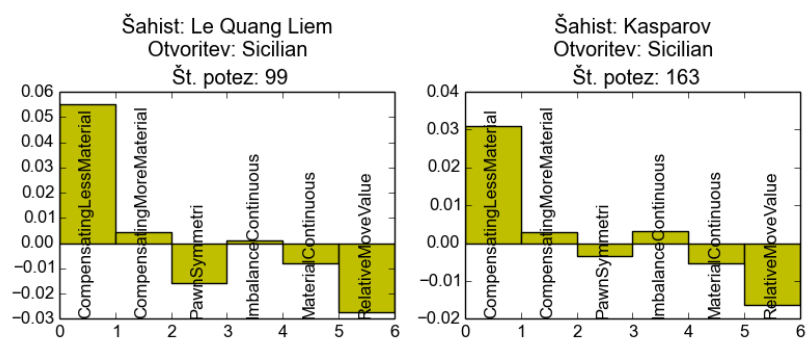
Slika A.87: Salem, AR., Euwe. Podobnost: 0.934



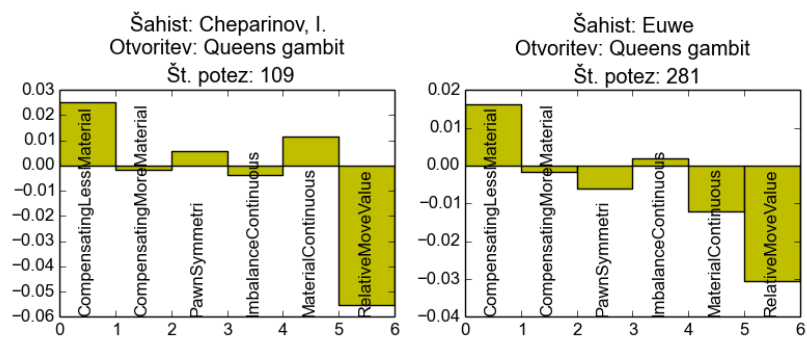
Slika A.88: Najer, E., Kasparov. Podobnost: 0.964



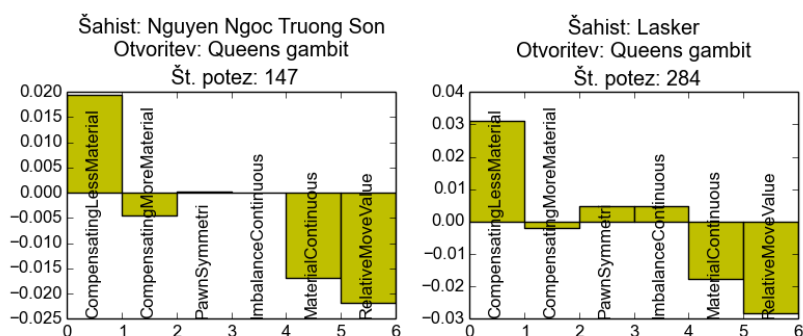
Slika A.89: Khismatullin, D., Botvinnik. Podobnost: 0.858



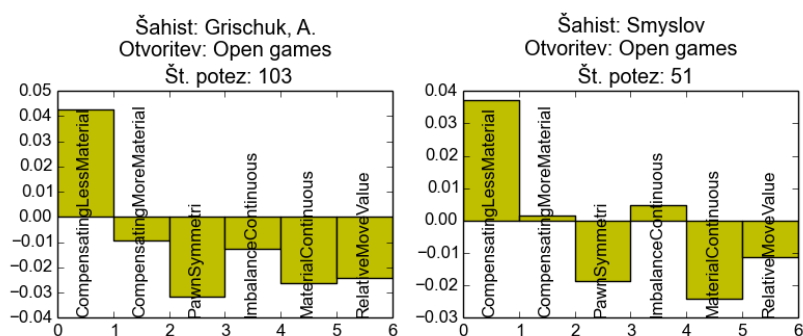
Slika A.90: Le Quang Liem, Kasparov. Podobnost: 0.985



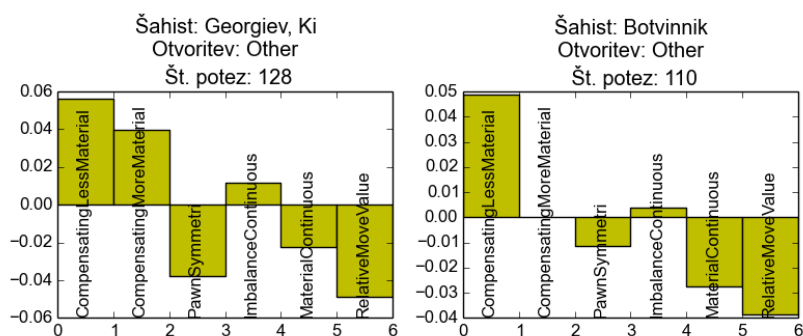
Slika A.91: Cheparinov, I., Euwe. Podobnost: 0.826



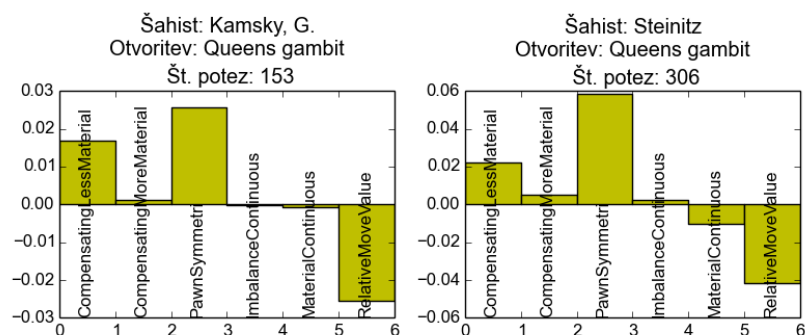
Slika A.92: Nguyen Ngoc Truong Son, Lasker. Podobnost: 0.973



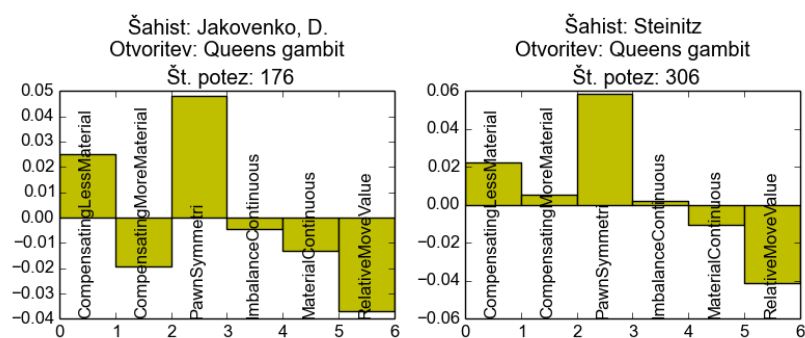
Slika A.93: Grischuk, A., Smyslov. Podobnost: 0.919



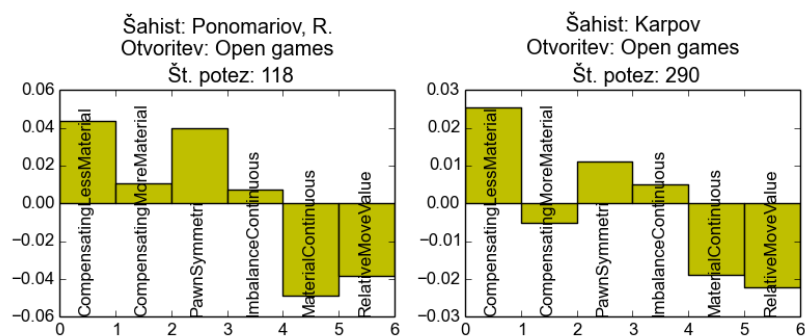
Slika A.94: Georgiev, Ki, Botvinnik. Podobnost: 0.864



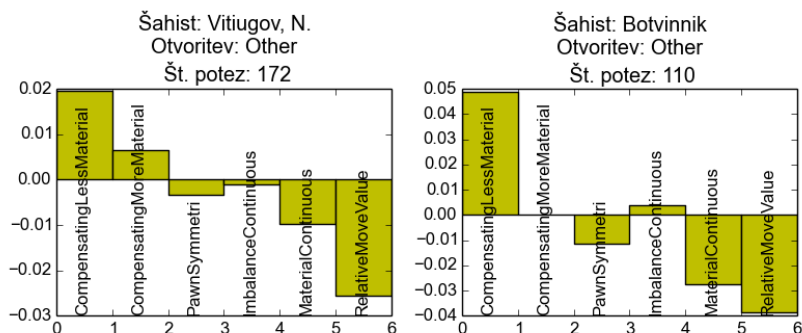
Slika A.95: Kamsky, G., Steinitz. Podobnost: 0.971



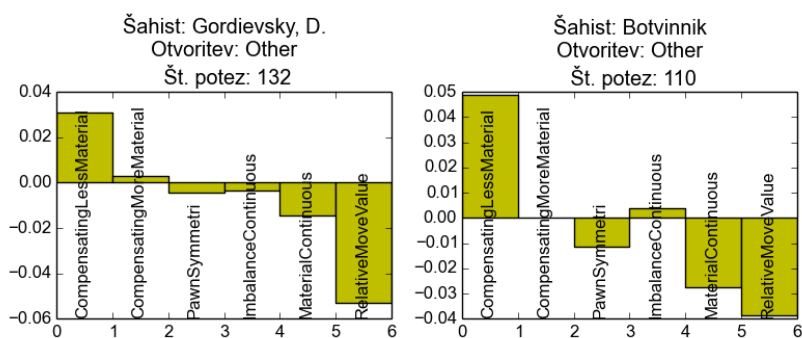
Slika A.96: Jakovenko, D., Steinitz. Podobnost: 0.929



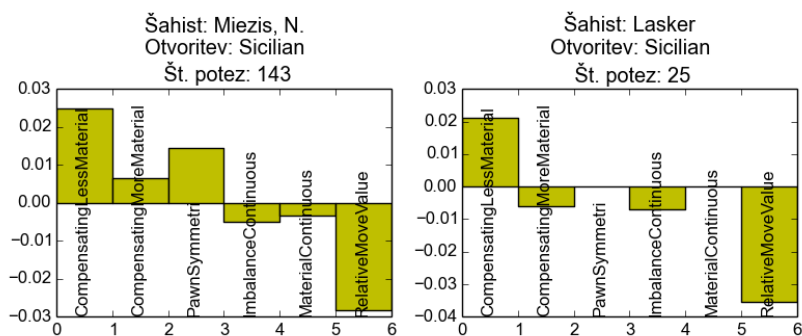
Slika A.97: Ponomarev, R., Karpov. Podobnost: 0.933



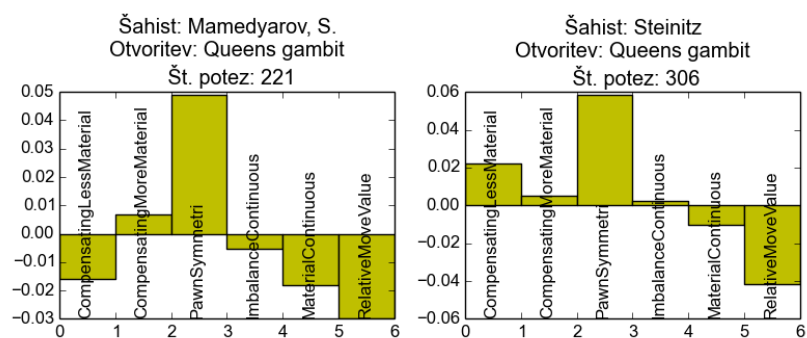
Slika A.98: Vitiugov, N., Botvinnik. Podobnost: 0.942



Slika A.99: Gordievsky, D., Botvinnik. Podobnost: 0.909



Slika A.100: Miezis, N., Lasker. Podobnost: 0.871



Slika A.101: Mamedyarov, S., Steinitz. Podobnost: 0.830