

УПУТСТВО ЗА ПИСАЊЕ ИЗВЕШТАЈА О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ
-обавезна садржина-

| |
|---|
| I ПОДАЦИ О КОМИСИЈИ |
| <p>1. Датум и орган који је именовао комисију</p> <p>25.05.2016., Наставно-научно веће Техничког факултета „Михајло Пупин“ у Зрењанину</p> <p>2. Састав комисије са знаком имена и презимена сваког члана, звања, назива уже научне области за коју је изабран у звање, датума избора у звање и назив факултета, установе у којој је члан комисије запослен:</p> <ul style="list-style-type: none">• Проф. др Ивана Берковић – редовни професор, Информатика и рачунарство, 16.05.2008. Универзитет у Новом Саду, Технички факултет „Михајло Пупин“ Зрењанин, председник,• Доц. др Жељко Стојанов - доцент, Информационе технологије, 27.02.2012., Универзитет у Новом Саду, Технички факултет „Михајло Пупин“, Зрењанин, члан,• Доц. др Далибор Добриловић - доцент, Информационе технологије, 01.10.2012., Универзитет у Новом Саду, Технички факултет „Михајло Пупин“, Зрењанин, члан,• Проф. др Милена Станковић – редовни професор, Рачунарство и информатика, 19.03.1999., Универзитет у Нишу, Електронски факултет, Ниш, члан,• Проф. др Владимир Бртка, ванредни професор, Информационе технологије, 12.12.2013. Универзитет у Новом Саду, Технички факултет „Михајло Пупин“ Зрењанин, ментор |
| II ПОДАЦИ О КАНДИДАТУ |
| <p>1. Име, име једног родитеља, презиме:</p> <p>Вишња, Мирослав, Огњеновић</p> <p>2. Датум рођења, општина, Република:</p> <p>02.10.1969., Кикинда, Србија</p> <p>3. Датум одбране, место и назив магистарске тезе:</p> <p>28.10.2004., Зрењанин, Математичка индукција у дедуктивним системима логичког програмирања</p> <p>4. Научна област из које је стечено академско звање магистра наука:</p> <p>Информатика</p> |
| III НАСЛОВ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ: |
| Апроксимативна дискретизација табеларно организованих података |

IV ПРЕГЛЕД ДОКТОРСKE ДИСЕРТАЦИЈЕ:

Навести кратак садржај са знаком броја страна поглавља, слика, шема, графикана и сл.

Дисертација се бави анализом утицаја расподела података на резултате алгоритама дискретизације у оквиру процеса машинског учења. На основу изабраних база и алгоритама дискретизације теорије грубих скупова и стабала одлучивања, истражен је утицај односа особина података и тачака реза одређене дискретизације. Праћена је промена конзистентности дискретизоване табеле у зависности од положаја редуковане тачке реза на хистограму. Дефинисане су одређене тачке реза у зависности од сегментације мултимодал расподеле на основу којих је могуће радити редукацију преосталих тачака реза. Циљ дефинисан овом дисертацијом је испуњен применом резултата истраживања за конструкцију алгоритама апроксимативне дискретизације APPROX MD. Алгоритам је компариран у односу на алгоритам максималне разберивости и у односу на алгоритам максималне разберивости са апроксимативним решењима за $\alpha=0,95$.

Докторска дисертација припада ужој научној области Информационе технологије. Написана је на српском језику, латиничним писмом, а апстракт у кључној документацијској информацији је дат на српском и енглеском језику.

Дисертација се састоји од осам поглавља. Укупно има 165 страна А4 формата, 18 табела, 232 слике, 10 прилога и 101 литературне референце.

Садржај дисертације обухвата следећа поглавља:

1. Увод
2. Методологија истраживања
3. Алгоритми дискретизације
4. Преглед стања у подручју истраживања
5. Истраживање
6. Алгоритам апроксимативне дискретизације максималне разберивости APPROX MD
7. Компарација алгорита APPROX MD са алгоритмом MD и алгоритмом MD са апроксимативним решењима за $\alpha=0,95$
8. Закључци и смернице будућих истраживања

Литература

Додаци

Испред основног текста се налази наслов рада, кључна документацијска информација, посвета, захвалност и садржај рада.

V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСKE ДИСЕРТАЦИЈЕ:

1. Увод

У првом поглављу наведена је проблематика истраживања. Дефинисани су основни појмови коришћени у дисертацији, наведени основни алгоритми са којима се радило као и теорије на којима су базиране. Прецизирани су битни правци рада дискретизације: испитивање законитости положаја тачака реза на расподели података у односу на одређени алгоритам дискретизације и разматрање инкзистентности дискретизоване табеле.

2. Методологија истраживања

У другом поглављу је наведен методолошки оквир рада. Дефинисани су циљеви и задаци истраживања, методе које су коришћене у истраживању, поступци и инструменти за вредновање класификација и формулисана је главна хипотеза истраживања са потхипотезама.

Главна хипотеза

(X): Могуће је изградити алгоритам за апроксимативне дискретизације у односу на расподеле из табеларно организованих података.

Потхипотезе

X1: Расподеле података утичу на алгоритам дискретизације.

X2: Расподеле података утичу на квалитет класификације.

X3: Дискретизације података може значајно да утиче на инконзистентност табеле код података са одређеним расподелама.

X4: На основу расподеле података могуће је смањити број тачака дискретизације тако да редукт скупа табеларно организованих података остане непромењен.

3. Алгоритми дискретизације

У трећем поглављу дат је приказ основних алгоритама дискретизације као и њихових подела. Посебно су анализирана два алгоритма дискретизације и теорије на којима су развијени:

- Алгоритам дискретизације базиран на ентропији настао у оквиру стабала одлучивања на основу примене алгоритама ID3 и C4.5, који користи принцип дужина минималних карактеристика (minimum description length - MDL), за заустављање рада.
- Алгоритам дискретизације максималне разберивости који је развијен у теорији грубих скупова и који проблем оптималне дискретизације решава на основу коришћења концепата теорије грубих скупова, трансформацијом у проблем минималног редукта.

4. Преглед стања у подручју истраживања

У четвртном поглављу дат је преглед истраживања дискретизација у односу на утицај расподеле података. Приказани су резултати истраживања утицаја расподела података вештачки генерисаних база, на алгоритме базиране на ентропији и друге алгоритме који нису базирани на теорији грубих скупова, односно нису компатибилни са концептом разберивости.

Такође дат је приказ неких алгоритама апроксимативне дискретизације. У оквиру алгоритма максималне разберивости показана је могућност редукције тачака реза на основу њиховог положаја на хистограму података по класама одлуке само код великих база и то само у односу на медијану.

5. Истраживање

У петом поглављу рађена је анализа положаја тачака реза добијених применом алгоритма дискретизације максималне разберивости и алгоритмом базираним на ентропији над десет база које имају различите расподеле. Све расподеле су се посматрале са становишта припадности одређеној дефинисаној групи расподела.

Прва група расподела чине расподеле сличне нормалној, униформној, leptokurtic, platykurtic, skewed, експоненцијалној, zipf, edge peak, comb и осталим расподелама које немају више изразитих максимума између којих су периоди минимума.

Друга група расподела чине расподеле сличне са бимодал и мултимодал расподелама са произвољним бројем локалних максимума.

Истраживање је подељено на три дела и то:

- I Утицај расподеле података на алгоритме дискретизације. Битан закључак је да се тачке реза добијене алгоритмом максималне разберивости налазе након локалних максимума или након дела хистограма мултимодал расподеле који је сличан некој расподели из прве групе.
- II Анализа класификација на основу расподела података. Битан закључак је да се над подацима са мултимодал расподелама који су дискретизовани алгоритмом максималне разберивости примећује већи редукт (са више атрибута) што резултује већим бројем правила и лошијом класификацијом у односу на податке са расподелама из прве групе.
- III Разрада идеје редукције тачака реза добијених алгоритмом максималне разберивости. Идеја је посебно разрађена за базе које имају добар резултат класификације, а посебно за базе са лошим резултатом класификације. За решавање ученог проблема редукције тачака реза

код расподела из друге групе, дат је приказ сегментације мултимодал расподеле и то на два начина. У складу са сегментацијом мултимодал расподела идеја редукције тачака реза је прецизно израђена дефинисањем фиксних тачака.

6. Алгоритам апроксимативне дискретизације максималне разберивости APPROX MD

У шестом поглављу дат је приказ конструкције алгоритма апроксимативне дискретизације максималне разберивости APPROX MD. Базиран је на резултатима – тачкама реза алгоритма максималне разберивости и на параметрима везаним за проценат непрецизних правила, укупни проценат класификације и број тачака редукције.

Два процеса су истакнута као битна и то:

1. Одабир тачака реза за редукцију на основу хистограма;
2. Утврђивање конзистентности дискретизоване табеле тренинг скупа на основу процента непрецизних правила.

Као основа конструкције процеса Одабир тачака реза за редукцију на основу хистограма, на основу претходног истраживања сумирани су и прецизирани услови редукције тачака реза. За одређивање фиксних тачака конструисан је алгоритам FixedPoints који одређује фиксне тачке у складу са грубом сегментацијом мултимодал расподеле и на одабиру тачака реза за редукцију на основу хистограма. За алгоритам FixedPoints процена и осмишљавање грубог „углачавања“ је рађено на основу анализе међусобних односа висине bar-ова хистограма елементарном геометријом. Урађена је имплементација алгоритма FixedPoints у Matlab-у.

Урађена је анализа примене алгоритма APPROX MD како над базом која има добар резултат класификације тако и над базом са лошим резултатом класификације.

7. Компарација алгоритма APPROX MD са алгоритмом MD и алгоритмом MD са апроксимативним решењима за $\alpha=0,95$

У седмом поглављу је урађена компарација алгоритма апроксимативне дискретизације APPROX MD са алгоритмом максималне разберивости и алгоритмом максималне разберивости са апроксимативним решењима за $\alpha=0,95$. Закључено је да постоји бољи резултат класификације алгоритма APPROX MD у односу на резултате добијене алгоритмом максималне разберивости са апроксимативним решењима за $\alpha=0,95$. Такође закључено је да постоји једнак или бољи резултат класификације алгоритма APPROX MD у односу на алгоритам максималне разберивости, на штету лошије конзистентности дискретизоване табеле.

8. Закључци и смернице будућих истраживања

У осмом поглављу су наведена закључна разматрања сумирана у односу на целу дисертацију. Она садрже потврду потхипотеза и главне хипотезе истраживања. Дате су смернице за даља истраживања у циљу имплементације алгоритма APPROX MD.

Литература

У списку литературе постоји 101 литературних референци. Међу списком литературе постоји неколико радова аутора који је креирао алгоритам максималне разберивости. Значајан број радова је новијег датума што потврђује да је тема докторске дисертације актуелна. Највећим делом наведени радови су објављени у врхунским међународним часописима.

Додаци

У оквиру додатака налазе се додаци за сваку од десет анализираних база. За сваку базу дате су тачке реза и хистограми са обележеним тачкама реза алгоритма максималне разберивости и алгоритма базираног на ентропији.

VI СПИСАК НАУЧНИХ И СТРУЧНИХ РАДОВА КОЈИ СУ ОБЈАВЉЕНИ ИЛИ ПРИХВАЋЕНИ ЗА ОБЈАВЉИВАЊЕ НА ОСНОВУ РЕЗУЛТАТА ИСТРАЖИВАЊА У ОКВИРУ РАДА НА ДОКТОРСКОЈ ДИСЕРТАЦИЈИ

уз напомену: Таксативно навести називе радова, где и када су објављени. У случају радова прихваћених за објављивање, таксативно навести називе радова, где и када ће бити објављени и приложити потврду о томе.

1. Brtka E., Ognjenovic V., Brtka V. (2012), *The evaluation of the overall knowledge of the students by usage Dynamic Reducts*, Technics Technologies Education Management, Vol7No4, 11/12.2012., Impakt faktor: 0,351 , ISSN: 1840-1503, pp. 1672-1680. [M23]
2. Brtka V., Brtka E., Ognjenovic V. and Berkovic I. (2011), *The Decision Rules Synthesis Based on Similarity Relation*, SCIENTIFIC BULLETIN of The "POLITEHNICA" University of Timișoara, Romania, Transactions on AUTOMATIC CONTROL and COMPUTER SCIENCE, Vol. 56 (70), No. 3, ISSN 1224-600X, pp. 97-104. [M51]
3. Brtka V., Ognjenovic V., Brtka E., Berkovic I. (2011), *The Rough Sets Based Data Analysis in Small and Medium Sized Enterprises*, 6th IEEE International Symposium on Applied Computational Intelligence and Informatics - SACI 2011, Timisoara, Romania, pp. 373-378. [M33]
4. Ognjenović V., Brtka V., Berković I. (2013), *Discretization influence on data reduction*, International Conference on Applied Internet and Information Technologies ICAИТ 2013, Zrenjanin, Proceedings, ISBN 978-86-7672-211-2, pp. 158-161. [M33]
5. Ognjenovic V., Brtka V., Berkovic I., Brtka E. (2012), *Comparison of the classification rules generated by See 5.0 and SSCO Systems*, Proceedings of the 23rd Central European Conference on Information and Intelligent Systems - CECIIS 2012, Varaždin, Croatia, ISSN 1847-2001, pp. 71-76. [M63]
6. Ognjenović V., Brtka V., Brtka E., Berković I. (2016), *Diskretizacija podataka redukcijom tačaka reza*, INFOTEH-JAHORINA Vol. 15, pp. 665-670. [M33]
7. Ognjenovic V., Brtka V., Jovanovic M., Berkovic I. (2011), *Supervised Discretization for Rough Sets – a Neighborhood Graph Approach*, , Proceedings of the 22nd Central European Conference on Information and Intelligent Systems - CECIIS 2011, Varaždin, Croatia, ISSN 1847-2001, pp. 265 – 272. [M63]
8. Ognjenovic V., Brtka V., Jovanovic M., Brtka E., Berkovic I. (2011), *The Representation of Indiscernibility Relation by Graph*, Intelligent Systems and Informatics (SISY), 2011 IEEE 9th International Symposium, Subotica, Serbia, IEEE Catalog Number: CFP1184C-CDR, ISBN: 978-1-4577-1973-8, pp. 91-94. [M33]

VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА

Анализом односа хистограма података и алгоритама за дискретизацију, истакнута је важност расподеле података на коначан резултат класификације. На основу положаја тачака реза на хистограму, истраживање се прво фокусирало на дефинисање група расподела битних за експеримент. Све расподеле су подељене у две групе.

Два битна правца у оквиру система одлучивања, стабла одлучивања и теорија грубих скупова имају директну корист од детаљних анализа одговарајућих алгоритама дискретизације. Овом дисертацијом је истражена област утицаја особина података датих хистограмима на добијене тачке реза одређене дискретизације: дискретизација алгоритмом максималне разберивости и дискретизација алгоритмом базираним на ентропији MDL.

Над десет изабраних база урађена је анализа хистограма помоћу софтвера EasyFit. Примењени су предвиђени алгоритми за дискретизацију података. Класификација је урађена предвиђеним софтвером Rosetta коришћењем Џонсоновог алгоритма за израчунавање редукта. Дат је одговор на питање колико је резултат класификације добар само на основу укупног процента матрице конфузије, а колико због инконзистентности дискретизоване табеле.

Тачке реза добијене алгоритмом максималне разберивости код расподела из прве групе углавном деле интервал на подједнак број подинтервала. У случају расподела из друге групе, тачке реза се налазе у малој околини тачака прагова сегментације мултимодал расподеле хистограма. Тачке реза добијене алгоритмом базираним на ентропији деле интервал хистограма врло неправилно код обе групе расподела. Нарочито код

мултимодал расподела постоји велика диспропорција у броју објеката које одређују тачке реза.

За већи број елемената редукта и мултимодал расподеле, углавном су се добијали лошији резултати класификације. Код алгоритма максималне разберивости резултат дискретизације генерисао је дискретизовану табелу са мањом инконзистентношћу него алгоритмом базираним на ентропији. Код оба алгоритма дискретизација над подацима који у већини припадају првој групи расподела резултовала је мањом инконзистентношћу дискретизоване табеле него код дискретизације над подацима који у већини припадају другој групи расподела.

Конструисан је алгоритам апроксимативне дискретизације APPROX MD за редукцију тачака реза, који користи тачке реза добијене алгоритмом максималне разберивости и параметре везане за проценат непрецизних правила, укупни проценат класификације и број тачака редукције.

За конструкцију алгоритма апроксимативне дискретизације, резултати истраживања су примењени у разради два процеса: Одабиру тачака реза за редукцију на основу хистограма, и Утврђивању конзистентности дискретизоване табеле тренинг скупа на основу процента непрецизних правила.

Редукција тачака реза добијених алгоритмом максималне разберивости је предложена на основу уочених фиксних тачака реза које одговарају грубој сегментацији мултимодал расподела. На примеру је показан утицај редукције фиксних тачака на промену инконзистентности дискретизоване табеле у односу на редукцију осталих тачака реза. Експериментом је потврђено да редукција фиксних тачака реза више утиче на инконзистентност табеле од редукције осталих тачака. За генерисање фиксних тачака од тачака реза направљен је алгоритам FixedPoints који је имплементиран у Matlab-у.

Алгоритам APPROX MD омогућује да редукција тачака реза само оних атрибута који нису део редукта и не производе промену редукта, а које задовољавају услове за редукцију дефинисане на основу хистограма, бивају допуњене следећом редукцијом тачака реза.

На основу потхипотеза и на основу урађене дисертације може се закључити:

1. Потврђено је да расподеле података утичу на алгоритме дискретизације.
2. Потврђено је да расподеле података утичу на квалитет класификације.
3. Потврђено је да дискретизација података може значајно да утиче на инконзистентност табеле код података са одређеним расподелама.
4. Потврђено је да је на основу расподеле података могуће смањити број тачака дискретизације тако да редукт скупа табеларно организованих података остане непромењен.

Потврђена је главна хипотеза укупним резултатима дискретизације:

Могуће је изградити алгоритам за апроксимативне дискретизације у односу на расподеле из табеларно организованих података.

Као битан резултат ове дисертације предвиђена је модификација бар једног алгоритма за дискретизацију података, тако да омогући рад са апроксимативним вредностима на основу расподела података. Конструкција алгоритма апроксимативне дискретизације APPROX MD у односу на резултате дискретизације максималне разберивости, у односу на конзистентност дискретизоване табеле и у односу на предложену редукцију тачака реза представља тражени резултат.

Добијени резултати утицаја хистограма на тачке реза могу се користити у даљем раду на развоју алгоритма дискретизације. Циљ дефинисан овом дисертацијом је испуњен применом резултата истраживања за конструкцију алгоритма апроксимативне дискретизације APPROX MD.

| |
|--|
| VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА |
| На основу прегледа и анализе докторске дисертације Комисија констатује да су спроведена истраживања приказана на прегледан, адекватан и систематичан начин. Тумачење резултата је у складу са савременим научним сазнањима, аргументовано и свеобухватно. У складу са наведеним, Комисија позитивно оцењује начин приказа и тумачења резултата истраживања. |
| IX КОНАЧНА ОЦЕНА ДОКТОРСKE ДИСЕРТАЦИЈЕ: |
| 1. Да ли је дисертација написана у складу са образложењем наведеним у пријави теме ДА - Дисертација је написана у складу са образложењем наведеним у пријави теме. |
| 2. Да ли дисертација садржи све битне елементе ДА - Дисертација садржи све битне елементе. |
| 3. По чему је дисертација оригиналан допринос науци Оригиналан допринос науци представља алгоритам апроксимативне дискретизације максималне разберивости APPROX MD који на основу резултата алгоритма максималне разберивости и одређивањем фиксних тачака мултимодал расподеле хистограма података на основу конструисаног алгоритма FixedPoints, одређује тачке реза за редукцију на основу дефинисаних услова редукције. Урађена је практична имплементација алгоритма FixedPoints. |
| 4. Недостаци дисертације и њихов утицај на резултат истраживања Нису уочени недостаци дисертације који би утицали на резултате истраживања. |
| X ПРЕДЛОГ: |
| На основу укупне оцене дисертације, комисија предлаже: |
| На основу укупне оцене дисертације, увида у истраживачки рад кандидата и сагласно свим претходно изнетим чињеницама у овом Извештају, Комисија предлаже да се докторска дисертација под називом: Апроксимативна дискретизација табеларно организованих података, кандидата мр Вишње Огњеновић прихвати, а кандидату одобри одбрана. |

ПОТПИСИ ЧЛАНОВА КОМИСИЈЕ

 Проф. Др Ивана Берковић, председник

 Доц. др Жељко Стојанов, члан

 Доц. др Далибор Добриловић, члан

 Проф. др Милена Станковић, члан

 Проф. др Владимир Бртка, ментор

НАПОМЕНА: Члан комисије који не жели да потпише извештај јер се не слаже са мишљењем већине чланова комисије, дужан је да унесе у извештај образложење односно разлоге због којих не жели да потпише извештај.