

ИЗВЕШТАЈ О ОЦЕНИ МАСТЕР РАДА

I ПОДАЦИ О КОМИСИЈИ
<ol style="list-style-type: none">1. Датум и орган који је именовao Комисију Дана 04.12.2019. веће Департмана за математику и информатику Природно-математичког факултета Универзитета у Новом Саду2. Састав Комисије са назнаком 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:<ol style="list-style-type: none">1. др Душан Јаковетић, ванредни професор, Природно-математичког факултета у Новом Саду, ужа научна област: математичко моделирање, изабран у звање: 15.11.2020.- председник2. др Сања Брдар, научни сарадник, ужа научна област: електроника, телекомуникације и информационе технологије, изабрана у звање: 26.10.2017. - члан3. др Татјана Лончар-Турукало, ванредни професор Факултета техничких наука у Новом Саду, ужа научна област: телекомуникације и обрада сигнала, изабрана у звање: 26.04.2017. - ментор
II ПОДАЦИ О КАНДИДАТУ
<ol style="list-style-type: none">1. Име, име једног родитеља, презиме: Наташа, Гордан, Топић2. Датум рођења, општина, република: 24.02.1994., Сомбор, Србија3. Година уписа на дипломске академске студије, смер/усмерење: 2017. Мастер математичар - примењена математика (модул: наука о подацима)
III НАСЛОВ МАСТЕР РАДА
Евалуација алгоритама кластеровања на подацима генских експресија
V ПРЕГЛЕД МАСТЕР РАДА
<p>Мастер рад „Евалуација алгоритама кластеровања на подацима генских експресија“ се састоји од четири поглавља: 1. Мотивација и увод, 2. Материјали и методе, 3. Резултати и 4. Закључци; подељених на мање секције. Рад садржи 41 слику, 4 табеле. Приложен списак коришћене литературе сачињава 23 референце.</p> <p>Тема овог рада је анализа резултата алгоритама за кластеровање на 33 скупа који садрже податке о експресије гена разних канцерогених ткива. Скупови података се разликују по технологијама којим су експресије процењене (чип технологије енгл. <i>Affymetrix</i> и <i>cDNA</i>) и врсти ткива, што захтева различите приступе у иницијалној обради података. У тези се евалуирају и пореде перформансе: алгоритама К средњих</p>

вредности, спектралног ко-кластерованја и спектралног би-кластерованја на овим скуповима података.

IV ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА МАСТЕР РАДА

Прво поглавље представља мотивацију за избор теме и увод у алгоритме бикластерованја и врсте бикластера.

Друго поглавље описује коришћене податке и примењене методе. Како би се што боље разумела природа података описана су укратко два приступа процени генских експресија, две чип технологије, затим основни подаци о коришћеним скуповима као и приступи за нормализацију података који ће бити коришћени. Други део поглавља описује коришћене алгоритме кластерованја и бикластерованја: К средњих вредности, Спектрални ко-кластеринг и Спектрални би-кластеринг. Перформансе ових алгорита на скуповима генских експресија упоређене су користећи једну екстерну и једну интерну меру за валидацију кластера: *Adjusted Rand Index* (ARI) и *Silhouette Index* (SI). Ова два индекса су описана на крају другог поглавља.

Треће поглавље приказује добијене резултате и подељено је у више делова. Поглавље почиње анализом и приказом оригиналних података. Анализирана је расподела узорака по лабелама и одређена је вредност силует индекса за стварну класну расподелу узорака, како би се проценило до које мере су узорци унутар исте класе слични. Након тога приказан је и начин идентификације и елиминације вредности експресија које значајно одступају (*outlier-a*). Следи опис избора нормализација у зависности од технике кластерованја и чип технологије, као и визуелизација нормализованих података. У наставку су приказани резултати кластерованја коришћењем алгоритма К средњих вредности, као стандардног и брзог алгоритма кластерованја, са бројем кластера подешеним на стваран број класа. Потом следе резултати примене алгорита спектралног ко-кластеринга и спектралног би-кластеринга, који врше истовремено груписање и узорака и гена. Добијени кластери и бикластери су визуелизовани коришћењем псеудо-колор репрезентације (*heatmap-a*), а након тога су партиције валидиране помоћу ARI и SI. Последње поглавље износи закључке на основу остварених резултата, као и предлоге за будући рад.

VI ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА

У раду су евалуиране перформансе алгоритама кластеровања који омогућавају симултано кластеровање колона и врста матрице података. Иницијално на подацима је одрађена нормализација да би се униформисао опсег вредности експресија гена и тако боље проценила сличност између узорака. Коришћена су два критеријума за валидацију добијених бикластера (*Adjusted Rand Index* и *Silhouette Index*).

Првобитном анализом на основу оригиналних лабела утврђено је да у просеку сличност узорака са узорцима из исте класе није много већа у односу на сличност са узорцима из других класа. Нехомогеност унутар стварних класа је сугерисала на потенцијалне потешкоће у кластеровању узорака које се врши ненадгледано, на основу сличности између узорака.

Резултати кластеровања над подацима добијеним различитим чип технологијама се значајно разликују. Такође, резултати кластеровања на различитим скуповима података нису конзистентни, одн. ниједан од алгоритама се није издвојио као најбољи на свим коришћеним скуповима података. Ова неконзистентност настаје вероватно услед разлике у врстама ткива и броју подтипова карцинома који су обухваћени једним скупом података. Како подаци експресије гена потичу од различитих ткива, различитих типова канцера и како су првобитно примењене различите редукције димензионалности, за детаљнију анализу и препоруку поступка који би требало пратити било би најпогодније прикупити и разматрати више скупова података са истог ткива и евалуирати више техника за редукцију димензионалности и нормализацију података.

VII КОНАЧНА ОЦЕНА МАСТЕР РАДА

Мастер рад је у потпуности урађен у складу са одобреном темом. Сви алгоритми наведени у пријави теме, су приказани и евалуирани на реалним подацима генских експресија. Рад је прегледно и добро написан, а резултати добијени различитим приступима су јасно приказани и упоређени.

VIII ПРЕДЛОГ

На основу укупне оцене, Комисија предлаже да се мастер рад прихвати, а кандидаткињи Наташи Топић одобри одбрана.

Нови Сад, 24.05.2021.

ПОТПИСИ ЧЛАНОВА КОМИСИЈЕ

др Душан Јаковетић
ванредни професор, ПМФ, председник

др Сања Брдар
научни сарадник, БиоСенсе Институт, члан

др Татјана Лончар-Турукало,
ванредни професор, ФТН, ментор