Стручни критички осврт

Дражен Д. Драшковић, *Софтверски систем за учење и примену алгоритама вештачке интелигенције*. Докторска дисертација. Београд: Електротехнички факултет Универзитета у Београду, 2018, 150 стр. (основни текст на штампаним странама 12–126). Научна област: Електротехника и рачунарство; ужа научна област: Софтверско инжењерство. Ментори: проф. др Бошко Николић и проф. др Милош Цветановић. Комисија: Б. Николић, М. Цветановић, З. Шеварац, П. Тадић, Ј. Протић. Отворени приступ: НаРДуС.

I Предмет, грађа и мерило

Дисертација узима за предмет оно што област образовне технологије ретко добија у једном заокруженом артефакту: софтверски систем који на јединственом корисничком интерфејсу и на више платформи симулира и визуелизује велику и по природи разнородну групу алгоритама вештачке интелигенције — од претраживања и теорије игара, преко резоновања и представљања знања, решавања проблема и рада у неизвесном окружењу, до индуктивног учења — и који се потом уводи у наставу и испробава на неколико примера ван ње. Тежиште рада је инжењерско (изградња и опис система SAIL), уз образовно вредновање у поглављу 7, и то раздвајање биће важно при мерењу свих замерки које следе.

Пре свега осталог ваља изрећи мерило, изричито и без изузетка. Рукопис је предат на преглед и оцену 28. јуна 2018. године. Рад се стога овде мери искључиво према литератури, стандардима, прописима и знању који су постојали и били доступни пре тог тренутка. Ништа што је настало касније није мерило: ни доцнија издања стандарда, ни смернице објављене после предаје, ни промене у алатима које су уследиле. То правило важи и за списак литературе овог осврта — свака његова јединица објављена је пре 28. јуна 2018, изузев саме дисертације, реферата комисије и чињеница које се временом не мењају (садржина објављених дела, историјски подаци), уз назнаку.

Наводи из дисертације преузети су дословно, у наводницима, с бројем штампане стране у загради; текст рада читан је из репозиторијумске верзије доступне у Националном репозиторијуму дисертација. Тврдње о својствима алгоритама и о техничким спецификацијама сравњиване су са стандардном стручном литературом и изворним радовима у верзији важећој до рока.

Најзад, једну ствар треба рећи на почетку јер одређује како се чита цео методолошки слој рада. Највећи део дисертације припада истраживању кроз изградњу артефакта — регистру у коме је конструкција и вредновање софтверског система легитиман научни исход, а не тестирање емпиријске хипотезе природнонаучним поступком (A. R. Hevner, S. T. March, J. Park, S. Ram, „Design Science in Information Systems Research“, *MIS Quarterly*, 28/1, 2004). Строго мерило доказане или оповргнуте хипотезе не сме се, дакле, применити механички на инжењерске доприносе; оно се пуном тежином односи само на онај део закључака који тврди узрок и учинак у настави (в. одељке III и VII). Тај оквир штити рад од једне врсте пренаглашене замерке, али не и од захтева да оно што тврди буде тачно, доследно и одмерено спроведено.

II Шта рад доноси

Осврт је дужан да почне оним што дисертација јесте, јер је то и највећи део онога што она вреди.

Прво, рад даје систематичну таксономију градива и везује је за међународне курикуларне препоруке. Тридесет алгоритама сврстано је у пет јасно разграничених група — „Алгоритми су подељени у пет група: алгоритми претраживања, алгоритми резоновања и репрезентације знања, алгоритми за решавање проблема, алгоритми за рад у неизвесном окружењу и алгоритми машинског учења“ (стр. 17) — а покривеност тема предмета Интелигентни системи изричито је и проверљиво мапирана на препоруке ACM/IEEE: „Из табеле 1 може се видети да су све четири обавезне теме препоручене од стране ACM потпуно покривене, а још три изборне су покривене у мањем или већем обиму“ (стр. 20). То је аналитички најјачи део уводних поглавља и веже наставни садржај за спољашњи, признати оквир.

Друго, рад доноси прегледан упоредни приказ постојећих решења сведен на упоредиве димензије. Четрнаест система приказано је уједначеном структуром и сведено у две табеле (покривеност главних тема и мерила к1–к9), одакле се изводи носећи емпиријски аргумент за нови систем — „ниједан систем нема све карактеристике“ (стр. 53) — при чему су захтеви з1–з9 изведени, а не постулирани: „Карактеристике к1-к9 приказане у табели 4 мапиране су у захтеве“ (стр. 53). Успостављена је, дакле, следива веза преглед → мерила → захтеви → систем.

Треће, рад образлаже избор модела развоја и не оставља га голим: „Приликом развоја софтверског система SAIL коришћен је хибридни приступ, комбиновањем више различитих модела развоја софтвера“ (стр. 56), а хибридни (итеративно-инкрементални с мини-водопадима по модулу) избор изричито се везује за сложеност система и приказује дијаграмом.

Четврто, и то је најизразитији изворни инжењерски допринос рада, за мобилну верзију решен је стваран проблем реализације. Пошто на Андроиду „није било задовољавајуће библиотеке за исцртавање графова и стабала“ (стр. 71), реализован је сопствени метод исцртавања: детекција преклапања чворова („растојање између центара чворова мање од 2R“, стр. 72) с прерачуном координата, детекција додира везе и унапред израчуната позиција стабла. Пројектне одлуке следе из захтева, а не из погодности алата — „Како је почетни циљ био да целокупан софтверски систем буде независан од интернет конекције, одустало се од овог решења да се прави клијент-сервер архитектура“ (стр. 71) — а архитектура је постављена тако да подупире проширивост: „Главна особина архитектуре овог система је модуларност, па се надоградња система може реализовати на веома ефикасан начин додавањем нових алгоритама“ (стр. 64).

Пето, дидактички примери у раду разрађени су и — где год су се могли поново прерачунати — тачни. Разрађени пример претраживања мреже путева даје оптимум који се слаже с графом и хеуристичким вредностима: „Овај метод даје најоптималније решење, дужине 17 километара“ (стр. 99); пример фактора извесности рачунски је доследан; формализација простора стања за образовну игру мисионара и људождера даје тачан број најкраћих решења. Педагошка вредност алата притом је прецизно именована: „У симулацији алгоритма са фактором извесности постоји велики број формула и прорачуна које би корисник система врло тешко савладао и израчунао без коришћења софтвера“ (стр. 90).

Шесто, рад показује научну честитост тамо где је најпотребнија. Слабији налаз извештава се отворено и с узроком — нижа оцена дата је на питању о самосталном учењу „јер систем не може потпуно да замени традиционалне материјале и учење из уџбеника и са слајдова“ (стр. 121); домашај доприноса формулисан је скромно — „позитивна оцена од стране студената који су били укључени у истраживање“ (стр. 125); примене ван наставе поштено су уоквирене („Рад у лабораторији је показног карактера и не оцењује се“, стр. 93); а у самом вредновању постоји свестан нагон да се промена изолује: „Треба напоменути да су у периоду од 2009. године предавања и аудиторне вежбе остале у истом формату, а уведене су лабораторијске вежбе у систему SAIL“ (стр. 118). Уз то, централни артефакт независно је потврђен рецензираним часописним радом [59].

Најзад, потребно је поштено рећи и оно што надилази појединачне вредности: где год су се техничке тврдње и разрађени примери могли сравнити са стандардном литературом и поново прерачунати, у највећем делу су се показали тачним. Изрази сложености минимакса и алфа-бета одсецања, деветокорачно свођење на конјунктивну нормалну форму, аритметика фактора извесности, оптимална дужина пута у примеру претраживања, дефиниције метода расплинуте логике по

спецификацији FCL и мапирање наставних тема на препоруке CS2013 — сравњени с изворима, држе. Ова поузданост фактографске и рачунске основе није споредна околност: она је оно што уопште омогућава да се о раду расправља овако детаљно.

III Циљеви, научна тврдња и метод

Хипотетички оквир рада ваља описати онаквим какав јесте. Рад **не формулише ниједну изричиту, нумерисану хипотезу**; уместо њих поставља три циља (реформа предмета Интелигентни системи по препорукама IEEE CS и ACM; реализација система који подржава кључне теме уводног курса уз електронско, мобилно и самостално учење; приказ примене и ван наставе) и четири узастопна корака истраживања који се поклапају с редоследом поглавља. Уз то стоји радна претпоставка целог рада — да „не постоји систем који обједињује све подржане алгоритме и њихов приказ“ (стр. 14). Прва два циља су пројектна и, у делу изградње артефакта, испуњена су; трећи и, изнад свега, прећутна главна тврдња вредновања носе узрочну компоненту (да студенти уз систем боље разумеју градиво и постижу бољи успех).

Управо ту је тежиште методолошке оцене. Пројектни доприноси мере се исходом реализовано или показано, што за изградњу и опис система стоји, и додатно је потврђено спољашњом верификацијом рада [59]. Али узрочна образовна тврдња мери се другим мерилом. Нацрт вредновања је поређење различитих генерација студената (јун 2009. као једина тачка пре, наспрам 2012–2015), без контролне групе. Из таквог нацрта узрочност не следи строго: разлике међу генерацијама могу потицати и од промене популације, промене структуре оцењивања, самоселекције на изборном предмету или општег тренда. Референтно дело о закључивању о узрочности без случајног распоређивања баш то и кодификује: селекција (различите генерације) и инструментација (промена онога што се бодује) именоване су претње унутрашњој ваљаности (W. R. Shadish, T. D. Cook, D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, 2002). Методолошки уџбеник експеримента у софтверском инжењерству признаје квазиексперименте као легитиман нацрт када рандомизација није изводљива — али уз изричиту обавезу да се именоване претње ваљаности размотре (C. Wohlin и др., *Experimentation in Software Engineering*, 2012); исто траже и смернице за емпиријско истраживање у струци (B. A. Kitchenham и др., „Preliminary guidelines for empirical research in software engineering“, *IEEE Transactions on Software Engineering*, 28/8, 2002). Читано, пак, као студија случаја једног предмета на једној установи, вредновање је ваљан нацрт, али онда га обавезује језик спољашње ваљаности студије случаја, а не језик доказаног ефекта (P. Runeson, M. Höst, „Guidelines for conducting and reporting

case study research in software engineering“, *Empirical Software Engineering*, 14/2, 2009). Појединости ове замерке дате су у одељку VII; овде је довољно рећи да су закључци о учинку каузални, а нацрт то не подупире.

Треба, ради правичности, одмах додати и оно што иде у прилог раду. Слаби нацрти чести су у истраживању рачунарског образовања (J. J. Randolph и др., „A Methodological Review of Computer Science Education Research“, *Journal of Information Technology Education: Research*, 7, 2008) — што није оправдање, али јесте контекст. И, што је важније, рад ради с целом популацијом генерације (а не с узорком), трендови су доследни кроз више година и на оба одсека, а аутор свесно наводи шта је остало непромењено (стр. 118). Ти чиниоци чине налаз уверљивим у оквиру предмета; они не претварају корелацију у доказан узрок.

IV Приказ алгоритама вештачке интелигенције

Приказ тридесет алгоритама носи изразито стручну и разнородну материју, и у целини је технички солидан: сравњени са стандардном литературом, описи су у највећем делу тачни и добро повезани с изворима (свођење на конјунктивну нормалну форму, резолуција с провером појаве у унификацији, метод задовољења ограничења с AC-3, стабла одлучивања и ID3 с информационом добити, сложеност минимакса и алфа-бета одсецања). Замерке које следе локалне су, углавном терминолошке или уџбеничко-поједностављујуће, и ниједна не погађа тезу — али их у монографском издању ваља отклонити.

Најозбиљнија међу њима тиче се Рете алгоритма, коме се приписује да „за разлику од класичног алгоритма директног уланчавања има константно време извршавања“ (стр. 30). То није тачно: Рете инкременталним чувањем стања упаривања битно смањује број поређења, али му време зависи од броја чињеница, правила и промена по циклусу — не постаје константно; сам рад у наредној реченици каже да алгоритам смањује број поређења (C. L. Forgy, „Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem“, *Artificial Intelligence*, 19/1, 1982). У истом реду прецизности су и: сложеност алфа-бета одсецања дата без услова под којим важи (то је сложеност најбољег случаја, уз близу оптималан редослед потеза, стр. 24); тврдња да је STRIPS планирање „NP-тежак проблем“ (стр. 36), што је тачна доња граница, али слабија од резултата извора — пропозиционо STRIPS планирање је у општем случају PSPACE-комплетно (T. Bylander, „The computational complexity of propositional STRIPS planning“, *Artificial Intelligence*, 69/1–2, 1994); и ограничење ID3, где су континуалне величине означене као редне: „садржи само дискретне податке, насупрот редним подацима као што су године, висина“ (стр. 47) — поента да ID3 захтева дискретне атрибуте остаје тачна и на истој страни изричито наведена.

Неколико поистовећивања захтева поправку. Стабла одлучивања уопштено се именују као CART — „Из тог разлога се често називају CART“ (стр. 114) — иако је CART конкретна методологија (L. Breiman и др., *Classification and Regression Trees*, 1984), а ID3 и C4.5 нису CART. Уводна дефиниција гејмификације — „Гејмификација у едукацији је процес који помаже ученицима и студентима да уче различита знања кроз игре“ (стр. 75) — заправо описује учење засновано на играма и одступа од дефиниције цитираног извора, који гејмификацију одређује као примену елемената дизајна игара у неиграчком контексту (S. Deterding и др., „From game design elements to gamefulness“, 2011). Дефазификација се на једном месту непрецизно изједначава с концентрацијом (стр. 69), појмом који је у расплинутој логици посебан прилог; оквири се описују као да „примењују само IS_A везе“ (стр. 32), што важи за међуоквирне релације приказаног модела, али не и за прегратке атрибута; а kNN, тачно сврстан у надгледано учење, касније се везује за „око 5000 кластера“ (стр. 116), појам ненадгледаног учења. Уз то, порекло STRIPS-а непрецизно је означено — „На Универзитету Стенфорд је седамдесетих година прошлог века развијен STRIPS алгоритам“ (стр. 36) — док је развијен у самосталном институту SRI International (R. E. Fikes, N. J. Nilsson, „STRIPS: A new approach to the application of theorem proving to problem solving“, *Artificial Intelligence*, 2/3–4, 1971).

Најзад, једна тврдња преслободно тумачи курикуларни извор: да студент област вештачке интелигенције изучава „најмање једном предмету, а највише три предмета“ (стр. 17) погрешно чита вредности 1–3 из препоруке CC2005, које су оцене релативног нагласка на скали 0–5, а не број предмета; суштинска поента о умереном простору за вештачку интелигенцију у курикулумима ипак остаје. Сви ови налази ослањају се на изворне радове и уџбеник на који се рад и сам претежно позива (S. J. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, 3. изд., 2009), и — да се понови — ниједан не дира носеће тврдње.

V Преглед постојећих система, захтеви и модел развоја

Упоредни преглед и из њега изведени захтеви носе аргумент новине целог рада, па је важно да им поступак буде проверљив. Ту је рад делом уредан, а делом недоречен.

Уредан је у томе што системе своди на упоредиве димензије и што захтеве изводи из уочених мерила. Недоречен је у три тачке које стоје у пуном обиму, мада ниједна не руши закључак. Прва: мерило по коме је баш четрнаест система ушло у преглед (обухват, база претраге, услови укључивања) није изричито наведено, нити је операционализовано ко је и када доделио троставне квалитативне оцене к1–к9. Ако се преглед не представља као систематски, то није прекршај — смернице за систематски преглед описују шта би такав преглед захтевао, а не обавезују несистематски (B. Kitchenham, S. Charters, „Guidelines for performing Systematic Literature Reviews in Software Engineering“, EBSE-2007-01, 2007) — али изричит опис

обухвата и мерила ојачао би проверљивост носеће тврдње о непостојању потпуног система. Да то није неизводљив захтев, сведочи објављен, рецензиран упоредни преглед образовних симулатора с изричитим мерилима из истог круга који рад цитира (B. Nikolić, Z. Radivojević, J. Đorđević, V. Milutinović, „A Survey and Evaluation of Simulators Suitable for Teaching Courses in Computer Architecture and Organization“, *IEEE Transactions on Education*, 52/4, 2009) — истовремено образац и позитиван преседан.

Друга тачка тиче се једног од мерила: к9 (реализована мобилна апликација) особина је коју ниједан постојећи систем нема, а коју нови систем по замисли има, па је закључак да ниједан постојећи систем нема баш све особине делом обезбеђен избором мерила скројеног према сопственом решењу. Ово ваља поштено сузити: мобилно и самостално учење независно је постављен педагошки циљ рада (стр. 16–17), а не накнадни критеријум, па је реч о ослабљеном, не и обореном аргументу.

Трећа тачка апаратске је природе, али погађа потпуност поређења: систем Minesweeper (University of Nebraska) уврштен је у обе упоредне табеле, али у тексту није појединачно описан, за разлику од осталих тринаест; будући да је реч о систему исте класе проблема и с исте установе као већ описани Sudoku Solver, оцене су аналогне и закључак остаје непромењен, али изостанак описа ваља отклонити. У истом реду су и ситне садржинске и апаратске омашке у прегледу: тврдња да се MIT-ов алат користи „и на Универзитету у Чикагу“ није поткрепљена наведеним извором ни табелом; подела тог алата на шест целина укључује Бајесове мреже, којих међу његовим демонстрацијама нема; систем AIMA3e у табели је позван на референцу [10] (анкету о симулаторима за архитектуру рачунара) уместо на [47] коју текст исправно користи; а установа система Chinese room дата је с обрнутим редом речи. Ниједна од тих омашки не дира садржину поређења.

Одељак о моделу развоја софтвера и извођење захтева методама случајева коришћења и прича коректно су изведени и почивају на добро изабраним примарним изворима.

VI Систем SAIL: архитектура, технологије, рад система и примене

Опис реализације носи главнину инжењерског доприноса и, на архитектонском нивоу, добро је изведен: модуларна подела на четрнаест модула по пакетима именованим према групама алгоритама, сопствени формат датотеке по модулу, вишеплатформска изведба (рачунарска и мобилна апликација из приближно исте спецификације) и, као издвојен изворни резултат, сопствено исцртавање графова и стабала на Андроиду разложено на потпроблеме и решено изведеном геометријом. Тај део рад савладава.

Три врсте замерки, ниједна тешка, ипак стоје. Прва је датираност приказа алата. Попис језика за Unity наводи и оне који су у тренутку предаје већ били укинути — „За писање пројеката могу се користити програмски језици JavaScript, C# или Boo“ (стр. 76) — при чему је званична подршка за Boo укинута још уз Unity 5, а C#, једини преостали и у раду коришћени језик, наведен је исправно; сличан је и датиран попис графичких интерфејса алата (изостају новији). То је чињенична замерка о ажурности, без стандарда иза себе, и погађа каталошки, не носећи слој.

Друга је поновљивост. Систем SAIL (рачунарска и мобилна апликација, игре), програмски код, примери, упитник и сирови подаци нигде у раду нису дати као јавно доступни (без адресе, складишта кода или продавнице), а верзије кључних алата (JUNG, ANTLR, Unity 3D) нису наведене. Захтев да се забележе тачне верзије свих спољних алата и да се код и подаци учине доступним није произвољан: он је изричит у смерницама за поновљивост рачунарских истраживања (V. Stodden и др., „Enhancing reproducibility for computational methods“, *Science*, 354, 2016; G. K. Sandve и др., „Ten Simple Rules for Reproducible Computational Research“, *PLoS Computational Biology*, 9/10, 2013), у начелима цитирања софтвера која траже верзију и трајни идентификатор (A. M. Smith, D. S. Katz, K. E. Niemeyer, „Software citation principles“, *PeerJ Computer Science*, 2, 2016) и у политици оцене артефаката (ACM, *Artifact Review and Badging*, в. 1.0, 2016). У истом реду је и то да најављени развој вођен тестовима и понашањем (TDD/BDD) не прати ниједан документован тест ни опис онога што је тестирано. Тежину ове замерке ваља, ипак, поштено сузити: централни артефакт спољашње је верификован рецензираним радом [59], што надомешта део изостале документоване поновљивости.

Трећа се тиче примена ван наставе. Оне су исправно уоквирене као приказ могућности, а не као вредновање, па им се одсуство мерења учинка не може узети за пропуст. Ипак, огледи над великом количином података (поглавље 6.3.3) користе kNN и линеарну регресију — технике које нису међу симулацијама самог система — а наведени бројеви дати су без дефиниције метрике, без описа поделе на скуп за учење и скуп за проверу и без доступних података, што њихову доказну снагу за трећи циљ чини ограниченом. Мерило је овде склад тврдње и вредновања: ако исход не одговара врсти потврде, тешко се пружа убедљив доказ (M. Shaw, „Writing good software engineering research papers“, *Proc. ICSE 2003*).

VII Вредновање у настави и закључци који се из њега изводе

Поглавље 7 је место на коме рад највише тврди и на коме мерило најстроже пита. Овде се сустиче неколико замерки, а једна је носећа за закључке рада.

Основа вредновања почива на поређењу успеха генерација пре и после увођења система. У истом раздобљу, међутим, промењена је и структура оцењивања: домаћи задатак уведен од школске 2009/2010. „носи 20% од укупног броја поена на предмету“ (стр. 120) и заменио је теоријска питања, па „просечан број поена на испиту“ из 2009. није у потпуности упоредив с оним из 2012–2015. Ту је важно рећи оно што иде у прилог раду: пролазност (однос положили/уписани) том променом није захваћена, јер мери положен/уписан однос, па се закључак сигурније ослања на њу него на просечне поене.

Друга замерка тиче се извештавања. Ни у једној табели вредновања нема мере расипања, теста значајности, интервала поверења ни величине ефекта; разлике се приказују само као сирови проценти и просеци, а узорак на изборном одсеку је мали (десетине студената по години), па се разлике у процентима теже могу назвати утврђеним ефектом. Водеће препоруке о статистичком извештавању траже да се увек наведе величина узорка и прикажу величине ефекта и мере расипања уз показатеље значајности (L. Wilkinson, Task Force on Statistical Inference, „Statistical methods in psychology journals“, *American Psychologist*, 54/8, 1999; APA Working Group on JARS, „Reporting Standards for Research in Psychology“, *American Psychologist*, 63/9, 2008), а на малим узорцима питање снаге и величине ефекта постаје пресудно (J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2. изд., 1988). Овде мерило иде и корак даље у корист прецизности приговора: пошто рад уопште не изводи тест значајности, ни само израчунавање показатеља значајности не би било довољно, јер он не мери величину ни важност ефекта — „A p-value, or statistical significance, does not measure the size of an effect or the importance of a result“ (R. L. Wasserstein, N. A. Lazar, „The ASA Statement on p-Values“, *The American Statistician*, 70/2, 2016, начело 5). Тежиште замерке је, дакле, изостанак величине ефекта и мере несигурности, а не пуко одсуство једне статистике.

Трећа замерка, и најозбиљнија, тиче се самог језика закључака. Из поређења генерација без контролне групе изводе се узрочне тврдње: „Као резултат увођења овакве симулације“ (стр. 117), а у закључку да је евалуација „показала да студенти постижу много већи број поена“ (стр. 124). Из таквог нацрта то не следи; тврдњу је требало исказати као повезаност доследну с увођењем система, не као доказан узрок. Исти помак понавља се код времена припреме — „студенти су трошили много мање времена за самосталну припрему овог испита, него студенти који нису користили софтверски систем SAIL“ (стр. 124) — где формулација сугерише паралелне групе, а реч је о поређењу кохорти по годинама, с умереним помаком и без теста значајности. Управо овде вреди подсетити на мерило по коме се и сам централни педагошки суд рада (да визуелизација води бољем разумевању) мора одмерити: мета-студија ефективности визуелизације алгоритама налази да „how students use AV technology has a greater impact on effectiveness than what AV technology shows them“ (C. D. Hundhausen, S. A. Douglas, J. T. Stasko, „A Meta-Study of Algorithm

Visualization Effectiveness“, *Journal of Visual Languages and Computing*, 13/3, 2002, стр. 259), а образовну вредност визуелизација везује за активно ангажовање ученика (T. L. Naps и др., „Exploring the role of visualization and engagement in computer science education“, *ITiCSE Working Group Reports*, 2002). То мерило раду делом иде у прилог — интерактивна симулација корак по корак и задаци управо омогућавају активно ангажовање — али и оштри тврдњу: оно што вредновање показује јесте систематски бољи исход уз увођење система, а не доказана веза ангажовања и ефекта.

Четврта замерка тиче се анкете, којом се поткрепљује задовољство студената. Инструмент је документован оквирно: наведен је само удео — „узело учешће око 80% студената“ (стр. 121) — без апсолутног броја, без разлагања по одсецима и годинама и без иједне речи о томе како је упитник састављен и да ли су проверене његова ваљаност и поузданост. Смернице за анкете у струци траже управо процену поузданости и ваљаности инструмента и његово предтестирање (B. A. Kitchenham, S. L. Pfleeger, „Principles of survey research“, део 4, *ACM SIGSOFT Software Engineering Notes*, 27/3, 2002; исти, „Personal Opinion Surveys“, у *Guide to Advanced Empirical Software Engineering*, 2008), а постојао је и готов, валидиран инструмент за оцену употребљивости (J. Brooke, „SUS: A ‘quick and dirty’ usability scale“, 1996). Најозбиљније: просечне оцене по анкетним питањима нигде нису дате бројем — „Просечна оцена по питањима приказана је у табели 9“ (стр. 121), а табела 9 приказује испитне поене, не оцене анкете; засебне табеле с тим оценама нема. Тиме носећа тврдња о задовољству остаје непроверљива из самог рада. Уз то се на једном месту мешају различити инструменти и трену̂ци (наменска анкета и акредитациона процена ЕСПБ, с неусклађеним годинама).

Пета замерка тиче се тога шта поједине мере заправо мере. Оцене наставника и асистента и процена ЕСПБ мере задовољство и оптерећење, а не непосредно педагошки учинак; та табела нема основицу пре увођења система, па не може поткрепити ефекат. У истом делу постоји и осетљива тачка коју рад не проблематизује: аутор је асистент на предмету чији рад анкета оцењује. Ту ваља бити одмерен — анонимност анкете и њен институционални карактер практично умањују тај ризик, а изричита сагласност испитаника није била уобичајена пракса у контексту тог доба — али једна реченица о сукобу улога била би на месту, тим пре што је етика студија у којима су испитаници студенти зависни од истраживача изричито обрађена у струци (J. Singer, N. G. Vinson, „Ethical issues in empirical studies of software engineering“, *IEEE Transactions on Software Engineering*, 28/12, 2002).

Најзад, поједине закључне тврдње уопштавају налазе с једног предмета на једном факултету на „студенте“ и „инжењере“ уопште; такав домашај превазилази прикупљене податке, иако модуларна и вишеплатформска изведба чини

преносивост начелно вероватном, али емпиријски непоказаном изван овог предмета. Границе таквог уопштавања управо су предмет спољашње ваљаности студије случаја (Runeson, Höst, 2009).

Треба, ради правичности, изричито поновити и супротну страну биланса, јер она није мала. Рад поштено извештава о слабијем налазу и његовом узроку (стр. 121), скромно формулише допринос (стр. 125), образлаже избор мерила — само први испитни рок, „јер статистика показује да између 60% и 75% студената излази на полагање овог предмета управо у првом испитном року“ (стр. 118) — и свесно наводи шта је остало непромењено да би издвојио увођење система (стр. 118). То је исправан методолошки нагон и најјачи ослонац одбране; он не претвара корелацију у узрок, али чини да замерка погађа формулацију и извештавање, а не поштење рада.

VIII Слике, табеле, формуле и језик

Презентациони слој рада тражи темељну редакцију, мада ниједна од омашки не дира садржину.

Снимци екрана рачунарске верзије приказани су у ширини стране, с врло ситним, местимично нечитким натписима, што смањује проверљивост описа. Језик натписа није усаглашен: делови интерфејса и легенде дати су на енглеском у раду писаном ћирилицом. То је најизразитије на слици броја студената, чија је легенда на енглеском, чије осе немају означену јединицу и чија друга серија у тексту није објашњена. Мешање писама и језика у натписима, као и писање децималног знака тачком уместо зарезом, приговори су с писаним ослонцем (Правопис српскога језика, Матица српска, 2010).

Унутрашње недоследности и погрешне унакрсне референце воде се као чињенични, а не нормативни приговор, јер за њих струковног стандарда нема сем општих упутстава за уредност. Носећи податак о порасту броја студената упућен је погрешно — „постепено се повећавао кроз школске године“ (стр. 117), уз упуту „што се може видети са слике 46“ (стр. 117), док прави приказ стоји на слици 49; притом опис „постепено“ благо преувеличава доследност, јер серија има два минорна пада. У садржају се, као лажан унос с бројем стране 74, појавила цела реченица из тела рада (последица погрешно додељеног стила наслова); у списку табела уз једну табелу стоји немогућ број стране; а проценат уз однос 86/113 требало је да гласи 76,11, а не 76,22 (сам однос је тачан, остале вредности се слажу, и закључак се не мења).

Најзад, низ ситних словних и терминолошких недоследности ваља лекторски отклонити: наслов трећег поглавља штампан је као „РЕАЛИЗИЈУ“ (док је у пододељку 3.3 иста реч исправна), уз „Имлементација“ на слици и неколико омашки у телу; страни термин писан спојено „selflearning“ (стр. 13); назив игре „Nine man morris“; исти систем зван „Searcher“ у тексту, а „Searchr“ у табелама; интерфејс

библиотеке JUNG наведен као „Rendered“ уместо „Renderer“; и алат писан „Graphiz“ уместо „Graphviz“. Ниједна од ових ситница не мења значење, али њихово мноштво нарушава уредност коју рад иначе показује.

IX Научни апарат

Научни апарат овог рада изнад је просека уредан, и то треба рећи пре замерки. Позивање је нумерисано по IEEE стилу, попис прати тај образац, редослед првог помена тачно одговара нумерацији, свако позивање има одговарајућу јединицу и свака јединица цитирана је бар једном; нема двоструких јединица ни посредног цитирања, а темељни историјски радови наводе се из примарних извора. Јака је и ажурност пописа.

Замерке су ограничене и претежно се тичу документованости. Кључни алати рачунарске апликације — JUNG, ANTLR, језик FCL и Swing — детаљно су описани у тексту, али нису унети у попис литературе с верзијом и извором, за разлику од алата Graphviz и пакета WEKA који јесу референцирани; та недоследност отежава поновљивост и мери се начелима цитирања софтвера (Smith, Katz, Niemeyer, 2016) и захтевом за потпуношћу библиографске јединице и датумом приступа за онлајн изворе (ISO 690:2010; IEEE Editorial Style Manual, изд. 2017 — образац који рад сам усваја као [103]). Један онлајн извор нема датум приступа. Уз то, за неколико темељних алгоритама рад се ослања на уџбеник и секундарне изворе без изворних радова — што је, с обзиром на предмет, легитиман избор, али би навођење примарних радова употпунило апарат. То што рад по IEEE стилу не означава изричито која су позивања сопствени ранији радови уобичајено је за тај стил и не води се као пропуст.

X Однос рада и реферата комисије

Реферат комисије оцењује да су циљеви дисертације прецизно формулисани и да су доприноси остварени и верификовани; тај део оцене, у делу који се тиче пројектних циљева, концепта и осам доприноса, одговара стању у тексту, а спољашња верификација централног артефакта рецензираним радом [59] стварна је и вредна потпора.

Две ствари, међутим, ваља изричито означити. Прво, реферат помиње полазне хипотезе, а рад ниједну изричиту, нумерисану хипотезу не поставља — он поставља циљеве и радну претпоставку. То није доказ да је рад мањкав (истраживање кроз изградњу артефакта не мора почивати на нумерисаним хипотезама), али напетост између формулације реферата и стања у тексту ваља забележити, а у монографском издању уклонити тако што ће се хипотезе или изричито поставити или ће се доследно говорити о циљевима. Друго, оцена да „не постоји одговарајући систем…

ниједан нема реализоване основне захтеве“ стоји за четрнаест анализираних система, али почива на самодефинисаним захтевима (в. одељак V), што реферат не проблематизује. Најзад, оцена о обиму апарата тачна је.

XI Шта остаје

Кад се све сабере, добија се оцена коју вреди изрећи у једном комаду, без ублажавања и без претеривања.

Дисертација Дражена Драшковића солидан је и добро изведен инжењерски рад који домаћој образовној технологији и настави вештачке интелигенције доноси заокружен, вишеплатформски софтверски систем за симулацију и визуелизацију тридесет разнородних алгоритама, изведен из проверљиве везе преглед → мерила → захтеви → систем, с изворним техничким доприносом (сопствено исцртавање графова и стабала на Андроиду) и с разрађеним, рачунски тачним дидактичким примерима. Централни артефакт независно је потврђен рецензираним часописним радом, а тамо где су се техничке тврдње могле сравнити и поново прерачунати, у највећем делу држе. Уз то, рад показује научну честитост: слабије налазе не прећуткује, домашај доприноса формулише скромно и свесно покушава да изолује промену коју вреднује.

Слабости су, по оцени овог осврта, тројаке, и ниједна не руши инжењерско језгро рада. Прва и најважнија је вредновање у настави: пре/после поређење генерација без контролне групе, с истовременом променом структуре оцењивања, без мере расипања и величине ефекта, с непотпуно документованом анкетом чије просечне оцене нигде нису дате бројем — из чега су, ипак, изведени узрочни закључци о учинку и уопштавања с једног предмета на студенте и инжењере уопште. Друга је проверљивост и поновљивост: мерило избора и оцењивања четрнаест система није изричито, а систем, његов код, упитник, подаци и верзије алата нису доступни, док најављени развој вођен тестовима не прати ниједан документован тест. Трећа је презентација и апарат: нечитки и вишејезични снимци и легенде, погрешне унакрсне референце, датиран приказ појединих алата и ситне терминолошке и словне недоследности, уз једну бројчану омашку у проценту пролазности.

Изнад свега тога стоји једна правилност: рад је брижљивији у изградњи и опису система него у закључивању о његовом учинку. Управо зато највећи део онога што треба поправити не тражи ново истраживање него доследно спровођење онога што је рад већ започео.

Пет измена биле би, по оцени овог осврта, најплодније у евентуалном монографском издању или наставку истраживања. Прво, каузални језик вредновања заменити језиком повезаности доследне с увођењем система, изричито размотрити именоване претње ваљаности (селекција, инструментација) и известити о величини ефекта и

мери несигурности, а просечне оцене анкетних питања дати бројем и исправити упуту на табелу. Друго, ако је изводљиво, ојачати нацрт — паралелном или квазиекспериментaлном контролом, односно доследним оквиром студије случаја — и документовати упитник (састав, ваљаност, поузданост, апсолутан број испитаника). Треће, учинити систем, његов код, примере, упитник и податке јавно доступним, навести верзије кључних алата и приказати најављене тестове. Четврто, у прегледу постојећих система изричито изрећи мерило избора и операционализовати доделу оцена к1–к9, а систем који је остао без описа или описати или изоставити из табела. Пето, презентациони слој и апарат довести до уредности коју рад иначе показује — ћирилична и читка легенда и натписи, исправне унакрсне референце, ажуриран приказ алата, тачне бројчане вредности и уједначена терминологија — и алате на које се систем ослања унети у попис с верзијом и извором.

Уз ту врсту редакције — претежно уредничку, а не истраживачку — ова дисертација остаје вредан и употребљив допринос, чији би систем и модел развоја имали шта да кажу и настави вештачке интелигенције и стручном читаоцу изван уже теме.

Литература

Предмет осврта

Драшковић, Дражен Д., *Софтверски систем за учење и примену алгоритама вештачке интелигенције* [докторска дисертација]. Електротехнички факултет Универзитета у Београду, 2018. (НаРДуС.)

Реферат Комисије о урађеној докторској дисертацији Дражена Драшковића, Електротехнички факултет Универзитета у Београду, 20. јул 2018.

Емпиријски и квазиекспериментални нацрт

Shadish, William R.; Cook, Thomas D.; Campbell, Donald T., *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin, Boston, 2002.

Wohlin, Claes; Runeson, Per; Höst, Martin; Ohlsson, Magnus C.; Regnell, Björn; Wesslén, Anders, *Experimentation in Software Engineering*, Springer, Berlin, 2012.

Kitchenham, Barbara A.; Pfleeger, Shari L.; Pickard, Lesley M.; Jones, Peter W.; Hoaglin, David C.; El Emam, Khaled; Rosenberg, Jarrett, „Preliminary guidelines for empirical research in software engineering“, *IEEE Transactions on Software Engineering*, 28/8, 2002, 721–734.

Runeson, Per; Höst, Martin, „Guidelines for conducting and reporting case study research in software engineering“, *Empirical Software Engineering*, 14/2, 2009, 131–164.

Randolph, Justus J.; Julnes, George; Sutinen, Erkki; Lehman, Steve, „A Methodological Review of Computer Science Education Research“, *Journal of Information Technology Education: Research*, 7, 2008, 135–162.

Статистичко извештавање

Wilkinson, Leland; Task Force on Statistical Inference, „Statistical methods in psychology journals: Guidelines and explanations“, *American Psychologist*, 54/8, 1999, 594–604.

Wasserstein, Ronald L.; Lazar, Nicole A., „The ASA Statement on p-Values: Context, Process, and Purpose“, *The American Statistician*, 70/2, 2016, 129–133.

Cohen, Jacob, *Statistical Power Analysis for the Behavioral Sciences*, 2. изд., Lawrence Erlbaum Associates, Hillsdale, 1988.

APA Publications and Communications Board Working Group on Journal Article Reporting Standards, „Reporting Standards for Research in Psychology: Why Do We Need Them? What Might They Be?“, *American Psychologist*, 63/9, 2008, 839–851.

Анкета као инструмент

Kitchenham, Barbara A.; Pfleeger, Shari L., „Principles of survey research“ (делови 1–6), *ACM SIGSOFT Software Engineering Notes*, 2001–2003 (део 4: „Questionnaire evaluation“, 27/3, 2002, 20–24).

Kitchenham, Barbara A.; Pfleeger, Shari L., „Personal Opinion Surveys“, у: F. Shull, J. Singer, D. I. K. Sjøberg (ур.), *Guide to Advanced Empirical Software Engineering*, Springer, London, 2008.

Brooke, John, „SUS: A ‘quick and dirty’ usability scale“, у: P. W. Jordan и др. (ур.), *Usability Evaluation in Industry*, Taylor & Francis, London, 1996, 189–194.

Ефективност визуелизације алгоритама

Hundhausen, Christopher D.; Douglas, Sarah A.; Stasko, John T., „A Meta-Study of Algorithm Visualization Effectiveness“, *Journal of Visual Languages and Computing*, 13/3, 2002, 259–290.

Naps, Thomas L. и др., „Exploring the role of visualization and engagement in computer science education“, извештај радне групе са скупа ITiCSE 2002, *ACM SIGCSE Bulletin*, 35/2, 2003, 131–152.

Етика студија са студентима као испитаницима

Singer, Janice; Vinson, Norman G., „Ethical issues in empirical studies of software engineering“, *IEEE Transactions on Software Engineering*, 28/12, 2002, 1171–1180.

Инжењерски допринос, вредновање артефакта и поновљивост

Shaw, Mary, „Writing good software engineering research papers“, *Proc. 25th International Conference on Software Engineering (ICSE 2003)*, 2003, 726–736.

Hevner, Alan R.; March, Salvatore T.; Park, Jinsoo; Ram, Sudha, „Design Science in Information Systems Research“, *MIS Quarterly*, 28/1, 2004, 75–105.

Stodden, Victoria и др., „Enhancing reproducibility for computational methods“, *Science*, 354/6317, 2016, 1240–1241.

Sandve, Geir Kjetil; Nekrutenko, Anton; Taylor, James; Hovig, Eivind, „Ten Simple Rules for Reproducible Computational Research“, *PLoS Computational Biology*, 9/10, 2013, e1003285.

Smith, Arfon M.; Katz, Daniel S.; Niemeyer, Kyle E., „Software citation principles“, *PeerJ Computer Science*, 2, 2016, e86.

ACM, *Artifact Review and Badging*, верзија 1.0, Association for Computing Machinery, 2016.

Систематски и упоредни преглед

Kitchenham, Barbara; Charters, Stuart, „Guidelines for performing Systematic Literature Reviews in Software Engineering“, EBSE Technical Report EBSE-2007-01, Keele University & University of Durham, 2007.

Nikolić, Boško; Radivojević, Zaharije; Đorđević, Jovan; Milutinović, Veljko, „A Survey and Evaluation of Simulators Suitable for Teaching Courses in Computer Architecture and Organization“, *IEEE Transactions on Education*, 52/4, 2009, 449–458.

Изворни радови о алгоритмима

Forgy, Charles L., „Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem“, *Artificial Intelligence*, 19/1, 1982, 17–37.

Bylander, Tom, „The computational complexity of propositional STRIPS planning“, *Artificial Intelligence*, 69/1–2, 1994, 165–204.

Fikes, Richard E.; Nilsson, Nils J., „STRIPS: A new approach to the application of theorem proving to problem solving“, *Artificial Intelligence*, 2/3–4, 1971, 189–208.

Breiman, Leo; Friedman, Jerome H.; Olshen, Richard A.; Stone, Charles J., *Classification and Regression Trees*, Wadsworth, Belmont, 1984.

Deterding, Sebastian; Dixon, Dan; Khaled, Rilla; Nacke, Lennart, „From game design elements to gamefulness: defining ‘gamification’“, *Proc. 15th International Academic MindTrek Conference*, 2011, 9–15.

Russell, Stuart J.; Norvig, Peter, *Artificial Intelligence: A Modern Approach*, 3. изд., Pearson, 2009.

Научни апарат и стил цитирања

ISO 690:2010, *Information and documentation — Guidelines for bibliographic references and citations to information resources*, ISO, Geneva, 2010.

IEEE Editorial Style Manual, IEEE Publishing Operations, изд. 2017.

Језик

Пешикан, Митар; Јерковић, Јован; Пижурица, Мато, *Правопис српскога језика* (изм. и доп. изд.), Матица српска, Нови Сад, 2010.