

14. Справка за оригиналните научни приноси

Справката цитира публикациите на кандидата, представени за участие в конкурса и дадени в т. 10б на комплекта с документи. Копия на публикациите са в папка 16 Научни трудове, представени за участие в конкурса.

Приноси, отразени в публикациите за хабилитационен труд по показател В в точка 4 на таблицата за Минималните национални изисквания по чл.2 б от ЗРАСРБ

1. Предложен е нов подход за идентификация на структурата на данни в реално време [Б1, Б2, Б9, Б10]

Повечето методи за клъстеризация са разработени за данни, които са предварително натрупани. Едно от големите предизвикателства на съвременните системи, обаче, е необходимостта данните да се обработват в реално време. Затова разработването на методи за откриване на структурата на данните динамично т.е. във времето на получаването им, е от особен интерес. Основен недостатък на съществуващите решения за динамична клъстеризация е, че се базират на дефинирано разстояние – най-често Евклидово, което в общия случай затруднява откриването на реалната структура на данните.

Предложен е нов подход за клъстеризация, който представя решение в отговор и на двата проблема [Б9]. Той идентифицира структурата на данни в реално време, когато данните се получават в поток. Подходът се базира на Густафсон-Кесел (Gustafson-Kessel (GK)) алгоритъма, който е приложим само към фиксирано множество от данни. Разработена е оригинална процедура, сведена до приложим *on-line GK-like* алгоритъм за клъстеризация, основана на възможността на GK алгоритъма да открива групи от данни, чиято структура е общо дефинирана. Процедурата локално адаптира метриката на разстоянието към формата на клъстера на база на текущата оценка на ковариационната матрицата на клъстера и на матрицата на разделянето на данните. Алгоритъмът реализира рекурсивната актуализация на съществуващите клъстери и с това тяхната адаптация към получаваните нови данни. За целта той преизчислява параметрите им – клъстерен център и ковариационната матрица, прилагайки правилото на Кохонен.

Разработеният подход е верифициран, както за генерирани данни, така и за реални данни на сигнали от електроенцефалограма (ЕЕГ) [Б9]. Той е приложен към редица практически проблеми като анализ на ефективността на динамични системи, задачи за диагностика, прогнозиране и моделиране на данни от времеви редове, класификация в реално време, както и идентификация на системи и управление на процеси.

Подходът е разширен като е предложено едно ново решение за точно общо дефиниране на клъстерния радиус в зависимост от приета стойност за стандартното отклонение [Б1]. По този начин решението е независимо от стойностите на подаваните данни. Разработен е подход, сведен до алгоритмична процедура - *Evolving GK-like* (eGKL) алгоритъм за клъстеризация, която прилага концепцията на статистически дефинирано управление на процесите (Statistical Process Control (SPC)). Идентификацията на структурата на данните в реално време се реализира чрез рекурсивна процедура за оценка на средната стойност (клъстерния център) и на ковариационната матрица на базата на Метода за експоненциално претеглена променлива средна стойност. Рекурсивната оценка на стойността на обратната ковариационна матрица се изчислява чрез приложение на лемата на Уудбъри (Woodbury) за обръщане на матрица. Процедурата динамично открива клъстери с различна форма и ориентация в пространството като по този начин определя точно групите от данни и успешно решава проблема за оценката на броя на клъстерите. Специфицирани са ограниченията и параметрите при приложение на представената процедура.

Подходът, реализиран с *on-line GK-like* алгоритъма, е използван при формулиране на нова процедура за динамично вземане на решение в класификационна задача в случай на данни, получавани като поток, с отчитане на неопределеността на информацията [Б10]. Разстоянието между текущата входна стойност и най-близкия клъстер, както и броят на данните в близост до тази стойност, се дефинират като лингвистични променливи със съответни лингвистични стойности. Въведена е система от база размити правила, чрез която могат да бъдат оценявани слабо разграничими случаи на включване към съществуващ клъстер, класифициране като шум или създаване на нов клъстер.

Представеният подход за откриване на структурата на данни в реално време е приложен успешно при структурната идентификация на модел на динамиката на растеж на софтуерната надеждност в процеса на тестване на софтуерния продукт [Б2]. Научна новост е и предложеният модел на динамика на нарастване на софтуерната надеждност, представен с размити правила от вид Такаги-Сугено (Takagi-Sugeno). Това позволява едновременно да се използва както съществуващата неточна информация от размит вид,

така и точната информация за описвания процес. Структурата на модела се открива автоматично на базата на клъстерен анализ на наличните данни. Използването на процедурата за динамична идентификация на клъстерите [Б9] позволява да се опише специфично явление в процеса на тестване, когато добавянето на значително количество нов код при отстраняване на грешки води до промяна в динамиката на растеж на надеждността [Б2].

2. Предложен е нов метод за неитеративна клъстеризация на данни, който открива значими клъстери сред данни от шум [Б3, Б21]

Много често задача на анализа на данни е не да бъде разделян целия набор от данни в клъстери, а да бъде открит интересен модел или структура като конкретна група, която може да обхваща само (малка) част от пълното количество данни.

Едно решение на тази задача е предложеният нов метод за клъстеризация на данни, сведен до алгоритмична процедура, наречена Dynamic Data Assigning Assessment (DDAA) [Б3]. Това е евристичен метод на неитеративна клъстеризация, който открива значимите (добрите) клъстери и отделя данните от шум. Групирането на данните се реализира чрез определяне на центроида и оценяване на плътността на добрия клъстер, а данните от шум се отделят в друг клъстер – клъстер на шума (noise cluster). Значимите клъстери се разпознават един по един при всяко отделно изпълнение на процедурата. Съществено предимство на DDAA алгоритъма е възможността да се визуализират получените резултати от клъстеризацията независимо от размера на пространството на данните. Методът не изисква специална процедура за настройка на параметрите на алгоритъма. В зависимост от задачата алгоритъмът може да се използва за откриване на един или няколко интересни клъстера, които покриват само част от набора от данни или самостоятелно за цялостно структуриране на пространството от данни.

Оригиналният DDAA алгоритъм е адаптиран за целите на анализ на данни в реално време. Evolving DDAA (EDDAA) алгоритъмът реализира инкрементален подход за групиране, който изследва всяка получена нова стойност за принадлежност към добрия клъстер или към клъстера на шума [Б3]. Принадлежността може да се дефинира точно или в термините на размита (fuzzy) оценка. Предложената процедура позволява да се разпознава развиваща се структура от данни, в която се откриват следните възможни промени: промяна на съществуващи клъстери; дефиниране на нови клъстери; сливане на клъстери. Методът на DDAA показва добри способности за групиране на данни от

различни области. Чрез него са открити специфични групи в данни за характеристики на вино и в метеорологични данни.

Групирането на училищата в страната според резултатите на учениците им е реализирано чрез DDAА алгоритъма като са отделени две основни групи от училища – група с добри резултати и група с лоша успеваемост [Б21]. Чрез промяна на параметрите на DDAА алгоритъма са идентифицирани пет подгрупи на групата с добра успеваемост. При отчитане на допълнителна информация за резултатите на учениците са открити седем значими групи училища. Получените резултати са в помощ на разработването на добри образователни политики.

3. Предложен е нов подход за избор на софтуерна услуга на базата на сравнението на интегрирана (обобщена) оценка на качеството на софтуерни услуги, предоставящи еднакви функционалности [Б18, Б20]

Изборът на подходяща софтуерна услуга е актуална задача особено в контекста на дигитализацията на съвременния живот. Ежедневието изисква ползването на все повече и различни софтуерни услуги, което от своя страна е улеснено от развитието на съвременните облачни технологии. Основен проблем е необходимостта при избора да се отчита повече от едно качество на услугите при това да се има предвид възможния им противоречив ефект върху поведението на услугата. Изисква се също бърза и динамична оценка на качеството, адекватна на текущото състояние на потреблението на услугата.

Предложен е нов подход за избор на уеб услуги, базиран на клъстерния анализ на данните от метрики за качествени характеристики на услугите, натрупани в определено време на функциониране [Б18, Б20]. Идеята е, че всеки клъстер, дефиниран в такова пространство от данни на отделна услуга, характеризира определено ниво на интегрирано качество на услугата. Изборът на най-подходяща услуга от услуги с еднакви функционалности, се реализира чрез сравнение на параметрите на идентифицираните клъстери. Сравняват се координатите на клъстерните центрове на откритите най-плътни клъстери в данните на отделните услуги. Избран е подходящ метод за клъстеризация – Subtractive clustering, който е адаптиран за реализацията на подхода. Подходът е сведен до приложима процедура за надежден и динамичен избор на услуга [Б18]. С цел по-информативно и точно сравнение на услугите са подбрани два допълнителни специфични индекса за определяне на компактността на клъстерите като процедурата за избор е разширена със сравнение и по тези стойности [Б20].

**Приноси, отразени в публикациите по показател Г в точка 7 на таблицата за
Минималните национални изисквания по чл.2 б от ЗРАСРБ**

4. Предложен е подход за разпознаване на състояния при хората чрез анализ на данни от активността на мозъка [Б4, Б5, Б11, Б13]

Последните интензивни изследвания на данните от мозъка като електроенцефалография (ЕЕГ) и данни от визуалните възприятия се налагат от търсенето на решения за подобряване на връзката човек-компютър, които да могат да се вложат в нови устройства, подпомагащи хората в ежедневието. Голямото предизвикателство е да се намерят и приложат подходящи алгоритми за обработка и извличане на нужната информация от данните на мозъчната дейност. В голямата част от публикуваните системи за разпознаване на тези сигнали се използват методи на супервайзорно обучение (обучение с учител), докато несупервайзорните (обучение без учител) са слабо изследвани.

Представен е подход за разпознаване на човешки емоции чрез анализ на ЕЕГ данни, получени при гледане на силно положително и силно отрицателно емоционално натоварени картини [Б4, Б13]. Научна новост е изследването на възможността на стандартните алгоритми за клъстеризация да се справят със задачата. Като задължителен етап на представения подход е изборът на подходящо пространство на данните. Предложена е рамка за предварителен избор на информативно пространство от данни с малък размер, базирана на детайлно проучване, изследване и анализ на различни по характер и изпълнение алгоритми за избор на характеристики. Вторият етап на подхода изисква приложение на ефективни алгоритми за разделяне. Показани са резултати и възможности на клъстеризация с методи на основата на минимизиране на критерий на разделянето. В [Б5] аргументирано е избран филтърен метод за редуциране на пространството от данни, върху което са приложени шест известни ефективни алгоритъма за клъстеризация (K-means, X-means, Farthest First, fuzzy-Cmeans, Expectation–Maximization and Hierarchical clustering) и са оценени по отношение на възможностите им да разграничават двата емоционални стимула.

Изследванията обхващат двете основни задачи на разпознаване - за отделни субекти, така и клъстеризация между субекти. Разграничаването на данни от различни субекти е трудната задача предвид променливостта на данните за отделните индивиди [Б4, Б5, Б13]. Резултатите ясно показват по-добра на дискриминация на състоянието на положителна емоция в сравнение с това на отрицателната. Основен резултат на анализа е предписание на начина, по който се подбират характеристиките. Комбинацията от характеристики на

малък брой времеви области (амплитуди и латентности) и пространствени (избрани канали) може да намали ефекта на променливост на данните от различни субекти и по този начин да подобри дискриминационните резултати на моделите за множество субекти [Б5].

Подходът на клъстерен анализ иновативно е приложен като аналитична стратегия за оценка на възможността за разграничаване на възрастовите от индивидуалните различия в обработката на визуална информация според данни, получени от психофизични експерименти за дискриминация по посоката на движение [Б11]. Изследванията са и във връзка с определянето на по-подходящите условия за оценка на тези различия. Показано е, че експеримент с най-голям брой движещи се елементи с различна скорост е най-дискриминативен. Интерпретацията на резултатите от групирането позволява да се открие нивото на отклонение на индивид от съответната възрастова група, което предполага възможности за използване на психофизични методи за ранна диагностика на влошаването на когнитивните способности на индивида с възрастта.

5. Разработен е размит лингвистичен модел за оценка на софтуерната надеждност [Б12]

Надеждността на софтуера е нефункционална характеристика от особена важност, защото дефинира способността на системата да предоставя услугите, за които е предназначена. Обикновено надеждността се дефинира чрез статистически анализ и вероятностни модели. Тези модели, обаче, се базират на предположения, което води до неточност в оценките на надеждността.

С цел да бъде отразена неточността в наличните данни е предложен нов модел за оценка на надеждността на софтуер с използване на теорията на размити множества и логика [Б12]. Аргументирано са избрани и са представени като лингвистични променливи характеристики, които са съществени за оценката на надеждността - брой на грешките, време и осреднено време за поява на грешка. Конструирани са база размити правила като модел за изчисляване на оценката. Реализирано е композиционно правило за извод на Мамдани като изчислителна процедура за получаване на оценка на времето за поява на грешка при подадена информация за стойностите на входните променливи. Точността на модела е илюстрирана върху достъпен набор от данни (benchmark case) за надеждност на софтуер. Стойностите, получени чрез модела, са съпоставими с получените от широко разпространения вероятностен модел на Джелински и Моранда. Размитият модел е точен в предсказването на надеждността в по-късната фаза на процеса на тестване.

6. Предложен е метод за избор на софтуерна услуга въз основа на определени нейни качествени характеристики, който отчита неточността в информацията за качеството на услугите [B7, B14, B15, B16, B17]

Качествените характеристики на софтуера са от основен интерес за голям брой области на софтуерното инженерство, свързани с различни приложения като например вградени и критични системи, уеб приложения. Системите, ориентирани към услуги, са основна парадигма на съвременните технологии. Изпълнението на основната цел да се доставят услуги, на които може да се вярва, е затруднено. Първо, поведението на уеб услугите по отношение на показаното качество е трудно да се предвиди, защото зависи от множество фактори като мрежовата връзка или съответния сървър, броя на едновременните извиквания, характеристиките на инфраструктурата на доставчиците. От друга страна, изискванията за качество на услугите от страна на клиентите са индивидуални и често неясно дефинирани.

Предложеният нов метод за избор на софтуерна услуга въз основа на определени нейни качествени характеристики въвежда концепция за оценка на качеството на софтуерните услуги, която комбинира два типа неопределеност на съществуващата информация, свързана с избора [B14, B16]. С цел да се отчете различния произход на неточността в данните и информацията методът предоставя рамка за формализация, която използва данни от метрики на качествени характеристики, натрупани по време на наблюдение на услугата, както и експертна информация за качеството ѝ. Използвана е теорията на вероятностите, за да се опише проявата на определени характеристики на дадената услуга, а теорията на размитите множества се прилага като теоретична база за описание на несигурността в изискванията на клиентите. Представен е формален начин, сведен до приложима процедура, за тяхното общо отчитане с цел получаване на надеждна оценка [B14]. За осигуряване на по-точен избор количественото измерение на едновременното изпълнение на показаното качество и исканото от клиента е обогатено с допълнителен индекс [B16]. Практическата приложимост на метода е показана по два начина. Реализиран е избор за казус с уеб услуги за проверка на електронен адрес по отношение на времето за отговор [B16]. Освен това методът е автоматизиран чрез специално разработен софтуерен инструмент RecSS, валидиран за пет услуги за изпращане на кратки съобщения, сравнени по отношение на осреднено време за отговор [B15].

Предложеният метод е разширен за избор според интегрирано на качество на уеб услуги, който отговаря на практическата ситуация, когато изборът трябва да се направи въз основа на повече от едно свойство на услугата [B17]. Представен е механизъм за

изчисляване на нивото на съвместна проява на няколко качествени характеристики на дадена услуга. Чрез приложението му към данните от всяка интересуваща ни уеб услуга, могат да се сравнят получените нива на интегрирано качество, за да се посъветва клиента за услугата с най-добро качество. Разширеният метод е обобщен в алгоритъм, проверен чрез реални данни за уеб услуги за проверка на електронен адрес, оценени по три качествени характеристики.

Публикация [Б7] дава едно пълно, последователно и обобщено представяне на метода за избор на уеб услуга според качествените характеристики. Подробно е представена теоретичната основа за отчитане на различната по характер неопределеност при решение на задачата. Представени са механизмите за избор в случай на оценка по едно качество и избор при интегрирано качество на услугите. Оценката на базата на интегрирано качество е допълнен с механизъм за оценка при сценарий, когато субективната оценка на клиента за искано качество също трябва да бъде отчетена.

7. Разработен е модел за разпознаване на стила на игра в образователни игри чрез линейна регресия [Б8]

Моделите, ориентирани към играчите, позволяват на дизайнерите на електронни игри да адаптират играта към специфичния стил на игра и така увеличат удовлетвореността от играта, а от там и ефикасността на играта. При образователните игри адаптацията на играта според стила на обучение води до по-добри резултати от обучението. Основният проблем е да се приложи подходящ подход за анализ на поведението на играча по време на играта и да се предостави начин за динамично приспособяване на играта към стила на игра.

Разработеният модел за адаптивна видео игра реализира модел за автоматично разпознаване на стилове на игра въз основа на множествени линейни регресии, използвайки показатели за ефективност на изпълнение на игровите задачи [Б8]. За разлика от известните подходи за прилагана линейна регресия, при които оценка на влиянието на играча е направена на базата на проявени физиологични характеристики, предложеният модел използва характеристики за стила на игра. Аргументирано са подбрани ключови показатели за изпълнение на игровите задачи, съответстващи на всеки един стил. Моделът разпознава четири стила на игра (състезател, мечтател, логик и стратег по теорията на Колб за експериментално обучение) по време на играта. Идентификацията на линейните регресии е реализирана по два начина – евристично - чрез структурирани интервюта и

формално - чрез метода на най-малките квадрати. Анализирани и обяснени са ефектите на адаптиране на афективната игра според показателите за изпълнение на игровите задачи и в двата случая на определяне на коефициентите на моделите. За всички стилове на игра регресионните коефициенти, изчислени чрез метода на най-малките квадрати, водят до по-голяма точност на модела. Разпознатите стилове на игра и/или учене чрез моделите са подходящи за реализиране на стилово приспособяване не само на дидактическо съдържание в образователните, но за адаптиране на различни характеристики на механиката на всяка приложна игра.

8. Разработен е невронен модел на определяне на стойностите на консумацията на електрическа енергия на електродъгова пещ в зависимост от химическия състав на заредената смес [Б6]

Производството на неръждаема стомана е сред най-енергоемките индустрии. Търсят се решения за икономия на енергия, които имат както икономически, така и значим екологичен ефект.

Предложен е модел, чрез който с голяма точност се изчислява какво е потреблението на електрическа енергия в зависимост от химическия състав на сместа на заредения материал [Б6]. За разлика от съществуващите решения моделът използва feedforward изкуствената невронна мрежа - многослоен персептрон (MLP). Предложеното решение е с трислойна архитектура, чиято оптимална структура с хиперболична тангента в скрития слой и линейна функция в изходния слой е получена след 89 цикъла на обучение. Постигната е точност на описание е такава, че 91% от вариациите в специфичната консумация на електрическа енергия могат да бъдат описани с помощта на предложения модел. Детайлният анализ на взаимните зависимости на отделните елементи на смесите показва, че най-голямо влияние върху консумацията на електрическа енергия има съдържанието на въглерод. Моделът може да се използва при изработване на стратегия за снижаване на разходите за електрическа енергия чрез избор на оптимална смес за зареждане.

9. Разработен е регресионен модел на за разпознаване на загубата на внимание [Б19]

При редица професии от особен интерес са внедряването на ефективни системи за управление на вниманието, които чрез психофизиологични индикатори да следят за

концентрацията на оператора и да прогнозира и алармира за моменти с ниска бдителност и пропуски във вниманието. Изграждането на приложима индивидуализирана система за наблюдение на състоянието на вниманието на субектите е предизвикателство, главно защото основният източник на данни – данни от сигналите на електроенцефалограма (ЕЕГ), са замърсени с артефакти и инструментални шумове.

За първи път е приложен регресионен подход за откриване на променливи нива на внимание, базиран на пространствено-времеви данни, извлечени от ЕЕГ записи [Б19]. Подбрани са характеристики на сигналите в специфични честотни диапазони и времеви прозорци като потенциални кандидати за дискриминация на нивата на внимание, свързани с появата на пропуски на вниманието, които са по-малко засегнати от артефакти и шум. Получени са надеждни модели на корелацията между измерената невронна активност през интервалите на стимулиране - преди (500 ms) и след (1000 ms) стимула и регистрираните времена на реакция. За целите на анализа са приложени три стандартни регресионни модела - линейна регресия (LR), изкуствени невронни мрежи (ANN) и машина на поддържащите вектори (SVM). Трите модела показват подобна възможност за описание на зависимостта, което е малко повлияно от различията в избраните характеристика на данните, получени при прилагане на Principal Component Analysis. Резултатът е обещаващ, като се има предвид, че осреднената квадратична грешка (RMSE) от около 14% е получена за различни сценарии и набори от характеристики. Показано е, че най-голяма точност се постига с ANN модел, но и LR проявява приемливи и близки по точност показатели на грешката. Разлики във времената на реакция на лявата ръка и дясната ръка на отделен субект налагат да се изградят различни регресионни модели за двете ръце, които обаче по точност те не се различават значително. Сценарият на дясната ръка изглежда малко по-подходящ за моделиране, предвид естествения ефект за хората, пишещи с дясна ръка. Откритата силна корелация между относителното намаляване на амплитудата на ЕЕГ сигнала след стимула (ERD) и времето за реакция също е нов резултат, който предполага по-нататъшно изследване на характеристиките на ERD не само при двигателни задачи, но и при когнитивни състояния и колебания.

26 Август, 2019 г.

Подпис:

/доц. Олга Георгиева/