

РЕЦЕНЗИЯ

на дисертацията " Автономни обучаващи се системи от потоци данни в реално време", представена от проф. д-р инж. Пламен Първанов Ангелов за присъждане на научната степен " доктор на науките" по професионалното направление 4.6 "Информационни и компютърни науки"

Рецензент : Академик Минчо Хаджийски

Въведение

Рецензирането е съгласно заповед 118 / 13.03.2015 г. на Директора на ИБФБМИ-БАН. Предварителен вариант на дисертацията е обсъден на разширен НС на ИБФБМИ (Протокол № 2 /11.03.2015 г.). В материалите за рецензиране са включени всички изискващи се от ЗРАСРБ документи. Дисертацията е представена на английски език. Авторефератът на дисертацията е на български език. Документите по процедурата са представени в съответствие с българското законодателство, съответните подзаконови разпоредби и специфичните правила на представящото кандидата самостоятелно научно звено на БАН - ИБФБМИ.

Проф. д-р Пламен Ангелов е роден през 1966 г. През 1989 г. завършва магистратура в ТУ - София като инженер по "Електроника и автоматика". В периода 1990 - 1993 г. е докторант в ЦЛБА - БАН и защитава успешно дисертация за получаване на научната степен "КТН" през 1993 г. През 1993 - 1999 г. е научен сътрудник I ст. в БАН. В периода след 1999 г. заминава в Англия, където е последователно научен сътрудник (1999 - 2002) и главен асистент (2003 - 2006). През 2007 г. е избран за доцент, а през 2013 г. - за професор по Интелигентни системи в Университета в Ланкастър, Англия.

Дисертацията представлява модифициран вариант на монографията на кандидата " Autonomous Learning Systems : From Data Streams to Knowledge in Real Time" , издадена от престижното издателство John Wiley през 2012 г. Дисертацията представлява обобщение на 46 подбрани от автора публикации с най- много цитирания в последните 15 години. Поради приетия характер на представяне на резултатите с различна степен на детайлизация в самата дисертация и публикациите към нея, рецензентът ще ги разглежда като взаимно допълващи се части на единно цяло. От приложения за сведение списък на всички 222 трудове на кандидата се вижда, че към темата на дисертацията могат да се отнесат и значителен брой други публикации (над 80 бр.). Това ясно показва значителния обем резултати, достигнали и оценени от специалистите в третираната област. Съгласно приложения списък, в трудовете, отнесени към представената дисертация, не са включени тези 10 публикации, които са използвани в дисертацията му за получаване на научната степен "д-р" . Няма сведения за независимо използвани публикации в сравнение с процедурата на кандидата за професор през 2013 г. Но от общия списък на публикациите не включени в настоящата процедура, се вижда, че има общо 49 публикации след 2012 г. (15 статии, 25 доклада и участие и редактиране на 9 книги). Основната част от тях са тясно свързани с тематиката на дисертацията, представена в настоящата процедура. Това показва категорично, че представената дисертация може да се разглежда (съвместно със значителния брой публикации, неизползвани в предварителни процедури) като самостоятелен научен труд, съгласно българското законодателство (ЗРАСРБ).

Обща характеристика на дисертационния труд

Дисертацията е изложена на 307 стр. Основното съдържание е на 256 стр. Включени са още: списък публикации по дисертацията (3 стр.), библиография (11 стр.),

индекси и означения (21 стр.), приложения (16 стр.), от които най-съществени са псевдо-кодовете на основните оригинални достижения на автора: AutoCluster, ELM, AutoSense, AutoClassify0 и AutoClassify1.

Материалът за дисертацията е структуриран в 15 глави, разпределени в три части- Основни положения (гл.1 - 4), Методология на автономните обучаващи се системи (гл. 5 - 10), Приложения (гл. 11 - 14) и Заключение (гл. 15).

Дисертацията е построена върху единна методическа основа - използване на рекурсивно оценявана текуща относителна плътност на размити облаци от данни на локално и глобално ниво. На тази основа са разработени варианти на основни алгоритмични процедури, необходими при наблюдение, оценка на параметри и състояния, предсказване, вземане на решение и управление на сложни системи от пределно общ вид: динамични, нестационарни, нелинейни, неопределени.

Работата е ориентирана изцяло към изчислителни операции върху поток от данни, като кандидатът си е поставил амбициозната задача получените системи да придобият следните важни качества в рамките на една единна процедура на проектиране:

- Автономни.
- Развиващи се.
- Самообучаващи се.

С решаване на тази комплексна конструктивна задача кандидатът се стреми да реализира следните цели:

- Системата да не се нуждае от априорна информация за обекта и входните данни.

- Да работи в реално време.
- Да бъде изчислително по-ефективна от съществуващите.

За достигане на поставените цели и задачи кандидатът е работил на три нива на абстрактност:

- **Първо ниво** - Развитие на новия подход за представяне на размити потоци и множества от данни с нов модел, основаващ се на локална и глобална плътност на данните и обособяване на облаци от размити данни или (ако е по-подходящо) на клъстери на същата основа.

- **Второ ниво** - Използване на предложения достатъчно общ подход за разработване на няколко функционално насочени процедури, които продължават да имат достатъчно общ характер:

- Елементи за оценяване на параметри на моделни структури.
- Филтри.
- Предсказващи елементи.
- Сензори за оценяване на променливи по косвени показатели.
- Класификатори.
- Системи за управление.

- **Трето ниво** - Приложение на разработените методи, техники и елементи към конкретни обекти:

- Софтсенсинг в нефтохимическата индустрия.
- Самообучаващи се мобилни роботи.
- Интелигентни системи за обработка на видеоинформация (подвижни обекти, потребители на Unix).

Оставайки все още изцяло в полето на размитите системи, авторът се стреми значително да разшири възможните приложения за случаи на резки изменения на операционните ситуации, такива като- промяна на оптимизационната повърхност поради изменения в състава на физико-химическите условия; непредвидени значителни изменения на входните въздействия поради повреди в съседни подсистеми

или резки принудени диспечерски решения; повреда в собственото технологично оборудване; груби операторски грешки и др. В тези ситуации конвенционалните размити системи за мониторинг, вземане на решение и управление не могат да се справят задоволително. При решаване на тези проблеми, възникващи в реалния живот, проф. П. Ангелов е получил голям брой убедителни научни и научно- приложни резултати.

Обзор на получените резултати и автореферат

Този обзор е направен от кандидата в автореферата към дисертацията и рецензентът го приема като достатъчно пълен и обективен. От материала на дисертацията в автореферата са намерили място преди всичко водещите принципи на кандидата, основните му постижения и личните му оценки за тяхната новост, значимост и приложимост. Авторефератът представлява обобщено резюме с акценти, а не редуциран вариант на дисертацията. Номерацията на главите до гл. 10 е запазена, но заглавията им са изменени. Третата част на дисертацията (гл. 11 - гл. 14) е представена с един единствен раздел в автореферата (гл. 11), като е запазена номерацията на използваните фигури и формули от основния текст. Авторефератът е в обем 75 стр., от които 17 стр. са лични данни. Като цяло авторефератът отразява пълно основното съдържание на дисертацията.

Приноси в дисертационния труд

Основните достижения на дисертацията могат да се обобщят така:

1. Предлагат се системи, изцяло базирани на данни, като не се предполагат никакви предварителни допускания относно структурата на данните и задаване на ограничения (брой, форма и параметри на функции на принадлежност, прагови стойности, брой и граници на клъстери).

2. Приета е непараметрична форма на описание на облаци от данни, чиято плътност се изчислява рекурсивно. Това е позволило да се предложи опростена размита структура, базирана на правила.

3. Системите са автономни и самообучаващи се, което позволява да се адаптират към текущата ситуация чрез адекватна промяна на параметри и структури, извличане на нови правила, отчитане на "възрастта" на всеки елемент от множеството размити облаци от данни.

4. В разработката не се налага да се спазват задължителните предпоставки на вероятностния подход- стационарност, ергодичност, достатъчна дължина на извадките. Комбинира се традиционната честотна вероятност с пространственото разпределение на данните чрез въвеждане на метрики за разстояние.

5. Алгоритмичните опростявания правят предложения подход изчислително ефективен и приложим за опериране в реално време.

Конкретните научни, научно- приложни и приложни достижения се изразяват в следното :

Научни приноси

1. Предложен е нов тип размита система A_nY , в която условната част на логическото отношение "Ако.....То" е непараметрична. Прилага се концепцията за размити облаци от данни и относителна плътност на данните. Облаците от данни са подмножество на всички предхождащи измерени стойности, близки помежду си в пространството на данните. Размитостта се запазва, защото всяко ново измерване принадлежи към облаците от данни с определена степен, но се избягва явното задаване на броя и параметрите на функциите на принадлежност за всяка входна променлива, необходими в класическите размити системи от типа на Мамдани или Такаги-Сугено. По този начин отпада субективността и предварителния опит като критично условие за успех при моделиране на системата.

2. Развитият подход за рекурсивно изчисляване на локалната плътност на поток от данни в облак. Това позволява бързо текущо определяне както на локалната, така и на глобалната плътност. На тази база могат да се описват нелинейни, нестационарни и недетерминистични системи само с използване на измерванията на входни и изходни променливи. Този нов принцип е защитен с два патента на кандидата..

3. Разработеният от автора подход за динамично развиващи се размити системи, при което са получени следните оригинални резултати:

- Нов метод за динамично и в реално време клъстеризиране. Това позволява множество от клъстери да се изгражда гъвкаво по хода на развитие на потока от данни. Не се изисква предварителна информация за броя, формата и размерите на клъстерите, които да зависят от конкретния случай или експертен опит.

- Създаденият еволюиращ нелинеен размит класификатор, който непрекъснато се адаптира по хода на постъпващите данни. Адаптацията обхваща както промяна в параметрите на класификатора, така и (възможно) неговата структура. За разлика от конвенционалните размити класификатори необходимите изисквания и ограничения (граници) се формират динамично и зависят от последователността на възникващите ситуации.

- Предложеният саморазвиващ се регулатор, който може да стартира дейността си без предварителна информация. По хода на подобрението на функционирането си той може да изменя както структурата си (напр. P, PI, PID), така и да оптимизира параметрите си на настройка автономно.

Научно- приложни приноси

Тук следва да се отнесат конкретно разработените градивни елементи за развиващи се системи, основаващи се на посочените по- горе нови подходи:

1. Методи и алгоритми за развиващ се (еволюиращ) клъстеринг : eClustering, AutoCluster, ELM, DEC, DDCAR.

2. Методи и алгоритми за развиващи се (еволюиращи) класификатори: AutoClassify 0, AutoClassify1, eClass.

3. Методи и алгоритми за мониторинг и диагностика:

- Сензор за оценка на състояния по косвени показатели - AutoSense.

- Методи за детектиране и изолиране на неизправности (два патента в САЩ).

4. Разработените елементи за предсказване, оценяване и косвено наблюдение на колективно действащи автономни подсистеми.

Приложни приноси

Разработените нови подходи, методи и алгоритми са верифицирани в четири класа реални приложения:

1. Завод от нефтохимическата индустрия- автономен самообучаващ се софтсенсинг (Autosense), промяна на структура, избор на входни променливи.

2. Автономно обучение на мобилни роботи- автономен класификатор AutoCluster, AutoClassify 0, AutoControl.

3. Автономно откриване на изменения и придвижване на обекти на видеоизображение - RDE.

4. Моделиране на еволюиращото поведение на ползвател - AutoClassify 0, AutoClassify1.

Публикации

Неразделна част от дисертационния труд на проф. Пламен Ангелов са представените 46 публикации, основните данни за които са систематизирани в приложената Таблица 1.

Таблица 1

	Монографии	Статии	Доклади	Глави от книги	Патенти	Общо
Брой	2	20	19	3	2	46
Брой цитирания	235	1599	422	232	-	2488
Импакт фактор	-	55,57	-	-	-	55,57
Издад.труд след 2012	-	7	8	1	2	15

Представените публикации са в престижни международни списания, за което говори сумарния импакт фактор 55,57. Статиите на кандидата са многократно цитирани: работа А 22 - 610 пъти, А 18 - 174 пъти, А 20 - 123 пъти, А 24 - 102 пъти. Отнесените към дисертацията статии са цитирани общо 1599 пъти и това представлява основната част (64 %) от общия брой 2488 цитирания на трудовете по дисертацията. Значителен е и броя на цитиранията на докладите (общо 422 пъти) , като А 40 е цитиран 142 пъти. Монографиите и главите от книги също са многократно цитирани (общо 467 пъти), като А 2 е цитирана 210 пъти, а А 46 - 176 пъти.

Кандидатът продължава активната си публикационна дейност и след удостоверяването му със звание "професор"- за три години е публикувал 15 труда в значими международни списания и научни форуми, което надхвърля с почти 50 % средната му годишна продукция в годините след получаване на научната степен "ктн" (д-р). Общият брой на цитиранията му е над 3860 .

Публикационната дейност и цитиранията на трудовете на кандидата, свързани с дисертацията, са на изключително високо ниво и свидетелстват за големия интерес, който се проявява по света към неговите изследвания.

Възможни разширения

Предложените в дисертацията подходи могат да бъдат приложени не само върху потоци данни, където се третира дискретни последователности от точки по време, но и върху съвкупности от точки, получени в различно време.

При достатъчно обща дефиниция на обекта за разглеждане, развитият в дисертацията непараметричен подход би могъл да намери приложение при автоматично изграждане на динамични онтологии.

Значителен интерес би представявало допълнителното разработване на установената в дисертацията връзка между предлагания рекурсивен подход и правилото на Бейс за априорна и апостериорна вероятности. Това би позволило след съответна модификация, разширение и навярно въвеждане на нови съотношения, да се използва съществуващото теоретично богатство на Бейсовските подходи.

Голямата популярност, която имат в момента различните направления на съвременен статистически анализ (advanced analytics), могат да се допълнят с алтернативни методи, основаващи се на предложените в дисертацията рекурсивни подходи и процедури. Предлаганите подходи, изискващи минимална или никаква предварителна информация, съдържат значителен потенциал за анализ, вземане на решение и управление на сложни неструктурирани системи с динамично променящи се структури, функции и параметри, такива като икономически системи, оперативно управление, военни действия, неврофизиологични системи, системи в предаварийно и аварийно състояние и мн. др.

Освен конкретните приноси бих желал да посоча някои достойнства на дисертацията на проф. П. Ангелов от по-общ характер:

- Дисертацията е подходящо структурирана, написана е ясно и в добър стил, лесно се чете, независимо от използвания обширен теоретичен и инструментален апарат, за което допринасят приложените списъци с термини и съкращения, както и приложенията.

- Историческите препратки са ограничени основно до хоризонт от около 20 години.

- Използва се единен подход, базиран на рекурсивно оценяване на плътността на потока от данни, който после се конкретизира в посоки клъстеризация, оценка на параметри, класификация, управление. Отделните глави и раздели са взаимзствани и това допринася за възприемане на дисертацията като цялостно изследване.

- Дисертацията отразява съвременното състояние на разработките в областта на автономните самообучаващи се системи, където кандидатът е между водещите автори в света. Личните му постижения са ясно конкретизирани.

- Приложенията са неелементарни и са достатъчно представителни.

Критични бележки

Рецензентът има следните забележки към материала, изложен в дисертацията :

1. В дисертацията са поставени значителен брой амбициозни цели и задачи с голяма сложност, които са решени с различна степен на пълнота. Голям интерес би представлявало задълбочаването на изследванията по някои от следните проблеми:

- Установяване на условията за устойчивост на редица от предложените в дисертацията процедури : оценка на параметри при зададена структура; адаптация в системите за управление с модел с опростена размитост от предлагания тип; координирано управление на многоагентни системи.

- Гарантиране на липса на колебания / робастност при превключване от една към друга структура, проблем, оказал се трудно преодолим за динамични експертни системи.

- Конкретизация на възможността за използване на предложения подход за описание на системи за управление в пространството на състоянията, която в дисертацията е само маркирана.

- Изследване на влиянието на нивото и свойствата на шума във входните данни върху ефективността на предложените on - line процедури.

2. Липса на цитиране и подробен анализ на предшестващи научни резултати по клъстеризация на базата на плътност на многодименсионни потоци от данни, появили се във втората половина на 90^{-те} години на миналия век, напр. DBSCAN (и неговите разширения) -[Anke 99,Sand 98, Bohm 00, Berk 02] и аналогично DBCLASD [Xu 98], DENCLUE [Hinn 98] и др.

3. В дисертацията се прави допускане за наличие на достатъчно подобие на данните, формиращи облаците от данни (или клъстерите, ако това е подходящо за дадената задача). Остава неясно как предложените подходи се справят със случаите, когато само малък брой признаци спомагат за идентификация на клъстерите и когато различни клъстери пребивават в различни подпространства (след редукционни процедури с PCA или SVD) на оригиналното пространство на признаците.

4. Обучаемостта на предложените процедури е третирана основно във връзка с изчислителните перспективи, но не и от гледна точка на необходимото време и наличие на достатъчно данни за вземане на адекватно решение при бързи значителни изменения в свойствата на входните въздействия или на самия обект (напр. авария).

5. При силно нестационарни потоци от данни проблемът със стареенето на информацията става критичен за адаптационните процедури. Поради това "хоризонтът

на забравяне" следва да се адаптира в рамките на същата процедура, с оценка на текущата плътност на данните.

Направените бележки в никаква степен не засягат основното съдържание на дисертацията и приносите на кандидата. Те биха могли да се окажат полезни при бъдещата му работа в ясно дефинираната от него област, получила одобрение и популярност от много изследователи по света, за което може да се съди по внушителния брой цитирания.

Лични впечатления

Познавам проф. П. Ангелов от 20 години във връзка с участието му в национални и международни конференции, провеждани в България, свързани с размито управление и управление на биотехнологични процеси. В следващите години той бе постоянен участник в иницирираната от български специалисти голяма международна конференция "Интелигентни системи " на IEEE, където е бил пленарен докладчик и организатор на поканени сесии. Съдейства за провеждането на конференции на IFAC в България, като на предстоящия през м. септември т.г. симпозиум на IFAC "TECIS'15" ще изнесе пленарен доклад.

Познавам изключителна активната му научно- организационна дейност в чужбина. Само в последните три години той е бил 18 пъти член на международни програмни комитети на големи научни конференции. Понастоящем заема отговорни постове в дружеството SMC на IEEE.. Той е главен редактор на основаната от него списание "Evolving Systems", което се превърна във важен център за обединяване на работещите в областта на развиващите се системи от различен тип.

Заклучение

Дисертацията "Автономни обучаващи се системи от потоци данни в реално време" е цялостно мащабно изследване, в което са получени значителни научни, научно- приложни и приложни резултати в твърде актуална научно- техническа област. Иновативният характер на изсследванията е получил широко признание в международната общественост, за което свидетелстват многобройните цитирания на публикациите на автора. Изследванията са в период на интензивно развитие , като в тях са приобщени научни колективи и отделни изследователи от много страни. Кандидатът е изтъкнат международно утвърден учен с респектираща научно- организационна дейност и с активни връзки с българските си колеги.

Въз основа на изложените по- горе аргументи убедено препоръчвам на уважаемото Жури по процедурата да присъди на **проф. д-р Пламен Първанов Ангелов** научната степен "Доктор на науките" по професионалното направление 4.6 "Информационни и компютърни науки".

26. 05. 2015 г.

Рецензент:

София

Академик Минчо Хаджийски