

Bírálói vélemény

Abonyi János:

„Technológiafejlesztési célú adatbányászati technikák” c.

doktori értekezéséről

A jelen bírálat tárgyát képező értekezését az „MTA doktora” cím megszerzése céljából Abonyi János 2000-ben megvédett, fuzzy modell alapú szabályozó algoritmusok fejlesztésével foglalkozó PhD disszertációjának elkészülte után jó 10 évvel nyújtotta be. Ekkora időtartam nyilván bőségesen elegendő a PhD értekezés témakörét, tárgyát meghaladó, új érdemi tudományos eredmények létrehozására és összefoglalására.

A PhD disszertáció lezárását követően a Jelölt figyelme a technológiai adatok kezelésének és adatbányászati algoritmusok célirányos fejlesztésének irányába fordult. A vizsgált tématerületek közül a dolgozat négy, némileg különálló, de a technológiafejlesztést egyaránt támogató témakörre fókuszál. A bírálat a továbbiakban csak az értekezésben közreadott eredményekre korlátozódik. Itt szeretném megjegyezni azonban, hogy a Jelölt átlagosnál messze magasabb aktivitása és munkabírása miatt az értekezés az azóta elért új eredményeinek csupán egy részét tartalmazza.

Az értekezés témaválasztásáról

A Jelölt munkahelyén, a Pannon Egyetem Folyamatmérnöki Tanszékén több mint három évtizede foglalkoznak vegyipari rendszerek modellezésével, modell-alapú szabályozásával és tervezésével. Ahogy a dolgozat bevezetése is rámutat, az információátvitel a különböző jellegű modellek/fejlesztési feladatok között e tekintetben kivétel nélkül jelentős sikertényező, azaz a meglévő és a potenciálisan elérhető információk hatékony kezelése kulcsfontossággal bír.

Abonyi János e műhely tagjaként a számítási intelligencia technikák (fuzzy modellek, neurális hálózatok, genetikus algoritmusok) folyamatmérnöki problémákban történő alkalmazhatóságát vizsgálja. Ezen témaválasztás rendkívül aktuális és szerencsés voltát minden kétséget kizáróan bizonyítja, hogy e munka eredményeként nem csupán új, jól dokumentált, modellalkotást, szabályozó tervezést és technológia-optimalizálást támogató alkalmazások jönnek létre, hanem a számítási intelligencia témakörében kutatók számára érdekesnek és hasznosnak bizonyuló új fuzzy csoportosítási és genetikus algoritmusok is.

Az értekezés szerkezetéről, nyelvezetéről, formájáról

A 156 oldal terjedelmű, kiváló angol nyelven írt értekezés 157 hivatkozást tartalmaz. Az érdemi rész 93 oldal, azaz teljesíti azt a követelményt, miszerint a Műszaki Tudományok Osztálya az érdemi fejezetek terjedelmét 100 oldalra korlátozta.

A szerző az Osztály előírásait követve a részletes levezetések, számítógépi algoritmusok bemutatását két mellékletbe szerkesztette. Az első melléklet az alkalmazások részleteinek

ismertetésére, a második melléklet pedig az alkalmazott technikák és modellek rövid elméleti háttérének bemutatására fókuszál.

Tekintettel arra, hogy az olvasónak e függelékek tanulmányozása nélkül is meg kell tudnia győződni az eredmények helyességéről, újdonságáról, a dolgozat egyes fejezetei szerkesztésileg is jelzett példákat tartalmaznak.

A dolgozat a szerző 67 cikkben közölt eredményeit dolgozza fel és foglalja egységes keretbe. Minden egyes fejezet összefoglalással, azaz a téziszfűzetben összefoglalt tézisek és a kapcsolódó publikációk ismertetésével zárul. A dolgozatnak előnyére válhatott volna egy záró, értékelő fejezet. Ennek mellőzésének oka feltételezhetően a terjedelmi korlát, illetve a téziszfűzet hasonló funkciója lehetett.

A dolgozat egy tömör, a technológiafejlesztés során az információátvitel fontosságát, és az ebből az elvárásból levezetett, a Szerző által javasolt módszertan ismertetését tartalmazó bevezetéssel indul. E bevezetés zárása (1.3 fejezet) próbál kísérletet tenni arra, hogy integrálja és gondolatilag összekapcsolja a következő négy, csaknem önálló részt: a folyamat adattárházakat (2. fejezet), csoportosítási algoritmusokat (3. fejezet), *a priori* információkon alapuló korlátokat alkalmazó identifikációs és adatsimítási technikákat (4. fejezet), valamint az evolúciós és genetikus algoritmusokkal kapcsolatos fejlesztéseket (5. fejezet) ismertető egységeket.

Az értekezés tartalmi értékeiről

A folyamatadatok kezelésére, azaz a folyamat-adattárakra vonatkozó fejezet egy, a Szerző által kidolgozott módszertant ismertet. A módszertani rész a konkrét eredményeket tartalmazó „A” függelék nélkül anekdotikusnak tűnhet. Ebben a mellékletben derül ki ugyanis, hogy – bár mára már széles körben megjelentek a folyamat- és termelésirányítással foglalkozó multinacionális vállalatok termékei között a folyamat-adatbázisok létrehozását és elemzését támogató eszközök –, ezek alkalmazásához elengedhetetlen a szerző által oly sokszor hangoztatott technológiai ismeretek felhasználása, amelyek pl. a technológia különböző pontjain keletkező adatok dinamikus összekapcsolásához is szükségesek.

A fuzzy csoportosítással kapcsolatos fejezet négy új, a folyamatmérnöki gyakorlatban jelentős probléma (regresszió, idősor elemzés, osztályozás, dinamikus rendszermodellek rendűségének meghatározása) kapcsán kidolgozott csoportosítási algoritmust ismertet. A kifejlesztett algoritmuscsalád alapja nemcsak a problémaorientált távolságmérték ügyes definiálása, hanem annak a mutatókészletnek a meghatározása is, mellyel a generált csoportok megfelelően jellemezhetők és ezen jellemző értékek az eredeti mérnöki feladat megoldásában (pl. modellrendűség meghatározás) közvetlenül alkalmazhatók.

A megoldásmód sikerességére a Jelölt releváns publikációiban fellelhető különböző paraméterek is garanciát jelentenek. A módosított Gath-Geva csoportosítási algoritmust kidolgozó munka önmagában 97, az osztályozásra alkalmazott módosítása 54, a többváltozós idősor elemzésre kidolgozott módszer 16, a modellrendűség meghatározásával kapcsolatos technika 20 független hivatkozást kapott.

A negyedik, az *a priori* információkat lineáris paraméterkorlátokként felhasználó rész két egymástól rendkívül távol álló területen, a Takagi-Sugeno (TS) fuzzy modellek identifikációjában és a többváltozós adatsorok spline simításában alkalmazza ugyanazt a

megoldási technikát. A 4.1 fejezet a TS fuzzy modellek felépítését részletesen nem ismerők számára nehezen követhető. A kapcsolódó elméleti háttérrel a szerző a B.2 függelékben ismerteti, de az erre való utalásról sajnos megfeledkezik a 4.1 fejezetben.

Az ötödik fejezet szintén önálló alfejezetekből épül fel. Az első alfejezet egy genetikusan programozás célirányos fejlesztését, míg a második alfejezet az evolúciós algoritmusok olyan módosítását mutatja be, ahol egy sokszor nehezen formalizálható célfüggvény megfogalmazása helyett inkább a felhasználó szubjektív értékelésére támaszkodó, interaktív optimalizációs algoritmust dolgoz ki a Szerző.

E fejezet kapcsán is felmerülhet az olvasóban, hogy a kevesebb néha több. A száz oldalas dolgozatban egy adattárház építési és elemzési módszertan, négy új csoportosítási algoritmus, egy a Takagi-Sugeno fuzzy modellek identifikációját javító megoldás, egy többváltozós spline simítási technika, egy új, paramétereiben lineáris egyenletek feltárását támogató genetikusan programozási algoritmus és végül egy interaktív evolúciós stratégia került ismertetésre.

Látszik ugyan, hogy a szerző törekszik a széles eszköztár követhető ismertetésére a függelékben megfelelő információk és fontos forrásmunkák megadásával, de az olvasóban néhány területen hiányérzet maradhat.

Az értekezésben összefoglalt új tudományos eredmények

A tudományos eredmények a dolgozat négy fejezetéhez és azok alfejezeteihez illeszkedő négy tételben lettek a szakterület kutatói közössége számára közérthetően és tömören megfogalmazva. (Az első tézispont eredményeit az A függelék mutatja be részletesen.)

Mindegyik altézispontot legalább egy referált nemzetközi folyóiratban megjelent publikáció támaszt alá.

A tézisfüzet jelentős része a III., az eredmények gyakorlati alkalmazhatóságát ismertető rész, amely jól mutatja, hogy a Szerző a dolgozatban „összezsúfolt” eredményein kívül további területeken is értékes eredményeket ért el (pl. kemometria, katalizátor-fejlesztés, többváltozós adatok megjelenítése).

A hivatkozó publikációk a kidolgozott megoldások szerteágazó alkalmazásairól számolnak be (rákkutatás, földrengési adatok előrejelzése, képfeldolgozás, térinformatika stb). Ezen alkalmazásokat a Szerző kutatócsapata által kidolgozott Fuzzy Clustering and Data Analysis Toolbox is támogatja. Ez önmagában is nagyon jelentős eredmény, hiszen a Szerző munkája jóvoltából a nemzetközi kutató közösség kiválóan alkalmazható eszközökhöz jutott hozzá.

Kérdések

Az értekezés olvasása közben a következő kérdések merültek fel bennem:

- Az információátvitel a különböző típusú modellek között egyre inkább kutatott terület. A Szerző ilyen szempontból az amúgy is transzparensnek és értelmezhetőnek tartott fuzzy modelleket vizsgálta. Milyen példákat tud mondani az amúgy fekete doboz modelleknek tekintett neurális hálózati modellekből történő információ kinyerésre?
- A 3.4 fejezetben kidolgozott technika a dinamikus rendszerek bemenet-kimenet adatai által definiált tér geometrikus jellemzésével foglalkozik. A kidolgozott jellemzőkön, módszertanon kívül van-e tudomása más, hasonlóan geometrikus alapokon nyugvó

mértékekről, melyek akár más, autonóm, illetve kaotikus rendszerek esetében is alkalmazhatók?

- Milyen gyakorlati jelentősége van a technológiai adatok splinokkal történő simításának? Milyen alkalmazásokban van/lehet e lépésnek gyakorlati haszna?
- Rejthet-e magában bizonyos veszélyeket a bizonyos tartományon simán illesztett spline függvények illesztési tartományhoz közeli tartományon való „hullámos” viselkedése? Hogyan kezelhetők általánosan ezek a problémák?
- Az interaktív evolúciós stratégia gyakorlati alkalmazhatóságát milyen eszközökkel célszerű támogatni? Hogyan lehet/kell mérni e megoldások minőségét?

Összefoglaló értékelés és javaslat

Az értekezés Abonyi János olyan saját kutatási eredményeit ismerteti, melyek tökéletesen elkülöníthetők a PhD fokozat megszerzéséig elért eredményeitől.

A fuzzy csoportosítási algoritmusok fejlesztésével, a Takagi-Sugeno fuzzy modellek identifikációjával és a genetikus programozással kapcsolatos eredményeknek a számítási intelligencia témakörében elméleti jelentősége van. A technológiai adatok kezelésével, az *a priori* információkat is figyelembe vevő spline simítást alkalmazó és a kapcsolódó technológiafejlesztést és működtetést támogató állapotbecslési megoldásai komoly gyakorlati jelentőséggel is bírnak.

A doktori munka eredményeit és azok tudományos értékét bőségesen elegendőnek tartom MTA doktora cím megszerzéséhez. A Jelölt ezen értekezésében összefoglalt eredményei, melyeket a PhD fokozat megszerzése után ért el, lényegesen hozzájárultak a szakterület továbbfejlesztéséhez.

A fentiek alapján javaslom a nyilvános vita kitűzését, a doktori mű elfogadását. Az értekezés minden egyes tézisét maradéktalanul elfogadom új tudományos eredménynek, és élek azzal a lehetőséggel, hogy már a bírálás jelen fázisában javasoljam Abonyi János részére az MTA doktora cím odaítélését.

Budapest, 2011. február 28.



Rudas Imre
Az MTA doktora