

Bírálati vélemény

Abonyi János: "Data Mining Techniques for Process Development"

(Technológiafejlesztési célú adatbányászati technikák)

c. MTA doktori (DSc) értekezésről

Az értekezés az utóbbi évtizedben divatossá vált adatbányászati technikák komponenseivel és azok integrálási lehetőségeivel foglalkozik. A megközelítések dominánsan a Takagi-Sugeno-féle (TS) fuzzy technikákra, az optimalizálási technikák az LS (Least Squares) és GP (Genetic Programming) alapú módszerekre, az esettanulmányok pedig vegyipari folyamatokra (polietilén gyártás, kémiai reaktorok) korlátozódnak.

Az értekezés 5 fejezetből és 2 függelékkel (A-B) áll. A négy tézis csoport rendre a 2-5 fejezetekhez kapcsolható.

Az 1. fejezet bevezetés, amely röviden bemutatja az életciklus-modellezés alapfogalmait, a modell-bányászat alapsémáját és metodikáját, valamint az értekezés felépítését.

A 2. fejezet a folyamat-adattárházak és a folyamat-adatbányászat koncepcióját mutatja be. A fő eredmény a 2.2 ábra blokkvázlatán látható integrált rendszer, amelynek feladata az operátori felügyelet (OSS, Operator Support System) és döntéshozatal (DSS, Decision Support System) támogatása. A megfogalmazott célok között van a hibadetektálás és hibadiagnózis, a termékminőség ellenőrzése és a következtetés a folyamatrendből a működési feltételekre. Az itt megfogalmazott know-how számos később tárgyalt módszert integrál, ezek szerepe és jelentősége itt még nem látható, ezért jobb lett volna a fejezetet az értekezés végére tenni.

A 3. fejezet a fuzzy klaszterezés kérdésével foglalkozik nemlineáris regresszió, idősor-szegmentálás és modell-rendszer szelekció céljára.

A nemlineáris regresszió területén a lokális és Gauss-féle kevert modellek Khambatala (1996) PhD disszertációjára vezethetők vissza. Az első jelentősebb fuzzy megközelítések Gustafson-Kessel (1979) és Gath-Geva (1989) munkái, melyeknek a Bayes-technikán alapuló valószínűségszámítási interpretáció adható. Ezekre LS és a sajátérték-sajátvektor technikán alapuló bemeneti transzformációt alkalmazó totális LS módszer épülhet. A fejezet a bemeneti transzformáció megkerülésére a módosított Gath-Geva technikát mutatja be, amely egy EM-elvű (E: Expectation, M: Maximization) klaszterezési módszer TS fuzzy rendszerek esetén, és amely Abonyi-Babuska-Szeifert (2002) eredménye. A módszert a szerző az Automobile MPG (miles per gallon) benchmark probléma megoldására használja és összehasonlítja más módszerekkel. A MATLAB környezetben jól ismert ANFIS-technika a módosított Gath-Geva módszerhez hasonló performanciát eredményez. Az antecedens és konzekvens változók számának csökkentésére az ortogonális LS módszert (OLS) és a módosított Fischer-féle szeparációs módszert (FIS) javasolja, melyeket a fenti benchmark probléma keretében vizsgál.

Az idősor-analízis célja folyamatok rejtett belső üzemmódjainak behatárolása (szegmentálása) a folyamatról rögzített jelek idősorainak analízise révén. A többváltozós (multivariate) adatanalízis fő célja az adatok középértékében, varianciájában és korrelációjában bekövetkező változások felismerése. A kifejlesztett megoldás a Bayes-elvű fuzzy klaszterezés, a valószínűségi főkomponens analízis (PPCA, Probabilistic Principal Component Analysis) és a log-likelihood technikákra épül. Az analízis két szinten folyik, a megfigyelt időtartományt egymáshoz csatlakozó diszjunkt intervallumokra bontja és a rendszer viselkedését az intervallumokon (szegmenseken) leírja, miközben egy költségfüggvényt optimalizál. Induláskor nagy-

számú csatlakozó időszakmenst választ, majd megkísérli a hasonló tulajdonságú és csatlakozó szegmenseket összevonni (merge), ezáltal csökkentve a megoldás komplexitását. Az idő skalárban és a vektor adatban Gauss-féle tagsági függvényeket használ, a paraméter-optimalizálásra a módosított Gath-Geva algoritmust használja, a PPCA komponenseket az adatok kovariancia mátrixából nyeri, amelyek részt vesznek a költségfüggvényben. Az előzmények tekintetében Keogh et al. (1998, 2001), Babuska-Veen-Kaymak (2002) és Tipping-Bishop (1999) munkái, a PCA modellek hasonlóságának modellezésében pedig Krzanowsky (1979) és Singhal-Seborg (2001) munkái fontosak. A döntéshozatal fuzzy technikán alapul. A klaszterek összevonása (cluster merging) Kelly (1994) módszerére épül. A szerző módszerét az Idősor-Szegmentálási Algoritmusban foglalja össze, melyet egy rejtett változásokat tartalmazó szintetikus adathalmazon mutat be.

A fuzzy klaszterezés optimális megoldásához meg kell találni az egyváltozós tagsági függvényeket. Ennek a projekciós problémának az eliminálására Abonyi-Seifert (2003) módszerét javasolja a szerző, amely bevezeti az osztály-cimkét (class label) az optimális klaszterhalmaz jellemzésére. A módszer hasonlít Rahman-Fairhurst (1997) Multi-Prototype Classifier technikájához, de eliminálja annak előnytelen (diszjunkt klaszterek, manuálisan választott klaszterszám) tulajdonságait. A klaszter távolság két komponenset foglal magába: az adat geometriai távolságát a klaszter centrumtól és annak valószínűségét, hogy a klaszter leírja az osztály adatsűrűségét, amely utóbbi az osztály-cimke értelmezésének alapja. A szerző a módszert a Wine Data benchmark probléma keretében mutatja be és összehasonlítja más módszerek eredményeivel.

A fuzzy klaszterezés fontos kiindulási feladata a modell-rendszám szelekció MIMO rendszerek esetén, ahol a modell-rendszám alatt a modellben figyelembe vett korábbi bemenőjelek és kimenőjelek száma (a modell memóriájának mérete) értendő. A szerző által kifejlesztett módszer előzménye a Kennel-Brown-Abarbanel (1992) által javasolt FNN (False Nearest Neighbor) módszer, amely az adatregresszorhoz meghatározza legközelebbi szomszédot, teszteli a kimenetek összetartozásának (távolságra normalizált) jóságát egy küszöbhez képest, kiszámítja a hibás legközelebbi szomszédok arányát, és növeli a rendszámot, amíg a hibaarány egy elfogadható korlát alá nem kerül. A szerző Fuzzy Klaszterezés-Alapú FNN algoritmusát az adaptív küszöb meghatározásába a klaszter-kovariancia mátrix minimális sajátértékéhez tartozó sajátvektort vonja be. A direkt modell-fokszám becslésére felhasználja Cadzow-Solomon (1988) és Liang-Wilkes-Cadzow (1993) eredményeit lineáris rendszerek esetére, amelyek, tekintettel az elsőrendű fuzzy TS-modellek lineáris konzekvens részére, átfogalmazhatók egy a priori valószínűségi mértékké, és amelyre alapozva az optimális rendszám egy optimum probléma megoldása. A szerző a direkt modell-fokszám szelekciót egy polimerizációs reaktor példáján keresztül mutatja be.

A 4. fejezet a paraméter identifikáció során figyelembe veendő korlátozásokkal és az adatok spline technikával történő simításával foglalkozik.

Szürke-doboz identifikáció számára a korlátozásokat lokális, globális és relatív esetekre bontja. Stabil diszkrétidejű lineáris alrendszereket feltételezve a TS-modell konzekvens részeiben megadja a parameterekre vonatkozó korlátozásokat egyenlőtlenségek alakjában Tulleken (1993) eredményeit felhasználva. A korlátozások kezelni tudják a stabilitás, dc-erősítés és szabályozási idő feltételeket. Alkalmazásukat két-tartályos rendszer fuzzy modellezésén illusztrálja hat háromszög-alakú tagsági függvényt feltételezve az antecedens részekben, melyek centruma ismert.

Spline-technikára épülő simítás a kis mennyiségű és zajos adatok miatt szükséges. A szerző standard köbös approximációt alkalmaz, szimultán approximálva a többdimenziós (multivariate) adatokat korlátozások figyelembevételével. A módszert kineti-

kus modellek sebesség konstansainak meghatározása keretében mutatja be, négy különböző becslési módszert vizsgálva. Egy további példa egy nempublikus (confidential) batch reaktor koncentráció profiljainak becslését mutatja be, ahol a hat koncentráció közül csak három mérhető, és a technológia leírása bonyolult és nem teljesen ismert. A probléma bilineáris korlátozásokat is tartalmaz az ismeretlen paraméterekre, ezért a megoldást alternáló optimalizálás határozza meg iterációval.

Az 5. fejezet a populáció-alapú (evolúciós) optimalizálási algoritmusok javítási kérdéseivel foglalkozik, ezen belül a Genetikus Programozás (GP) alkalmazásával a rendszer-identifikációban és az interaktív optimalizálással.

Az irodalomból jól ismert GP módszer a nemlineáris modellekben szükséges többváltozós függvénykapcsolatok felismerésére kerül felhasználásra, ennek speciális esete a többváltozós polinomiális modell. A feladat a modellben szükséges függvénykapcsolatok automatikus megkeresése, illetve polinomiális modell esetén a modellben szereplő többváltozós polinom termék meghatározása. Erre a célra az irodalomban a Genetikus Algoritmusokétól (GA) eltérő keresztezési és mutációs operátorok, valamint visszahelyettesítési technikák rendelkezésre állnak. A populáció egyedei fastruktúrával írhatók le, az optimalizálás során a fa-struktúra OLS technikával minősíthető a fitness számításakor. A módszert a szerző egy egyszerű analitikus példa és a gyakorlatból vett folytonos polimerizációs reaktor modellezése keretében mutatja be.

A GP megközelítés több alternatív megoldást is fel tud kínálni a tanítás egyes epochaiban, amelyből egy humán döntéshozó szelektálhat és befolyásolhatja a populációba visszakerülő egyedeket (IEC, Interactive Evolutionary Computation). IEC esetén a populáció mérete kicsi a könnyű döntéshozatal érdekében. A módszert a szerző egy MPC (Model Predictive Control) MIMO szabályozó paramétereinek hangolására alkalmazza szimulációs vizsgálat keretében. Az MPC szabályozó a Skogestad (1997) által publikált modellű bináris desztillációs oszlopot szabályozza zavarások (táplálási ráta és összetétel) jelenlétében, a fölérendelt IEC algoritmus a költségfüggvényben szereplő négy skalár súlyozó tényezőt, valamint a predikciós és irányítási horizontok méreteit hangolja. A döntéshozatalt az IEC vizuális felülete segíti.

Függelék A alkalmazási példákat mutat be: A.1 folyamat adat-tárház, A.2 folyamat tranzíciók monitorozása állapotbecsléssel, A.3 szemi-mechanisztikus modellek a folyamatminőség becslésére.

Függelék A.1 a TVK polietilén gyár termelési folyamatát analizálja. A folyamat részei a plomerizációs gyártási egység (PPU), a granulációs gyártási egység (GPU) és a polietilén minőségellenőrzési labor (PE Lab). A teljes folyamatról a Honeywell Distributed Control System (DCS) a Process History Database (PHD) adatbázison keresztül mért adatokat, az Advanced Process Control (APC) modulján keresztül pedig számított adatokat ad. Ezek az adatok 15s-ként keletkeznek és 13 jelet tartalmaznak, tipikusan áramlási sebességeket és hőmérsékleteket. A GPU 1 vagy 2 óránként ad adatokat riportok formájában. A PE Lab követi a polimerpor, granulátum és batch minőség jellemzőket riportok formájában, továbbá 1 óránként méri a folyamatminőséget. A melt-index (MI) és sűrűség 0.5–5 óránként kerül meghatározásra. A szerző és társai által kifejlesztett kiegészítő rendszer MATLAB scriptekből épül fel. Megoldja a diamikus adatintegrációt, a polimerpor mennyiségének és korának számítását, a silók tartalmának és összetételének (átlagos molekulatömeg, melt-index) dokumentálását, a végtermék minőségének becslését és validálását, valamint az egyes termelési periódusok befolyását a végtermék minőségére. Egy féléves termelési folyamat analízisének tapasztalatain alapulva vizuális információs eszköz került kifejlesztésre az operátorok online támogatására, amely a PCA analízisen, valamint a klasszikus Hotelling T2 és a Q hibamérték jellemzőkön alapul. A szerző bemutatja, hogy a reaktorállapotok és be-

meneti változók kiértékelése révén szegmentálhatók a folyamatban bekövetkezett jelentősebb változások.

Függelék A.2 folyamatátmenetek monitorizását vizsgálja állapotbecslés felhasználásával. A cél változásdetektálás a 2. fejezet eredményei alapján. Három algoritmust mutat be, az első a bemenet-kimenet adatok szegmentálásán alapul, a második nemlineáris állapotbecslést használ a változások detektálására a becsült állapotokban, a harmadik változások korrelációját vizsgálja a modellezési hibával. A DCS méri a 4 bemenet áramlási sebességét és hőmérsékletét 15s-ként, a katalizátor áramlási sebességét, valamint a belépő és kilépő hűtővíz áramlási mennyiségét és hőmérsékletét. A modell épít a tömeg, komponens és energia mérlegekre. Az állapotváltozók a reaktorban lévő folyadék és polimer mennyisége, a láncátviteli ágens, a monomer, a komonomer és a katalizátor koncentrációja, valamint 3 reakcióráta együttható, összesen 10 állapotváltozó. A modell 5 fő állapotegyenlettel jellemezhető, melyek normalizálva vannak. A kezdeti állapot és a zajok kovariancia mátrixait induláskor meg kell választani a szegmentálási algoritmus számára, további paraméter a szegmensek száma. A szegmentálásra az egyszerűsített Vasko-Toivonen (2002) módszer kerül felhasználásra. A vizsgált példa 100 órás üzemvitelt analizál, ahol a 45. órában változás lép fel az üzemben. A három módszer hasonló eredményeket produkál, de a második érzékeny a felejtési együttható megválasztására. Egy másik példa illusztrálja, hogy termékváltás nélküli esetekben is képes a harmadik módszer belső változások detektálására, a példában a katalizátor táplálás megváltozására.

Függelék A.3 szemi-mechanisztikus modellek alkalmazását vizsgálja a termékminőség becslése keretében. A polimerizációs egység nemlineáris modelljébe egy állapot és bemenet függő önálló blokk épül be visszacsatolás nélküli neurális hálózat (NN) formájában. A bemenetek között szereplő állapotok az aktuális etilén koncentráció, a reaktor hőmérséklet és a hexán koncentráció a polietilénben. Az NN közreműködik a melt-index predikciójában és csökkenti igyekszik a laboratóriumi mérések 1 órás holtidejének hatását. Az NN hangolása kvázi-Newton módszerrel történik, ahol a deriváltak számításához a simítást köbös spline-interpoláció végzi. Az állapotbecslést Norgaard-Poulsen-Ravn (2000) módszerén alapuló DD1-szűrő végzi, amely biztosítja az approximációt a ritkán mért melt-index értékek között.

Függelék B az értekezés elméleti alapjait foglalja össze, így bevezet a fuzzy klaszterezésbe (B.1: fuzzy c-mean és Gath-Geva algoritmus), a fuzzy modellezésbe (B.2: TS fuzzy rendszer), az osztályozáshoz használható fuzzy modell-struktúrákba (B.3: Bayes-osztályozó, fuzzy osztályozó, kevert sűrűségű modellek, fuzzy döntéssfa), a populáció-bázisú optimalizálásba (B.4: evolúciós algoritmus, evolúciós stratégia, genetikus programozás) és a paraméterekben lineáris modellek identifikációjába (B.5: NAARX, Volterra, polinomiális ARMA modellek, LS és OLS becslés).

Formai észrevételek:

Az értekezés terjedelme (156 oldal) indokolt a tág problémakörre és a vizsgált problémák összetettségére való tekintettel. Az értekezés angol nyelvezete és stílusa jó. Gépelési hiba csak elvétve akad (például p.63, p.71, p.89, p.100, p.116, p.107). Van néhány szóiméltés (p.88) és dupla tagadás (p.49). Néhány helyen ?? áll (p.117, p.137, p.140). Nincs hivatkozás a Figure B.3 ábrára a szövegben (p.124). Sok sorból álló komplett részek a p.117 oldalról megismételve szerepelnek a p.118 oldalon két esetben is. Nem világos, hogy mi a sárgával jelölt adat Figure A.5-ben (p.98). Nem világos, hogy a Figure A.8/(b)-ben szereplő változások kb. 5 és 12.5 óránál (figyelembe véve az időtengely 15s-os léptékét), milyen kapcsolatban vannak a szövegben jelzett termékváltásokkal a 24, 54 és 86 óra körül.

A tézisek értékelése:

A szerző 4 téziscsoportban fogalmazta meg új tudományos eredményeit, melyek további altézisekre vannak bontva. A tézisek értékelésénél fenntartom az egyes fejezeteknél már megtett megállapításaimat, melyeket további észrevételekkel egészíték ki. A tézisek értelmezésénél problémát okoz, hogy a téziszüzet nem rendeli össze a téziscsoportokat és az értekezés fejezeteit. Ezért a tézisek értelmezésénél azzal a hipotézissel élek, hogy az 1. téziscsoport a 2. fejezettel és Függelék A-val áll kapcsolatban, a 2-4. téziscsoportok pedig rendre a 3-5. fejezetekkel. A téziseket az alábbi formában és pontosításokkal tudom elfogadni.

1. téziscsoport: Folyamat-adattárházak egy lehetséges struktúrájának meghatározása, amely integrálja az operátor felügyeleti (OSS) és a döntéshozatalt támogató (DSS) alrendszereket a hibadetektálás/hibadiagnózis, termékminőség ellenőrzés és folyamat-trend vizsgálatokban. A téziscsoportot ebben a megfogalmazásban tudom elfogadni. Az 1.1 tézis mérnöki szempontból triviális megállapítás. 1.2-1.3 tézisek nem mondanak többet, mint hogy nemlineáris modellekbe önálló egységként NN modellek is beágyazhatók. Függelék A.3 ezt szemi-mechanisztikus modellnek nevezi, lényegében azonban ez felfogható alkalmasan választott bemenőjel transzformációnak is. Nem világos az sem, hogy az integrált rendszer, mint szoftvertechnológiai termék, milyen formában és általánosságban valósult meg.

2. téziscsoport: Regressziós, osztályozási és idősor-szegmentálási feladatok megoldására fuzzy csoportosítási algoritmusok kidolgozása. A 2.1 tézis a Gath-Geva módszer módosításával és az OLS módszer bevonásával hatékony eljárást ad fuzzy TS rendszerek lokális lineáris modelljeinek meghatározására és a feltételrendszerének redukciójára. A 2.2 tétel, szintén a Gath-Geva módszer általánosításával, hatékony módszert ad fuzzy szabályalapú osztályozási modellek identifikációjára és interpretálására. A 2.3 tétel a Gath-Geva módszer és a PPCA hibamértékek bevonásával módszert ad többváltozós idősorok szegmentálására. A 2.4 tétel bemenet-kimenet modellek rendszámának (memóriaméretének) meghatározására ad új módszert, amely javítja az FNN módszer képességeit, és a generált klaszterek kovariancia-mátrixai tulajdonságainak figyelembevételével csökkenti a számításgényes lépések volumenét. A 2.1-2.4 téziseket eredeti formájukban elfogadom.

3. téziscsoport: Stabil folyamatok diszkrétidejű fuzzy TS approximációs modelljeinek meghatározására korábbi irodalmi eredmények bevonásával identifikációs módszer kidolgozása, amely a bemenőjelek tagsági függvényeinek ismerete esetén QP optimalizálást alkalmaz, és korlátozásokként figyelembe tudja venni a stabilitási, dc-erősítés és felfutási idő elvárásokat a paraméterterben (3.1 tétel). Módszer kidolgozása vegyipari folyamatok reakciókinetikai paramétereinek meghatározására köbös spline-technika és alternatív optimalizálás bevonásával, amely figyelembe tudja venni az a priori anyagmérlegeket és bilineáris korlátozásokat (3.2 tétel). A téziseket a fenti megfogalmazásban fogadom el.

4. téziscsoport: A megfogalmazott két tézis az irodalomból már ismert módszerek javítását célozza meg. Nemlineáris modellek struktúrájának meghatározására GP alkalmazása javasolható, amelyben a fa-struktúra OLS technikával minősíthető az egyedek fitness értékének számításakor (4.1 tézis). Folyamatmérnöki problémák keretében a csak kompromisszum formájában kielégíthető ellentmondó célok és korlátozások esetén interaktív EC módszer alkalmazása javasolható, amely az egyes op-

timalizálási lépésekben a populáció egyedeinek tulajdonságait vizuálisan könnyen értelmezhető formában felkínálja a humán szakértő számára, amely ezeket felülbírálván befolyásolhatja a populációba visszakérülő egyedeket, lehetővé téve alacsony szintű tervezési folyamatok paramétereinek magasszintű optimalizálását (4.2 tétel). A tézisek kevés új eredményt tartalmaznak. A GP alkalmazása a modellezésben már kb. tíz éve ismert a gyakorlatban és megjelent a hazai egyetemi oktatásban is (pl. BME IIT diplomaterveben 2003-ban). A vektorkritériumú optimalizálás köréből szintén ismert, hogy az ott nem összehasonlítható Pareto-optimumok közül egy döntéshozó szakértőnek kell kiválasztani a megoldást. A fentiek tükrében a szerző eredménye igéretes kísérlet, de nem átütő új tudományos eredmény.

A tézisek publikálása rangos nemzetközi folyóiratokban és nemzetközi konferenciák kiadványaiban megtörtént, a tézisek megfogalmazása a vonatkozó publikációkat is megjelöli. A tézisekben szereplő publikációk kevés kivétellel többszerzősek és csupán az 1. téziscsoport tartalmaz 2005 utáni publikációkat. A téziszüzetben megjennek olyan publikációk is, amelyek nem szerepelnek az értekezésben.

Kérdések:

1. Ismertesse, mi a saját eredménye a téziszüzet 12 sorszámú Abonyi-Babuska-Szeifert (2002) és 18 sorszámú Abonyi-Babuska-Verbruggen-Szeifert (2000) publikációiban.
2. Ismertesse, hogyan alkalmazhatók az értekezésben kidolgozott folyamatfelügyeleti technikák valós idejű hibadetektálás és hibadiagnózis céljára.
3. A szegmentálási módszerek koncepciója feltételezi, hogy a viselkedésben az egyes szegmenseket bizonyos állandóság jellemzi az átlagban, kovariancia mátrixban stb. Jelent-e ez korlátot gyorsan változó folyamatok esetén, pl. repülőgépek vagy más járművek idősorainak elemzésében?
4. Az idősorok szegmentálásakor az időintervallumok diszjunktan csatlakoznak és az idő-változóban az egyes intervallumok centrumától mért távolság exponenciálisan változik. Ez részben ellentmond annak, hogy olyan intervallumokat kell keresni, amelyben a folyamat minősége "homogénnek" tekinthető. Hogyan értékeli ezt a problémát az eredmények interpretálásakor?
5. Hosszabb technológiai idősorok esetén egy korábbi viselkedésmód később, időben távolabb ismét előfordulhat. A szomszédos intervallumok merzselési technikája lehetővé teszi-e a távoli intervallumban megismétlődő hasonló viselkedésmódot egy korábbi viselkedésmód replikációjának tekinteni, hiszen intervallumaik nem csatlakoznak?
6. A Gath-Geva módszer és rokon technikák alkalmazásakor eltűnik az eredményekben az egyenlőtlenség típusú korlátozások hatása, ami visszavezethető a Kuhn-Tucker-Karush tételre. Ennek tükrében mi a véleménye a módszerek kiterjesztésének lehetőségéről Gauss-típusú helyett más, véges tartójú tagsági függvények esetére?
7. Tud-e becslést adni arra, milyen változószámig képes a GP technika valóban megtalálni a globálisan optimális nemlineáris dinamikus modell jó approximációját, értelmes epochaszám korlát mellett.
8. AZ IEC technika folyamatosan minősíti az egyedeket és befolyásolja a visszahegyettesítést a populációba. Mi a garancia, hogy a populáció nem konvergál lokális optimum felé?

Összefoglalva megállapítom, hogy az értekezés fontos, a kutatások középpontjában álló rendszermodellezési és vegyipari folyamatirányítási kérdésekkel foglalkozott, és a nemzetközi kutatások figyelembevételével is jelentős új saját eredményeket fogalmazott meg a technológiafejlesztési célú adatbányászati technikák területén, melyeket külföldi és hazai társszerzőkkel közösen rangos konferenciákon és folyóiratokban publikált, és amelyekre több rangos külföldi hivatkozás történt.

Az értekezés hiteles adatokat tartalmaz. A téziseket (korábbi észrevételeim fenntartása mellett és a 4. téziscsoport kivételével) a fenti megfogalmazásban elfogadom. Az értekezés a PhD fokozat 2000-ben történt megszerzését követően jelentős eredeti tudományos eredménnyel gyarapította a rendszermodellezést és a folyamatirányítási célú adatbányászati technikákat, hozzájárult a tudományág fejlődéséhez, ezért az értekezés elfogadását és a nyilvános vita kitűzését javaslom a műszaki tudományok területén.

Budapest, 2011. április 14.

Lantos Béla
a műszaki tudomány (MTA) doktora