

Korszerű adatelemzési technikák és modell alapú algoritmusok a kísérlettervezésben és kiértékelésben

OTKA 49534 Zárójelentés

Készítette: D. Abonyi János, témavezető

2008 november 30.

Bevezető, általános áttekintés

Az elmúlt évtizedekben a vegyipari technológiák, folyamatok, illetve a vegyipar gazdasági környezete olyan irányban változott, hogy egyre égetőbbé vált az igény az olyan eszközök és módszertanok iránt, melyek a technológia fejlesztést, üzemeltetést és a termékfejlesztés támogatják, integrálják, azaz e feladatok ellátásának hatékonyságát növelik. E kihívásra a nagyobb vegyipari konszernek fejlesztési osztályain egymástól eltérő megoldások születtek, illetve születnek, mégis a megoldások közös jellemzője, hogy a kutatás, a fejlesztés, a tervezés, a gyártás, a marketing és a menedzsment feladatokat ellátó egységek intenzív kommunikációjára épülnek. Ezen egységek közti információátvitel alapvető eszköze a modellezés és a szimuláció, azaz a termelő technológiáknak és a technológiához kapcsolódó folyamatoknak olyan színvonalú modellezése, mely képes arra hogy integrálja a termék és technológia fejlesztési láncot a termelő folyamat, illetve műveleti egység minden idő és méretskálájára vonatkozóan.

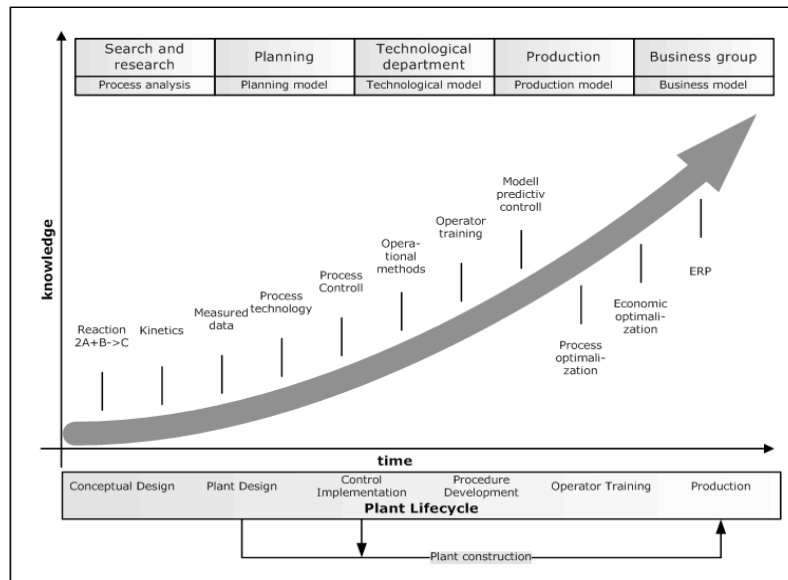
A termék- és technológia fejlesztési feladatokhoz szükséges ismeretek feltárására, a különböző szinteken jelentkező részfeladatok közti információátvitelre, és ezen heterogén forrásokból származó tudások integrálására alkalmas eszközök fejlesztése volt kutatások alapvető célkitűzése. Ahhoz, hogy az ezen eszközök által nyújtott lehetőségeket a lehető legjobban ki lehessen használni, a K+F tevékenységnek a következő két lépését kell nagyon tudatosan elvégezni:

- Kísérlettervezés, azaz hipotézisek generálása és annak biztosítása, hogy kísérleteink a lehető legtöbb információt nyújtsák a hipotézisünkről, illetve a fejlesztés „keresési terében” a leghatékonyabban közelítsük meg a keresett optimumot.
- Mérési eredmények elemzése: azaz a mérési eredmények megbízhatóságának minősítése, új összefüggések feltárása, hipotézisek generálása.

A javasolt megközelítés alapgondolata az volt, hogy a kísérlettervezés iteratív folyamatának, tágabb értelemben a teljes technológia és termékfejlesztési folyamatnak minden lépésében az összes elérhető releváns mérnöki és természettudományos ismeret (first-principle knowledge) alkalmazható legyen, illetve ezen ismeretek tárháza a kísérleti eredmények hatékony értékelése alapján minden lépésben a lehető legnagyobb mértékben bővüljön.

Ez a koncepció leginkább a piacvezető vállalatok (DuPont, Dow Chemical) mérnökeiben és vezetőiben kristályosodott ki a "modell integrálja a teljes szervezetet" jelmondatban. Ezek a piaci szereplők komolyan hisznek abban, hogy az adatoknak, az információknak és a tudásnak teljes mértékben végig kell követniük a termék, illetve a technológia fejlődését a kutatás munkafázisától a gyártásig, illetve az eladásig. [Krieger JH: Process Simulation Seen As Pivotal In Corporate Information Flow, Chemical & Engineering News, March 27, 1995]. Természetesen ez a koncepció egyre nagyobb szerepet tulajdonít a modellezésnek és a szimulációnak a bio-, vegyész-, illetve folyamatmérnöki

tudományokban, prognosztizálható például, hogy a működő technológiák optimalizálásában és karbantartásában kulcselem lesz e technikák alkalmazása. Az AspenTech vállalat stratégái ma már "életciklus modellezésről" beszélnek, mely ebben az esetben a "modell életciklusának" hatékony követését jelenti, azaz annak a kezelését, hogy amennyiben a modell integrálja a szervezetet, a termelő folyamatot, akkor annak életciklusát a modell életciklusának is követnie kell, integrálva a különböző termék- és technológiafejlesztési lépéseket mind időben, mind pedig méret szerinti léptékekben (1. ábra)



1. ábra. "Modell életciklus" integrált modellezés alapú tudásmenedzsment során. A technológiafejlesztés minden lépése modellekre épül. E modellek az adott részproblémák megoldásán túl az egyes lépések közti információátvitel során is felhasználásra kerülnek. Amennyiben ez az információátvitel sikeres, joggal mondható el, hogy a koncepcionális tervezéstől a működő technológia optimalizálásáig a modell integrálja a szervezetet.

A kutatás elsődleges célja tehát olyan szürke-doboz modell (gray-box model, hybrid model) alapú kísérlettervezési és adatelemzési technikák kidolgozása volt, amelyek "csak" a fejlesztési folyamat aktuális lépésében jelentős problémakör nem ismert jelenségeit írják le adat-alapú fekete doboz modellek segítségével. E modellek identifikálásához, azaz a kísérleti eredményekben rejtett információk feltárásához új állapotbecslő és adatbányászati technikákat dolgoztunk ki, melyek a számítási intelligencia (fuzzy modellek, neurális hálózatok) legújabb kutatási eredményeit is magukba olvasztják.

A kidolgozott eszközök, módszerek

A pályázatban rögzítetteket maradéktalanul követve a kutató munka a következő két tématerületre fókuszált:

- A. Kísérlettervezés, azaz hipotézisek generálása és annak biztosítása, hogy kísérleteink a lehető legtöbb információt nyújtsák a hipotézisünkről, illetve a fejlesztés „keresési terében” a leghatékonyabban közelítsük meg a keresett optimumot.
- B. Mérési eredmények elemzése: azaz a mérési eredmények megbízhatóságának minősítése, új összefüggések feltárása, hipotézisek generálása.

Az (A) tématerülettel kapcsolatban egy új, szekvenciális, illetve párhuzamos kísérlettervezési techniká(ka)t dolgoztunk ki és publikáltunk (nemzetközi folyóiratokban). A kidolgozott algoritmus identifikációs problémák esetében a kísérletek információ tartalmát Evolúciós Stratégia és/vagy kvadratikus programozás segítségével maximalizálja. A szekvenciális kísérlettervezési technika előnyének a tervezett kísérletek egymásra épülésének biztosítása bizonyult. A hipotézisek generálása kapcsán a feltáró jellegű adatelemzés eszköztára mutatkozott rendkívül hasznosnak, melyet működő technológiák energetikai elemzésére is sikeresen felhasználtunk.

A (B) tématerületet már maga a kutatási terv is több részre osztotta.

B.1 Laboratóriumi és technológiai rendszerek szimulátorának kidolgozása és K+F tevékenységbe történő integrálása. Információ átvitel különböző modellek között az adat-alapú és hagyományos mérnöki ismeretek kombinációjaként felépülő „szürke doboz” modellek segítségével, azaz a fejlesztés aktuális szakaszában nem ismert jelenségek adat-alapú modellezése (pl. neurális hálózattal, fuzzy modellekkel). A kutatási tervben megfogalmazottakon túl rendkívül fontos, gyakorlati szempontból értékes eredményeket értünk el a technológia szabályozó rendszerének integrálása kapcsán (Balaskó, 2008, Moldoványi, 2008).

B.2 Üzemeltetési és tervezési változók és a termék tulajdonságok közti kapcsolat becslésére alkalmas modellek fejlesztése és alkalmazása. E téren elért eredményeink négy folyóiratcikkben publikáltuk, illetve a projektvezető által szerkesztett (tan)könyvben (Adatbányászat - a hatékonyság eszköze, Abonyi, 2006) is ismertettük. Az algoritmusok többsége fuzzy csoportosítási technikákon alapul. Ezeket az eredményeinket egy, a Birkhauser kiadó által megjelentetett monográfiában foglaltuk össze (Fuzzy Clustering for Data Mining and System Identification, Abonyi, 2007).

Az asszociációs szabályok és a gyakori elemhalmaz kereső algoritmusok alkalmazása rendkívül meglepő és hasznos lehetőségnek bizonyult. Ezen eredményeket két referált nemzetközi folyóiratcikkben és egy sikeresen megvédett PhD dolgozatban foglaltuk össze (Pach Ferenc munkái).

A modellek és a modellezési technikák közti átjárhatóságot Kenesei Tamás publikációiban tanulmányoztuk, újfajta értelmezését adva a Hinging Hiperplane, az SVM (Support Vector Machines) és a neurális hálózati modelleknek.

B.3 Idősor (process trend) elemzés nemlineáris, adat-alapú modellek segítségével, alkalmazás lehetséges üzemeltetési módok és hibák feltárásában. A kutatás során epizód alapú technikákat dolgoztunk ki. Az alapgondolat az volt, hogy a technológiai idősorok szegmentálását követően a tipikus epizódok, illetve epizód szekvenciák feltárása, és ezen szekvenciák részletes elemzése alkalmas az idősorokból a potenciálisan hasznos információk feltárására. Az eredmények Balaskó és Bankó (2007) publikációiban kerültek ismertetésre.

B.4 Optimalizációs technikák fejlesztése témakör olyan korszerű keresési eljárások alkalmazási lehetőségeinek vizsgálatát foglalta össze, amelyek alkalmasak a sok tervezési változó szimultán történő optimalására, különös tekintettel a többcélúságra és a műszaki korlátokra. Mivel célunk, hogy a szakértői ismeret és heurisztika az optimalizációs folyamatban is alkalmazható legyen, ezért modell alapú (szimplex, szekvenciális kvadratikus programozás és heurisztikus (pl. chemotaxis, genetikus algoritmus) algoritmusokat és ezek kombinációját interaktív és diagnosztikai funkciókkal egészítettük ki. E cél alapján e témakörhöz kapcsolódnak a folyamat- és adat-vizualizációval kapcsolatos kutatásaink, mely témakörben szintén sikerült publikációképes eredményeket elérnünk. Ezek közül Vathy gráf alapú adatmegjelenítési munkái bizonyultak a legértékesebbnek (két nemzetközi folyóiratcikk).

2007 novembere végén a projektben résztvevő holland partnernél munkamegbeszélést tartottunk a folyamatban lévő publikációk lezárásának és további kutatási ötletek, témák átbeszélésének céljából. E munka eredményeként hamarosan két nemzetközi folyóiratcikk lezárása és egy particle filter technika alkalmazásával kapcsolatos vizsgálat indulása várható.

Összefoglalás

E kidolgozott tudásfeltárási keretrendszer a folyamatmérnöki tudományok és az informatika fontos, aktuális rész-problémáinak megoldása során alakult ki, illetve fejlődik. A folyamatmérnökség, a rendszeridentifikáció, a számítási intelligencia és az adatbányászat eszköztárának talán szokatlanul tűnő kombinálásán alapuló fejlesztés eredményei azonban nem csupán ezeken a területeken alkalmazhatók, hanem akár a gazdasági életben a marketing és a folyamatmenedzsment feladatainak támogatásában, vagy szélesebb körű mérnöki, illetve tudományos feladatok megoldásában.

A publikációk mellett MATLAB programcsomagok kerültek előkészítésre, melyek némi finomítást követően elérhetőek a kutatócsoport és a Mathworks honlapján (www.fmt.uni-pannon.hu/softcomp és <http://www.mathworks.com/matlabcentral/fileexchange/7486>, illetve <http://www.mathworks.com/matlabcentral/fileexchange/authors/17446>).

Az OTKA támogatását a kutatás témavezetőjeként ezúton is köszönöm, megemlítve azt, hogy e támogatás közvetett módon két megvédett (Pach Ferenc Péter és Moldoványi Nóra), egy benyújtott (Fogarassyné Vathy Ágnes), és három előkészítés alatt álló (Balaskó Balázs, Kenesei Tamás és Varga Tamás) PhD dolgozat, egy sikeresen lezárult habilitációs cselekmény (Abonyi János) téziseinek illetve két könyv elkészítéséhez is hozzájárult.