

Bírálati vélemény

**Bántay László**

**„Frequent pattern and process mining-based process operation analysis” c.**

PhD értekezéséről

### **1. Témaválasztás, a téma aktualitása**

A dolgozatban a jelölt az ipari folyamatok elemzésére irányuló módszertanokat dolgoz ki, különös figyelmet szentelve a riasztások kezelésre, a párhuzamosan futó folyamatok azonosítására, a gyakori eseménysorozatok feltárására és az elemzések vizuális támogatására.

A dolgozat témája kifejezetten aktuális az Ipar 4.0 és 5.0 korában, ahol a digitalizáció és az automatizálás egyre nagyobb szerepet kap. Az ipari naplófájlok hatékony elemzése lehetőséget biztosít a termelési folyamatok optimalizálására, a biztonság növelésére és az operátori terhelés csökkentésére, amelyek kulcsfontosságúak a modern, intelligens gyárak működésében. A jelölt által fejlesztett módszerek által kinyert információk felhasználhatók prediktív rendszerek fejlesztéséhez, amelyek a jövőbeni eseménysorozatok előrejelzésével segítik a gyártási folyamatok optimalizálását és az ipari folyamatok hatékonyságának növelését.

### **2. A dolgozat szerkezete és nyelvezete**

A dolgozat angol nyelven íródott, tartalmi részét tekintve 96 oldal. Szerkezete a kapcsolódó publikációk alapján tagolt, melyből fakadóan némely részek (pl. az ipari alkalmazási példa bemutatása) kissé redundáns, de ez nem ró jelentős terhet az olvasóra. Nyelvezete és stílusa megfelelő. A dolgozat ábrái és táblázatai megfelelően támasztják alá az aktuális mondanivalót.

### **3. A szakmai fejezetekhez tartozó tartalmi észrevételek és kérdések**

#### **Introduction (Chapter 1) és Theoretical Background (Chapter 2)**

Az Introduction fejezet a később bemutatandó eredmények struktúrájában vezeti fel a dolgozatot, s ebből fakadóan jól strukturált, átfogó képet ad a dolgozat további tartalmára vonatkozóan.

A 2. fejezet a későbbi fejezetekhez szükséges alapfogalmakat és jelölésrendszert vezet be, különös figyelmet fordítva a szekvenciális mintázatok bányászata és a folyamatbányászat közti megfeleltethetőség és párhuzam bemutatására. Bár a szerző gondolatmenete követhető, a fejezet sajnos számos hibát tartalmaz a formalizmus tekintetében. Példaként említve: a 2.1 Táblázat-ban definiált Process Mining jelölések nincsenek összhangban a 10. oldal alján bevezetett jelölésekkel (lásd  $L$  és  $T$ ); a halmaz és a szekvencia jelölése nem következetes – sok helyen a szekvencia halmazként jelenik meg (pl. 2.8 egyenletben a Trace, vagy  $\phi_i$  és  $\phi_h$  definiálása a 2.2 egyenlet alatt); az indexkezelés nem következetes (pl.  $T_i$  és  $\phi_i$  definíciói a 2.1 egyenlet felett jobb oldalon

nem tartalmaznak  $i$ -re vonatkozó indexet; a 2.3 egyenletben az  $i$  index vélhetően a szekvencia hosszára is utal, holott ez nem lett definiálva).

Bár a formalizmusban előforduló hibák és pontatlanságok nem akadályozzák a témában jártas olvasót a szerző gondolatmenetének követésében, a fejezet szakszerűsége és tudományos igényessége szempontjából kívánatos lett volna egy következetesebb és precízebb jelölésrendszer alkalmazása.

### Chapter 3: Process mining of industrial log files

A fejezetben a jelölt olyan módszertant mutat be, amelynek segítségével az ipari alarm- és eseménynaplók (alarm & event log files) folyamatbányászatra alkalmassá tehető. A megközelítés célja a riasztásmenedzsment támogatása, a kezelői beavatkozások mélyebb elemzése, valamint a folyamatbiztonság növelése. A módszertan alapját az eseménynaplókban szereplő adatokból származó eseménysorozatok (trace-ek) azonosítása képezi, különös tekintettel a riasztások és az operátori beavatkozások közötti kapcsolatok feltárására. A trace-ek meghatározása a bejegyzések időbélyege, az eseménytípusok, valamint az események közötti időintervallumok alapján történik.

Bár a 3. fejezetben bemutatott módszertan újszerű és releváns ipari alkalmazásokat céloz, a fejezet egyik lényeges hiányossága az irodalmi kontextus részletes feltárásának elmaradása. A szerző nem tér ki arra, hogy a nemzetközi szakirodalomban milyen korábbi megközelítések léteznek az ipari alarm- és eseménynaplók folyamatbányászati célú előfeldolgozására, különösen a trace-generálás tekintetében.

A jelölt a kidolgozott módszer gyakorlati alkalmazhatóságát egy valós ipari esettanulmányon keresztül mutatja be. A gyakorlati példa megfelelően szemlélteti a kidolgozott módszertan alkalmazhatóságát.

Kérdés: A kidolgozott módszertan egyik kulcseleme a  $dt_w$  időablak méretének meghatározása, amely meghatározza a trace-ek elkülönítését. A szerző ezt empirikus alapon közelíti meg: az esettanulmányban alkalmazott 220 másodperces küszöbértéket az alarm események kezdési időpontjai közötti időintervallumok mediánja alapján választotta meg. Ez a megközelítés azonban több kérdést is felvet. Implicit módon azt feltételezi, hogy az eseményközök fele külön trace-eket határoz meg, ami nem feltétlenül tükrözi a valós eseményláncolatok szerkezetét. Véleménye szerint hogyan lehetne ezt a megközelítést finomítani? Mit gondol például egy, az eseményközök gyakorisági eloszlása alapján történő „rés”-detektálási technika alkalmazásáról?

### Chapter 4: Sequence compression and alignment-based process alarm prediction

A fejezet egy olyan riasztás-előrejelző eljárást ismertet, amely korábbi eseménynaplók feldolgozása révén gyakran előforduló eseményszekvenciákat azonosít, majd az aktuális folyamat során bekövetkező eseményláncokat ezekhez a mintákhoz illeszti előrejelzési célból. A szerző a módszer alkalmazhatóságát egy valós ipari környezetből származó esettanulmányon keresztül demonstrálja, melynek során különböző teljesítménymutatókkal értékeli az algoritmus előrejelző képességét.

A javasolt módszer dokumentáltsága megfelelő, a bemutatott példák világosan szemléltetik a mögöttes elméleti megfontolásokat. Véleményem szerint a kidolgozott algoritmus a gyakorlatban is jól alkalmazható, amit a jelölt meggyőzően mutat be az ipari esettanulmányon keresztül. A

példa impozáns, a módszer hatékonyságának értékelése kiegyensúlyozott és kritikus szemléletű, a szerző pedig megfelelően reflektál az alkalmazás korlátjaira is.

Sajnálatos módon azonban, jelen fejezetből is hiányolom a javasolt módszer pozícionálását a szakirodalomba, valamint az eredmények összevetését más módszerek hatékonyságával. Egy benchmark jellegű összehasonlítás (pl. PrefixSpan, SPADE) értékes kiegészítése lett volna a fejezetnek, különösen a módszer gyakorlati hatékonyságának alátámasztásához.

Kérdés: A gyakori szekvenciák kinyerésére alkalmazott GoKimp algoritmus előnye, hogy nem igényel előzetesen megadott paramétereket, mivel a mintakiválasztást tömörítési nyereség alapján végzi. Ez az „önhangoló” működés egyes esetekben előnyt jelenthet, ugyanakkor az automatizmusból fakadóan a felhasználói kontroll csökkenése potenciális kockázatot is hordoz. Kérdésem, hogy milyen típusú logfájl-struktúrák vagy alkalmazási környezetek esetén ajánlott különösen a GoKimp algoritmus alkalmazása, illetve milyen jellegű hibák vagy torzítások kockázatával kell számolni az algoritmus használatakor?

## Chapter 5: Frequent pattern mining-based log file partitioning

A fejezet egy új módszert mutat be a több folyamatot tartalmazó ipari eseménynaplók strukturálására: a javasolt módszer a gyakori mintázatbányászatot és a hagyományos folyamatelemzési kockaműveleteket ötvözi oly módon, hogy iteratívan szűri ki a szükségtelen eseményeket és hoz létre résznaplókat. A cél az, hogy az eseménynaplókban ismétlődő, strukturált viselkedési egységeket különítsen el, ezzel támogatva a további bányászati, vizualizációs vagy diagnosztikai lépéseket.

A javasolt módszer előnye az adatvezérelt jellegében rejlik, azonban éppen ebből fakadóan általánosan alkalmazható paraméterbeállításokra nem lehet javaslatot tenni és teljesen automatizálható módon nem hajtható végre. A *MinSup* érték beállítása például véleményem szerint minden esetben manuális hangolást igényel. Ugyancsak kritikus lépés az irreleváns események kiszűrése.

A javasolt módszer alkalmazhatóságát a jelölt a korábbi fejezetekben látható ipari példán keresztül szemléltette. Az esettanulmány ismertetésekor az volt az az érzésem, hogy a módszertan valóban „goal oriented”, s alkalmazása nagy mértékben függ az alkalmazó szemléletétől. Ugyanakkor szeretném kiemelni az eredményezett folyamatmodellek kvantitatív kiértékelését, amely objektív módon is alátámasztja a részmodellek helyességét és felhasználhatóságát.

Kérdés: A 3. fejezetben bemutatott eljárások célja az előfeldolgozás és a releváns eseményláncok azonosítása, míg az 5. fejezetben a logfájl gyakori minták mentén történő strukturálása történik meg. Véleménye szerint e két módszertani megközelítés integrálható-e egy közös módszertanba, és ha igen, milyen szinergiák vagy nehézségek adódhatnak a kombinálásukból?

## Chapter 6: Network-based visualisation of frequent sequences

A 6. fejezet 3 gráf alapú vizualizációs módszert javasol a folyamatok feltárásának támogatására oly módon, hogy a vizualizációs módszerek alapját a szekvenciák gráfalapú ábrázolása képezi. A három javasolt módszer alapjaiban megegyezik (a szerző is csak 1 módszerként hivatkozik rá), a gráfok struktúrájának kialakításához viszont a gyakori szekvenciák közötti kapcsolatok eltérő

típusú és mélységű relációi kerülnek figyelembevételre (rákövetkezési valószínűség, szekvenciák tranzitív megbízhatósága, támogató tranzakciók Jaccard hasonlósága).

A javasolt módszerek jól dokumentáltak, s véleményem szerint jól alkalmazhatók a gyakorlatban is, mivel különféle aspektusból nyújtanak rálátást a folyamatokban megjelenő gyakori szekvenciákra. Külön pozitívum, hogy a javasolt vizualizációs technikák alkalmazhatóságát a jelölt két mintapéldán keresztül is szemléltette.

Kérdés: A 4-es algoritmusban, ha a  $\text{TransDist}(A)$  tranzitív távolságokat ad vissza, akkor a következő lépésben miért szükséges ezen értékeket 1-ből kivonni? A kivonást követően ugyanis hasonlósági értéket kapnánk vissza, s nem különbözőség értékeket; holott az MDS távolságmátrixot vár bemenetként.

## **Chapter 7: Machine Learning-supported designing of Human-Machine Interfaces**

A fejezet a Human-Machine Interface (HMI) tervezés gyakori mintázatok bányászati és folyamatbányászati alapú támogatására tesz javaslatot. A javasolt módszer általános megközelítést, módszertani lépéseket fogalmaz meg a HMI tervezésre vonatkozóan, az egyes lépések megvalósítási részleteit a későbbi alkalmazókra bízta. A módszer alkalmazhatóságát a jelölt újfent a korábban ismertetett ipari környezetben szemlélteti.

Úgy vélem a javasolt módszer valóban hatékonyan támogathatja a kijelzők tartalmának tervezését, azonban ahogy a szerző is rámutat, a módszer számos kontrollt, és szakértői kiegészítést igényel. Emellett fontos kiemelni, hogy a módszertan alkalmazása komoly jártasságot igényel a szekvenciabányászatban, a folyamatbányászatban és a csoportosítás témakörében.

### **4. Tézisek**

Az eredményekhez kapcsolódóan a jelölt 4 tézist fogalmaz meg. A tézisek a publikációk révén megfelelően alátámasztottak.

A tézisek szakmailag megalapozottak, koherensen illeszkednek a dolgozat fő célkitűzéseire, és gyakorlati példákon keresztül jól demonstráltak. Véleményem szerint a 2-4 tézisek megfogalmazása megfelelő, az első tézis azonban tényszerűségében kissé javítható lett volna. A munka eredetisége, ipari alkalmazhatósága és módszertani konzisztenciája miatt a tézisek érvényesnek és védhetőnek tekinthetők.

### **5. Eredmények publikálása**

A jelölt az eredményeit 5 publikációban tette közzé, melyek mindegyikében első szerzőként szerepel. A publikációk közül 4 db Q1-es, 1 db pedig Q2-es besorolású. Ezáltal a jelölt messzemenően teljesíti a Vegyész-mérnöki- és Anyagtudományok Doktori Iskola publikációs követelményeit. Mivel minden tézisponthoz rangos publikáció tartozik, ezért a tézispontok is megfelelően alátámasztottak.

## 6. Összegzés

Összességében megállapítható, hogy Bántay László PhD disszertációja értékes, új tudományos eredményekkel járul hozzá a folyamatbányászat, folyamatanalízis és alarm management témaköréhez. A dolgozat jelentősége abban rejlik, hogy gyakorlati ipari problémákra ad adataalapú válaszokat, különös figyelmet fordítva a diszkrét folyamatesemények és folyamatszekvenciák strukturálására, előrejelzésére és vizualizálására. A tudományos eredményeket a jelölt a Doktori Iskola követelményrendszerének megfelelően, a minimális követelményeket jelentősen meghaladó szinten publikálta.

**A bemutatott eredményeket a jelölt saját eredményeinek fogadom el. Javaslom az értekezés nyilvános vitára bocsátását, és sikeres védés esetén támogatom a PhD fokozat odaítélését a jelölt számára.**

2025. június 05.



Dr. Fogarassyné dr. Vathy Ágnes  
tanszékvezető egyetemi docens  
Pannon Egyetem  
Rendszer- és Számítástudományi Tanszék

