

Process Mining Manifesto

Manifest je „javna deklaracija principa i namera“ od strane grupe ljudi. Ovaj manifest napisali su članovi i podržavaoci IEEE Radne grupe za rudarenje procesa (*IEEE Task Force on Process Mining*). Cilj ove radne grupe jeste promovisanje istraživanja, razvoja, obrazovanja, implementacije, evolucije i razumevanja rudarenja procesa.

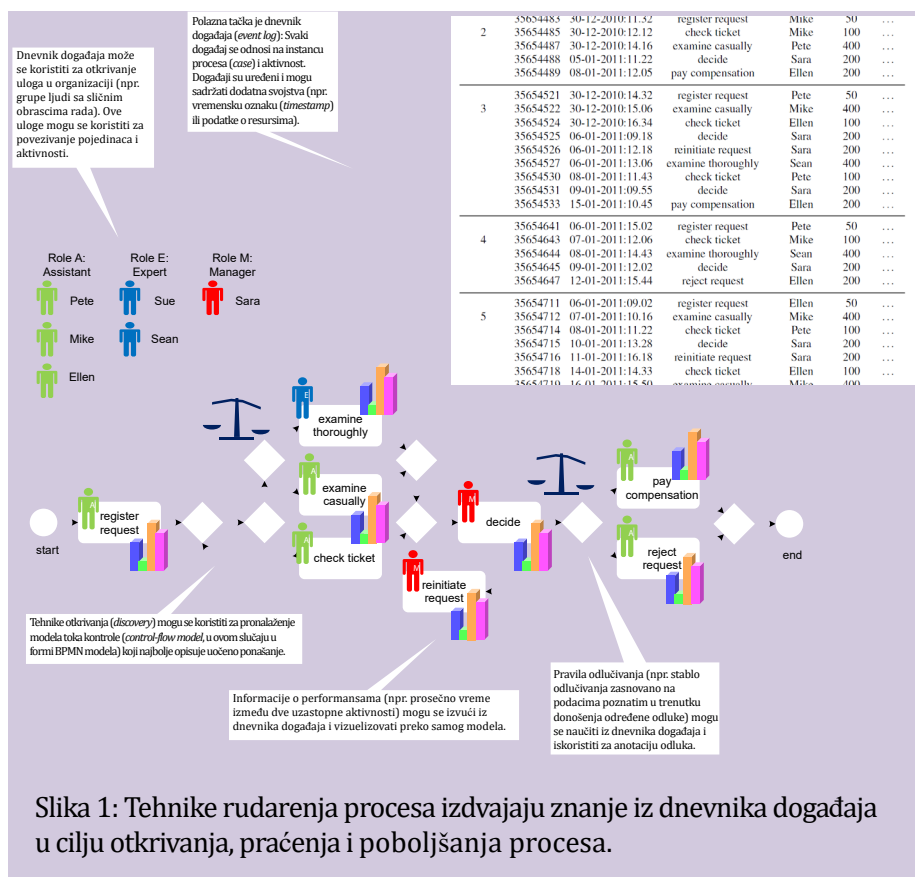
Rudarenje procesa (*Process mining*) relativno je mlada istraživačka disciplina koja se nalazi između računske inteligencije i rudarenja podataka (*data mining*) sa jedne strane, i modelovanja i analize procesa sa druge strane. Ideja rudarenja procesa jeste otkrivanje, praćenje i poboljšanje stvarnih procesa (tj. procesa koji nisu pretpostavljeni) izdvajanjem znanja iz dnevnika događaja (*event logs*) koji su danas lako dostupni u (informacionim) sistemima. Rudarenje procesa obuhvata (automatizovano) otkrivanje procesa (*process discovery* — tj. izdvajanje modela procesa iz dnevnika događaja), proveru usklađenosti (*conformance checking* — tj. praćenje odstupanja upoređivanjem modela i dnevnika), rudarenje društvenih mreža / organizaciono rudarenje, automatizovanu konstrukciju simulacionih modela, proširenje modela,

popravku modela, predviđanje ishoda instanci (*case prediction*) i preporuke zasnovane na istorijatu izvođenja.

Sadržaj:

Rudarenje procesa – Stanje u oblasti	3
Vodeći principi	6
Izazovi	10
Epilog	13
Rečnik pojmova	14

Tehnike rudarenja procesa sposobne su da izdvoje znanje iz dnevnika događaja koji su uobičajeno dostupni u današnjim informacionim sistemima. Ove tehnike pružaju nova sredstva za otkrivanje, praćenje i poboljšanje procesa u širokom spektru oblasti primene. Dva su glavna pokretača rastućeg interesovanja za rudarenje procesa: s jedne strane, beleži se sve više događaja, čime se obezbeđuju detaljne informacije o istorijatu procesa; s druge strane, postoji potreba za poboljšanjem i podrškom poslovnim procesima u konkurentnim i brzorastućim promenljivim okruženjima. Ovaj manifest kreirala je IEEE Radna grupa za rudarenje procesa sa ciljem da promoviše ovu temu. Štaviše, definisanjem skupa vodećih principa i popisivanjem važnih izazova, ovaj manifest teži da posluži kao vodič za softverske inženjere, naučnike, konsultante, poslovne menadžere i krajnje korisnike. Cilj je podizanje nivoa zrelosti rudarenja procesa kao novog alata za unapređenje (re)dizajna, kontrole i podrške operativnim poslovnim procesima.



Rudarenje procesa obezbeđuje važan most između rudarenja podataka (*data mining*) i modelovanja i analize poslovnih procesa. Pod krovim pojmom poslovne inteligencije (*Business Intelligence – BI*) uvedene su mnoge popularne kovanice koje se zapravo odnose na prilično jednostavne alate za izveštavanje i kontrolne table (*dashboards*). Praćenje poslovnih aktivnosti (*Business Activity Monitoring – BAM*) odnosi se na tehnologije koje omogućavaju praćenje poslovnih procesa u realnom vremenu.

Konkretni ciljevi radne grupe:

- 1) Upoznavanje krajnjih korisnika, programera, konsultanata, poslovnih menadžera i istraživača sa stanjem u oblasti rudarenja procesa.
- 2) Promovisanje upotrebe tehnika i alata za rudarenje procesa i podsticanje novih primena.
- 3) Učešće u naporima za standardizaciju evidentiranja podataka o događajima.
- 4) Organizovanje tutorijala, specijalnih sesija, radionica i panela.
- 5) Objavljivanje članaka, knjiga, video-materijala i specijalnih izdanja časopisa.

Obrada složenih događaja (*Complex Event Processing – CEP*) odnosi se na tehnologije za obradu velikih količina događaja, koje se koriste za praćenje, usmeravanje i optimizaciju poslovanja u realnom vremenu. Upravljanje korporativnim performansama (*Corporate Performance Management – CPM*) još je jedan termin za merenje performansi procesa ili organizacije. Takođe su povezani i menadžerski pristupi kao što su kontinuirano poboljšanje procesa (*Continuous Process Improvement – CPI*), poboljšanje poslovnih procesa (*Business Process Improvement – BPI*), upravljanje totalnim kvalitetom (*Total Quality Management – TQM*) i Šest sigma (*Six Sigma*). Ovi pristupi imaju zajedničku crtu da se procesi „stavljaju pod mikroskop“ kako bi se videlo da li su moguća dalja poboljšanja. Rudarenje procesa predstavlja tehnologiju koja omogućava i podržava CPM, BPI, TQM, Six Sigma i slične pristupe.

Dok BI alati i menadžerski pristupi poput Six Sigma i TQM imaju za cilj poboljšanje operativnih performansi (npr. smanjenje vremena protoka i defekata), organizacije takođe stavljaju sve veći naglasak na korporativno upravljanje, rizike i usklađenost (*corporate governance, risks, and compliance*). Zakonski propisi kao što su Sarbanes-Oxley Act (SOX) i sporazum Bazel II ilustruju fokus na pitanja usklađenosti. Tehnike rudarenja procesa nude način

da se rigoroznije proveriti usklađenost i utvrdi validnost i pouzdanost informacija o ključnim procesima organizacije.

Tokom protekle decenije, podaci o događajima postali su lako dostupni, a tehnike rudarenja procesa su sazrele. Štaviše, kao što je pomenuto, trendovi upravljanja koji se odnose na poboljšanje procesa (npr. Six Sigma, TQM, CPI i CPM) i usklađenost (SOX, BAM, itd.) mogu imati velike koristi od rudarenja procesa. Na sreću, algoritmi za rudarenje procesa implementirani su u različitim akademskim i komercijalnim sistemima. Danas postoji aktivna grupa istraživača koji rade na rudarenju procesa i to je postala jedna od „vrućih tema“ u istraživanjima upravljanja poslovnim procesima (*Business Process Management – BPM*). Pored toga, postoji ogromno interesovanje industrije. Sve više proizvođača softvera dodaje funkcionalnosti rudarenja procesa u svoje alate. Primeri softverskih proizvoda sa mogućnostima rudarenja procesa su: ARIS Process Performance Manager (Software AG), Comprehend (Open Connect), Discovery Analyst (StereoLOGIC), Flow (Fourspark), Futura Reflect (Futura Process Intelligence), Interstage Automated Process Discovery (Fujitsu), OKT Process Mining suite (Exeura), Process Discovery Focus (Iontas/Verint), ProcessAnalyzer (QPR), ProM (TU/e), Rbminer/Dbminer (UPC) i Reflect|one (Pallas Athena). Rastuće interesovanje za analizu procesa zasnovanu na dnevnicima motivisalo je osnivanje Radne grupe za rudarenje procesa.

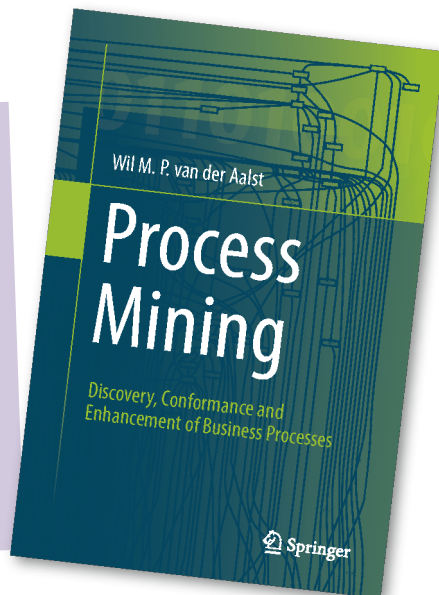
Radna grupa je osnovana 2009. godine u okviru Tehničkog komiteta za rudarenje podataka (*Data Mining Technical Committee – DMTC*) Društva za računsku inteligenciju (*Computational Intelligence Society – CIS*) pri Institutu inženjera elektrotehnike i elektronike (IEEE). Trenutna radna grupa ima članove koji predstavljaju proizvođače softvera (npr. Pallas Athena, Software AG, Futura Process Intelligence, HP, IBM, Infosys, Fluxicon, Businesscape, Iontas/Verint, Fujitsu, Fujitsu Laboratories, Business Process Mining, Stereologic), konsultantske kuće / krajnje korisnike (npr. ProcessGold, Business Process Trends, Gartner, Deloitte, Process Sphere, Siav SpA, BPM Chile, BWI Systeme GmbH, Excellentia BPM, Rabobank) i istraživačke institute (npr. TU/e, Univerzitet u Padovi, Univerzitet Politècnica de Catalunya, New Mexico State University, IST – Tehnički univerzitet u Lisabonu, Univerzitet u Kalabriji, Penn State

University, Univerzitet u Bariju, Humboldt-Universität zu Berlin, Queensland University of Technology, Univerzitet ekonomije i biznisa u Beču, Stevens Institute of Technology, Univerzitet u Haifi, Univerzitet u Bolonji, Ulsan National Institute of Science and Technology, Cranfield University, K.U. Leuven, Tsinghua University, Univerzitet u Inzbruku, Univerzitet u Tartuu).

Od svog osnivanja 2009. godine sprovedene su različite aktivnosti u vezi sa navedenim ciljevima. Na primer, radna grupa je (ko)organizovala nekoliko radionica i specijalnih sekcija, npr. radionice o inteligenciji poslovnih procesa (BPI'09, BPI'10 i BPI'11) i specijalne sekcije na glavnim IEEE konferencijama (npr. CIDM'11). Znanje je šireno putem tutorijala (npr. WCCI'10 i PMPM'09), letnjih škola (ESSCASS'09, ACPN'10, CICH'10 itd.), video-zapisa (v. www.processmining.org) i nekoliko publikacija, uključujući prvu knjigu o rudarenju procesa koju je nedavno objavio Springer. Radna grupa je takođe (ko)organizovala i prvi *Business Process Intelligence Challenge (BPIC'11)*: takmičenje u kojem su učesnici morali da izvuku smisleno znanje iz velikog i složenog dnevnika događaja. Godine 2010. radna grupa je takođe standardizovala XES (www.xes-standard.org), proširivi standardni format za beleženje dnevnika podržan od strane OpenXES biblioteke (www.openxes.org) i alata kao što su ProM, XESame, Nitro itd.

Process Mining Book

www.processmining.org/book/
W.M.P. van der Aalst. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer-Verlag, Berlin, 2011.



Čitalac se poziva da poseti www.win.tue.nl/ieetfpm/ radi više informacija o aktivnostima radne grupe.

2. Rudarenje procesa: Stanje u oblasti (*State of the Art*)

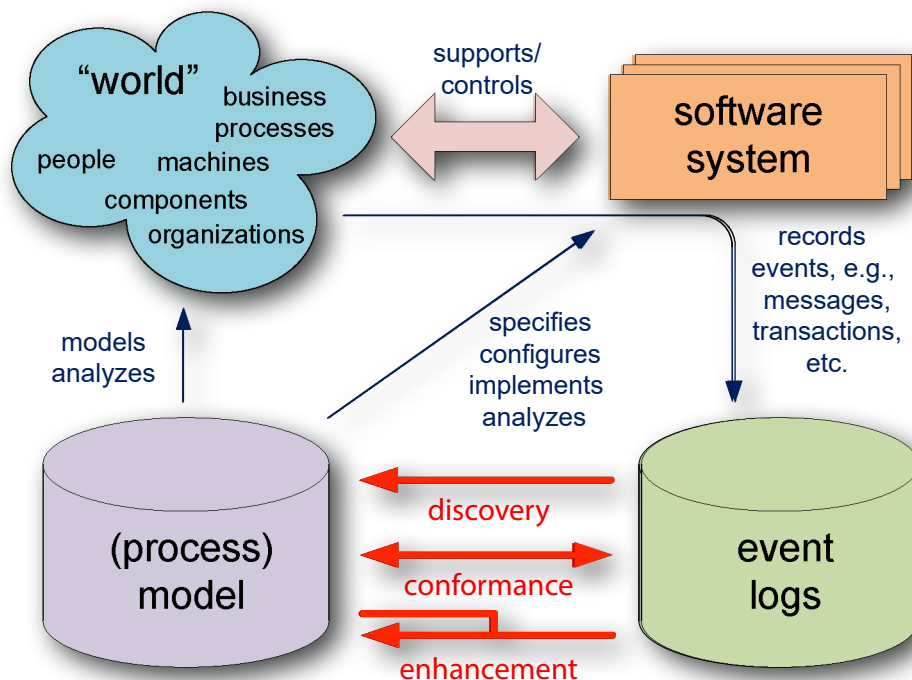
Rastuće mogućnosti informacionih sistema i drugih sistema koji zavise od računarstva dobro su opisane Murovim zakonom. Gordon Mur, suosnivač Intela, predvideo je 1965. godine da će se broj komponenti u integrisanim kolima udvostručavati svake godine. Tokom poslednjih pedeset godina rast je zaista bio eksponencijalan, doduše nešto sporijim tempom. Ovaj napredak

rezultirao je spektakularnim rastom „digitalnog univerzuma“ (tj. svih podataka koji se elektronski čuvaju i/ili razmenjuju). Štaviše, digitalni i stvarni univerzum postaju sve više usklađeni.

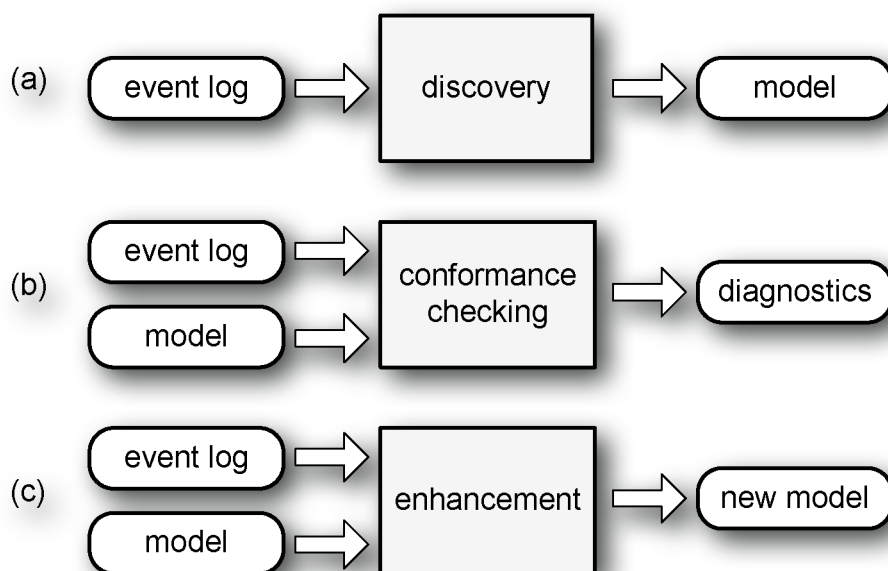
Rast digitalnog univerzuma koji je usklađen sa procesima u organizacijama omogućava beleženje i analizu događaja. Događaji mogu varirati od podizanja gotovine sa bankomata, lekara koji podešava rendgen aparat, građanina koji podnosi zahtev za vozačku dozvolu, podnošenja poreske prijave, do prijema broja elektronske karte od strane putnika. Izazov je iskoristiti podatke o događajima na smislen način, na primer, za pružanje uvida, identifikovanje uskih grla, predviđanje problema, evidentiranje kršenja pravila, preporučivanje protivmera i pojednostavljenje procesa. Rudarenje procesa ima za cilj upravo to.

Polazna tačka za rudarenje procesa jeste dnevnik događaja. Sve tehnike rudarenja procesa pretpostavljaju da je moguće sekvencijalno beležiti događaje tako da se svaki događaj odnosi na aktivnost (tj. dobro definisan korak u nekom procesu) i da je povezan sa određenom instancom procesa (case). Dnevnik događaja mogu čuvati dodatne informacije o događajima. Zapravo, kad god je moguće, tehnike rudarenja procesa koriste dodatne informacije kao što su resurs (tj. osoba ili uređaj) koji izvršava ili pokreće aktivnost, vremenska oznaka događaja ili elementi podataka zabeleženi uz događaj (npr. veličina porudžbine).

Kao što je prikazano na Slici 2, dnevnik događaja mogu se koristiti za sprovođenje tri tipa rudarenja procesa. Prvi tip rudarenja procesa jeste otkrivanje (*discovery*). Tehnika otkrivanja uzima dnevnik događaja i proizvodi model bez korišćenja bilo kakvih



Slika 2: Pozicioniranje tri glavna tipa rudarenja procesa: (a) otkrivanje (*discovery*), (b) provera usklađenosti (*conformance checking*) i (c) unapređenje (*enhancement*).



Slika 3: Tri osnovna tipa rudarenja procesa objašnjena u terminima ulaza i izlaza: (a) otkrivanje, (b) provera usklađenosti i (c) unapređenje.

a-priori informacija. Otkrivanje procesa je najistaknutija tehnika rudarenja procesa. Za mnoge organizacije je iznenađujuće kada vide da su postojeće tehnike zaista u stanju da otkriju stvarne procese samo na osnovu primera izvršavanja u dnevnicima događaja. Drugi tip rudarenja procesa jeste usklađenost (*conformance*). Ovdje se postojeći model procesa upoređuje sa dnevnikom događaja istog procesa. Provera usklađenosti može se koristiti za proveru da li stvarnost, zabeležena u dnevniku, odgovara modelu i obrnuto. Treba napomenuti da se mogu razmatrati različiti tipovi modela: provera usklađenosti može se primeniti na proceduralne modele, organizacione modele, deklarativne modele procesa, poslovna pravila/politike, zakone itd. Treći tip rudarenja procesa jeste unapređenje (*enhancement*). Ovdje je ideja proširiti ili poboljšati postojeći model procesa korišćenjem informacija o stvarnom procesu zabeleženih u nekom dnevniku događaja. Dok provera usklađenosti meri stepen slaganja između modela i stvarnosti, ovaj treći tip rudarenja procesa ima za cilj promenu ili proširenje *a-priori* modela. Na primer, korišćenjem vremenskih oznaka u dnevniku događaja model se može proširiti kako bi prikazao uska grla, nivoe usluga, vremena protoka i frekvencije.

Slika 3 opisuje tri tipa rudarenja procesa kroz ulaze i izlaze. Tehnike otkrivanja uzimaju dnevnik događaja i kreiraju model. Otkriveni model je tipično model procesa (npr. Petrijeva mreža, BPMN, EPC ili UML dijagram aktivnosti), međutim, model može

opisivati i druge perspektive (npr. društvenu mrežu). Tehnike provere usklađenosti zahtevaju dnevnik događaja i model kao ulaz. Izlaz se sastoji od dijagnostičkih informacija koje prikazuju razlike i zajedničke tačke između modela i dnevnika. Tehnike za unapređenje modela (popravka ili proširenje) takođe zahtevaju dnevnik događaja i model kao ulaz. Izlaz je poboljšani ili prošireni model.

Rudarenje procesa može pokriti različite perspektive. Perspektiva toka kontrole (*control-flow perspective*) fokusira se na tok kontrole, tj. na redosled aktivnosti. Cilj rudarenja ove perspektive jeste pronalaženje dobre karakterizacije svih mogućih putanja. Rezultat se obično izražava u vidu Petrijeve mreže ili neke druge procesne notacije (npr. EPC, BPMN ili UML dijagrama aktivnosti). Organizaciona perspektiva fokusira se na informacije o resursima skrivenim u dnevniku, tj. koji su akteri (npr. ljudi, sistemi, uloge ili odeljenja) uključeni i kako su međusobno povezani. Cilj je ili strukturirati organizaciju klasifikovanjem ljudi po ulogama i organizacionim jedinicama, ili prikazati društvenu mrežu. Perspektiva instance procesa (*case perspective*) fokusira se na svojstva instanci. Očigledno, slučaj se može okarakterisati svojom putanjom u procesu ili akterima koji na njemu rade. Međutim, slučajevi se takođe mogu okarakterisati vrednostima odgovarajućih elemenata podataka. Na primer, ako slučaj predstavlja porudžbinu za popunu zaliha, može biti zanimljivo znati dobavljača ili broj naručenih proizvoda. Vremenska

Karakteristike rudarenja procesa:

1. Rudarenje procesa nije ograničeno na otkrivanje toka kontrole.

Otkrivanje modela procesa iz dnevnika događaja raspiruje maštu i praktičara i akademske zajednice. Zbog toga se otkrivanje toka kontrole često posmatra kao najuzbudljiviji deo rudarenja procesa. Međutim, rudarenje procesa nije ograničeno na to. S jedne strane, otkrivanje je samo jedan od tri osnovna oblika rudarenja procesa (otkrivanje, usklađenost i unapređenje). S druge strane, obuhvat nije ograničen na tok kontrole; organizaciona, procesna i vremenska perspektiva takođe igraju važnu ulogu.

2. Rudarenje procesa nije samo specifičan tip rudarenja podataka.

Rudarenje procesa može se posmatrati kao „karika koja nedostaje“ između rudarenja podataka i tradicionalnog BPM-a vođenog modelima. Većina tehnika rudarenja podataka uopšte nije procesno usmerena. Modeli procesa koji potencijalno pokazuju konkurentnost (*concurrency*) neuporedivi su sa jednostavnim strukturama rudarenja podataka kao što su stabla odlučivanja i asocijativna pravila. Zbog toga su potrebni potpuno novi tipovi reprezentacija i algoritama.

3. Rudarenje procesa nije ograničeno na oflajn analizu.

Tehnike rudarenja procesa izdvajaju znanje iz istorijskih podataka o događajima. Iako se koriste „*post mortem*“ podaci, rezultati se mogu primeniti na tekuće slučajeve u realnom vremenu. Na primer, vreme završetka delimično obrađene narudžbine kupca može se predvideti korišćenjem otkrivenog modela procesa.

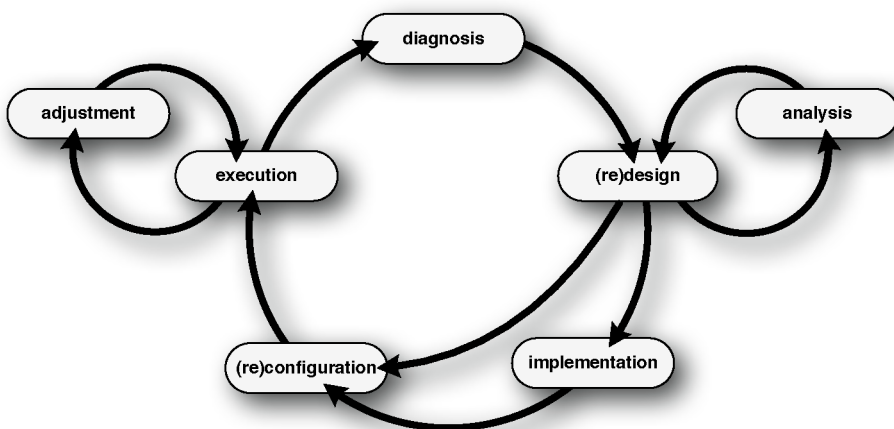
perspektiva bavi se vremenom i učestalošću događaja. Kada događaji nose vremenske oznake, moguće je otkriti uska grla, meriti nivoe usluga, pratiti iskorišćenost resursa i predvideti preostalo vreme obrade tekućih slučajeva.

Postoje određene uobičajene zablude u vezi sa rudarenjem procesa. Neki proizvođači, analitičari i istraživači ograničavaju delokrug rudarenja procesa na posebnu tehniku rudarenja podataka za otkrivanje procesa koja se može koristiti samo za oflajn analizu. To nije tačno, i zato naglašavamo tri karakteristike navedene u okviru na prethodnoj strani.

Da bismo pozicionirali rudarenje procesa, koristimo životni ciklus upravljanja poslovnim procesima (BPM) prikazan na Slici 4. BPM životni ciklus prikazuje sedam faza poslovnog procesa i odgovarajućeg informacionog sistema. U fazi (re)dizajna kreira se novi model procesa ili se prilagođava postojeći. U fazi analize analiziraju se kandidatski model i njegove alternative. Nakon faze (re)dizajna, model se implementira (faza implementacije) ili se postojeći sistem rekonfiguriše (faza rekonfiguracije). U fazi izvršavanja projektovani model se sprovodi u delo. Tokom faze izvršavanja proces se nadgleda. Štaviše, mogu se izvršiti manja podešavanja bez ponovnog redizajniranja procesa (faza podešavanja). U fazi dijagnoze analizira se izvršeni proces, a izlaz ove faze može pokrenuti novu fazu redizajna. Rudarenje procesa je vredan alat za većinu faza prikazanih na Slici 4. Očigledno je da faza dijagnoze ima koristi od rudarenja procesa. Međutim, rudarenje procesa nije ograničeno na fazu dijagnoze. Na primer, u

fazi izvršavanja tehnike rudarenja procesa mogu se koristiti za operativnu podršku. Predviđanja i preporuke zasnovane na modelima naučenim korišćenjem istorijskih informacija mogu se koristiti za uticaj na tekuće slučajeve. Slični oblici podrške odlučivanju mogu se koristiti za prilagođavanje procesa i vođenje (re) konfiguracije.

Dok Slika 4 prikazuje opšti BPM životni ciklus, Slika 5 se fokusira na konkretne aktivnosti i artefakte rudarenja procesa. Slika 5 opisuje moguće faze u projektu rudarenja procesa. Svaki projekat rudarenja procesa počinje planiranjem i opravdanjem tog planiranja (Faza 0). Nakon pokretanja projekta, podaci o događajima, modeli, ciljevi i pitanja moraju se izvući iz sistema, od stručnjaka iz domena i menadžmenta (Faza 1). Ovo zahteva razumevanje dostupnih podataka („Koji podaci se mogu koristiti za analizu?“) i razumevanje domena („Koja su važna pitanja?“) i rezultira artefaktima prikazanim na Slici 5 (tj. istorijski podaci, ručno izrađeni modeli, ciljevi i pitanja). U Fazi 2 konstruiše se model toka kontrole i povezuje sa dnevnikom događaja. Ovde se mogu koristiti tehnike automatskog otkrivanja procesa. Otkriveni model procesa već može pružiti odgovore na neka od pitanja i pokrenuti akcije redizajna ili podešavanja. Štaviše, dnevnik događaja može se filtrirati ili prilagoditi korišćenjem modela (npr. uklanjanjem retkih aktivnosti ili ekstremnih vrednosti slučajeva i ubacivanjem događaja koji nedostaju). Ponekad su potrebni značajni naponi da se korelišu događaji koji pripadaju istoj instanci procesa. Preostali događaji se povezuju sa entitetima



Slika 4: BPM životni ciklus koji identifikuje različite faze poslovnog procesa i odgovarajućih informacionih sistema; rudarenje procesa (potencijalno) igra ulogu u svim fazama (osim u fazi implementacije).

Vodeći principi:

VP1: Podaci o događajima moraju se tretirati kao građani prvog reda

VP2: Izdvajanje dnevnika mora biti vođeno pitanjima

VP3: Konkurentnost, izbor i druge osnovne konstrukcije toka kontrole moraju biti podržani

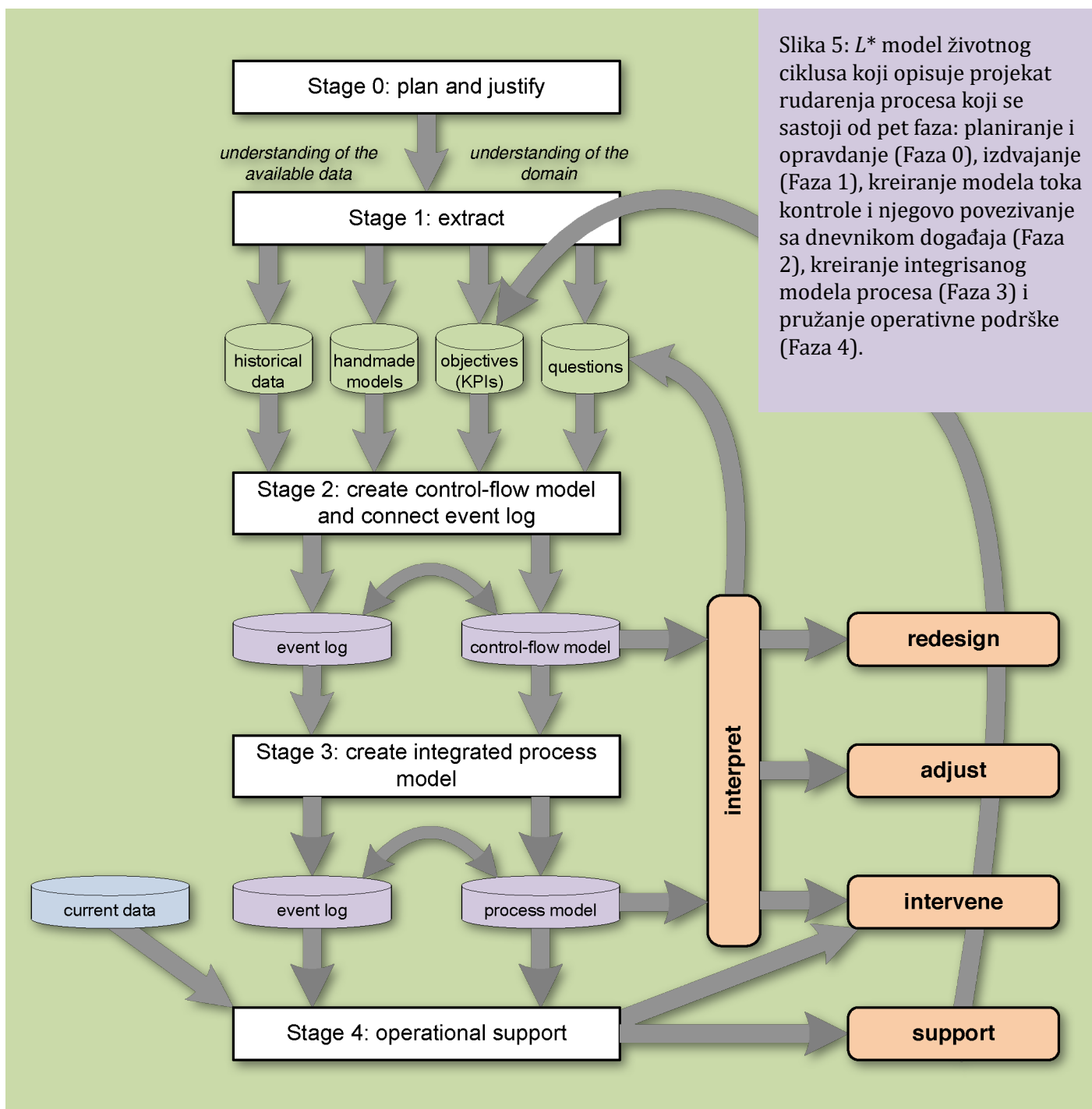
VP4: Događaji moraju biti povezani sa elementima modela

VP5: Modeli se moraju tretirati kao svrsishodne apstrakcije stvarnosti

VP6: Rudarenje procesa treba da bude kontinuiran proces

modela procesa. Kada je proces relativno strukturisan, model toka kontrole može se proširiti drugim perspektivama (npr. podaci, vreme i resursi) tokom Faze 3. Veza između dnevnika događaja i modela uspostavljena u Fazi 2 koristi se za proširenje modela (npr. vremenske oznake povezanih događaja koriste se za procenu vremena čekanja na aktivnostima). Ovo se može iskoristiti za odgovaranje na dodatna pitanja i može pokrenuti dodatne akcije. Na kraju, modeli konstruisani u Fazi 3 mogu se koristiti za operativnu podršku (Faza 4). Znanje izvučeno iz istorijskih podataka o događajima kombinuje se sa informacijama o tekućim slučajevima. Ovo se može koristiti za intervencije, predviđanja i preporuke. Faze 3 i 4 mogu se dostići samo ako je proces dovoljno stabilan i strukturisan.

Trenutno postoje tehnike i alati koji mogu podržati sve faze prikazane na Slici 5. Međutim, rudarenje procesa je relativno nova paradigma i većina trenutno dostupnih alata je još uvek prilično nezrela. Pored toga, potencijalni korisnici često nisu svesni mogućnosti i ograničenja rudarenja procesa. Zbog toga ovaj manifest katalogizuje vodeće principe (v. sledeći odeljak) i izazove (v. stranu 10) kako za korisnike tehnika rudarenja procesa, tako i za istraživače i programere koji su zainteresovani za unapređenje stanja u ovoj oblasti.



Slika 5: L^* model životnog ciklusa koji opisuje projekat rudarenja procesa koji se sastoji od pet faza: planiranje i opravdanje (Faza 0), izdvajanje (Faza 1), kreiranje modela toka kontrole i njegovo povezivanje sa dnevnikom događaja (Faza 2), kreiranje integrisanog modela procesa (Faza 3) i pružanje operativne podrške (Faza 4).

3. Vodeći principi (Guiding Principles)

Kao i kod svake nove tehnologije, postoje očigledne greške koje se mogu napraviti prilikom primene rudarenja procesa u realnim okruženjima. Zbog toga navodimo šest vodećih principa kako bismo sprečili korisnike/analitičare da prave takve greške.

VP1: Podaci o događajima moraju se tretirati kao građani prvog reda (First-Class Citizens)

Polazna tačka za svaku aktivnost rudarenja procesa jesu zabeleženi događaji. Skupove događaja nazivamo dnevnicima događaja, međutim, to ne implicira da događaji moraju biti uskladišteni u namenskim datotekama dnevnika. Događaji mogu biti sačuvani u tabelama baze podataka, evidencijama poruka, arhivama elektronske pošte, transakcionim logovima i drugim izvorima podataka. Od samog formata skladištenja važniji je kvalitet takvih dnevnika događaja. Kvalitet rezultata rudarenja procesa u velikoj meri zavisi od ulaza. Stoga, dnevnik događaja treba tretirati kao građane prvog reda u informacionim sistemima koji podržavaju procese koji se analiziraju.

Nažalost, dnevnik događaja su često

samo „nusproizvod“ koji se koristi za otklanjanje grešaka (*debugging*) ili profajlsanje. Na primer, medicinski uređaji kompanije Philips Healthcare beleže događaje jednostavno zato što su programeri ubacili „print naredbe“ u kod. Iako postoje neke neformalne smernice za dodavanje takvih naredbi u kod, potreban je sistematičniji pristup kako bi se poboljšao kvalitet dnevnika događaja. Podaci o događajima moraju se posmatrati kao građani prvog reda (a ne kao građani drugog reda).

Postoji nekoliko kriterijuma za procenu kvaliteta podataka o događajima. Događaji moraju biti pouzdani, tj. mora se sa sigurnošću moći pretpostaviti da

Nivo	Karakterizacija	Primeri
★★★★★	Najviši nivo: dnevnik događaja je odličnog kvaliteta (tj. pouzdan i potpun) i događaji su dobro definisani. Događaji se beleže na automatski, sistematičan, pouzdan i bezbedan način. Aspekti privatnosti i bezbednosti su adekvatno rešeni. Štaviše, zabeleženi događaji (i svi njihovi atributi) imaju jasnu semantiku. Ovo podrazumeva postojanje jedne ili više ontologija. Događaji i njihovi atributi ukazuju na ovu ontologiju.	Semantički anotirani dnevnic BPM sistema.
★★★★	Događaji se beleže automatski, na sistematičan i pouzdan način, tj. dnevnic su pouzdani i potpuni. Za razliku od sistema na nivou ★★★, pojmovi kao što su instanca procesa (case) i aktivnost podržani su na eksplicitan način.	Dnevnic događaja tradicionalnih BPM / workflow sistema.
★★★	Događaji se beleže automatski, ali se ne prati sistematičan pristup pri beleženju. Međutim, za razliku od dnevnika na nivou ★★, postoji određeni nivo garancije da zabeleženi događaji odgovaraju stvarnosti (tj. dnevnik je pouzdan, ali ne nužno i potpun). Uzmimo, na primer, događaje zabeležene ERP sistemom. Iako događaje treba izvući iz različitih tabela, može se pretpostaviti da su informacije tačne (npr. bezbedno je pretpostaviti da uplata zabeležena u ERP-u stvarno postoji i obrnuto).	Tabele u ERP sistemima, dnevnic CRM sistema, transakcioni dnevni- ci sistema za razmenu poruka, dnevnic događaja visokoteh- noloških sistema itd.
★★	Događaji se beleže automatski, tj. kao nusproizvod nekog informa- cionog sistema. Pokrivenost varira, tj. ne prati se sistematičan pristup u odlučivanju koji se događaji beleže. Štaviše, moguće je zaobići infor- macioni sistem. Otuda događaji mogu nedostajati ili biti nepravilno zabeleženi.	Dnevnic sistema za upravljanje do- kumentima i proizvodima, dnevnic grešaka ugrađenih (<i>embedded</i>) sistema, radni listovi serviseri itd.
★	Najniži nivo: dnevnic događaja su lošeg kvaliteta. Zabeleženi događaji možda ne odgovaraju stvarnosti, a događaji mogu nedostajati. Dnevnic u kojima se događaji beleže ručno obično imaju ovakve karakteristike.	Tragovi ostavljeni na papirnim doku- mentima koji se kreću kroz organi- zaciju („žuti samolepljivi papirići“), papirni medicinski kartoni itd.

Tabela 1: Nivoi zrelosti dnevnika događaja

su se zabeleženi događaji zaista dogodili i da su atributi događaja tačni. Dnevnic događaja moraju biti potpuni, tj. unutar određenog obuhvata nijedan događaj ne sme nedostajati. Svaki zabeleženi događaj mora imati dobro definisanu semantiku. Pored toga, podaci o događajima moraju biti bezbedni u smislu da su pitanja privatnosti i sigurnosti rešena prilikom njihovog beleženja. Na primer, akteri bi trebalo da budu svesni vrste događaja koji se beleže i načina na koji se oni koriste.

Tabela 1 definiše pet nivoa zrelosti dnevnika događaja u rasponu od odličnog kvaliteta (★★★★★) do lošeg kvaliteta (★). Na primer, dnevnic događaja kompanije Philips Healthcare nalaze se na nivou ★★★, tj. događaji se beleže automatski i zabeleženo ponašanje odgovara stvarnosti, ali se ne koristi sistematičan pristup za

dodeljivanje semantike događajima i za osiguranje pokrivenosti na određenom nivou. Tehnike rudarenja procesa mogu se primeniti na dnevnicke nivoe ★★★★★, ★★★★★ i ★★★. U principu, rudarenje procesa je moguće primeniti i na dnevnicke nivoe ★★ ili ★. Međutim, analiza takvih dnevnika je tipično problematična i rezultati nisu pouzdani. U stvari, nema mnogo smisla primenjivati rudarenje procesa na dnevnicke nivoe ★.

Da bi imale koristi od rudarenja procesa, organizacije treba da teže dnevnicima događaja najvišeg mogućeg nivoa kvaliteta.

VP2: Izdvajanje dnevnika mora biti vođeno pitanjima (*Log Extraction Should Be Driven by Questions*)

Kao što je prikazano na Slici 5, aktivnosti rudarenja procesa moraju biti vođene pitanjima. Bez konkretnih pitanja veoma je teško izvući smislene podatke o događajima. Uzmimo, na primer, hiljade tabela u bazi podataka ERP sistema kao što je SAP. Bez konkretnih pitanja nemoguće je odabrati tabele relevantne za ekstrakciju podataka.

Model procesa kao što je onaj prikazan na Slici 1 opisuje životni ciklus slučajeva (tj. instanci procesa) određenog tipa. Prema tome, pre primene bilo koje tehnike rudarenja procesa potrebno je odabrati tip slučajeva koji će se analizirati. Ovaj izbor treba da bude vođen pitanjima na koja je potrebno odgovoriti, a to može biti netrivialno. Razmotrimo, na primer, obradu narudžbina kupaca. Svaka narudžbina može se sastojati od više stavki, jer kupac može naručiti više proizvoda u jednoj

porudžbini. Jedna narudžbina kupca može rezultirati višestrukim isporukama. Jedna isporuka može se odnositi na stavke iz više različitih narudžbina. Dakle, postoji relacija više-prema-više ($M:N$) između narudžbina i isporuka, kao i relacija jedan-prema-više ($1:N$) između narudžbina i stavki narudžbine. Na osnovu baze podataka sa podacima o događajima koji se odnose na porudžbine, stavke i isporuke, mogu se otkriti različiti modeli procesa. Podaci se mogu izdvojiti sa ciljem da se opiše životni ciklus pojedinačnih narudžbina. Međutim, podaci se takođe mogu izdvojiti i sa ciljem otkrivanja životnog ciklusa pojedinačnih stavki narudžbine ili životnog ciklusa pojedinačnih isporuka.

VP3: Konkurentnost, izbor i druge osnovne konstrukcije toka kontrole moraju biti podržani

Postoji mnoštvo jezika za modelovanje procesa (npr. BPMN, EPC, Petrijeve mreže, BPEL i UML dijagrami aktivnosti). Neki od ovih jezika pružaju mnogo elemenata za modelovanje (npr. BPMN nudi više od 50 različitih grafičkih elemenata), dok su drugi veoma bazični (npr. Petrijeve mreže su sastavljene od samo tri različita elementa: mesta, prelaza i usmerenih lukova). Opis toka kontrole je okosnica svakog modela procesa. Osnovne radne konstrukcije (poznate i kao obrasce / *patterns*) koje podržavaju svi vodeći jezici jesu sekvenca, paralelno usmeravanje (AND-razdvajanja/spajanja), izbor (XOR-razdvajanja/spajanja) i petlje. Očigledno, ove obrasce moraju podržavati tehnike rudarenja procesa. Međutim, neke tehnike nisu u stanju da se bave konkurentnošću i podržavaju samo Markove lance / sisteme tranzicije.

Slika 6 prikazuje efekat korišćenja tehnika rudarenja procesa koje nisu u stanju da otkriju konkurentnost (nemaju AND-razdvajanja/spajanja). Razmotrimo dnevnik događaja:

$$L = \{ \langle A, B, C, D, E \rangle, \langle A, B, D, C, E \rangle, \langle A, C, B, D, E \rangle, \langle A, C, D, B, E \rangle, \langle A, D, B, C, E \rangle, \langle A, D, C, B, E \rangle \}$$

Dnevnik L sadrži slučajeve koji počinju sa A i završavaju se sa E . Aktivnosti B , C i D javljaju se u bilo kom redosledu između A i E . BPMN model na Slici 6(a) prikazuje kompaktnu reprezentaciju osnovnog procesa koristeći dva AND račvanja (*gateways*). Pretpostavimo

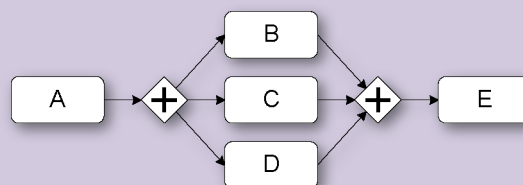
da tehnika rudarenja procesa ne podržava AND račvanja. U tom slučaju, druga dva BPMN modela na Slici 6 su očigledni kandidati. BPMN model na Slici 6(b) jeste kompaktan, ali dozvoljava previše ponašanja (npr. slučajevi kao što je $\langle A, B, B, B, E \rangle$ mogući su prema modelu, ali nisu verovatni prema dnevniku događaja). BPMN model na Slici 6(c) dozvoljava slučajeve iz L , ali eksplicitno kodira sve sekvence, tako da ne predstavlja kompaktnu reprezentaciju dnevnika. Primer pokazuje da za realne modele koji imaju desetine potencijalno konkurentnih aktivnosti, rezultujući modeli drastično pate od nedovoljnog prilagođavanja (*severely underfitting*, tj. dozvoljavaju previše ponašanja) i/ili postaju izuzetno složeni ako konkurentnost nije podržana.

Kao što je ilustrovano Slikom 6, važno je podržati barem osnovne tokove rada. Pored pomenutih osnovnih obrazaca, poželjno

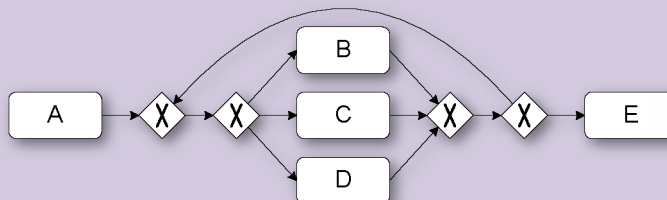
je podržati i OR-razdvajanja/spajanja, jer ona pružaju kompaktnu reprezentaciju inkluzivnih odluka i delimične sinhronizacije.

VP4: Događaji moraju biti povezani sa elementima modela

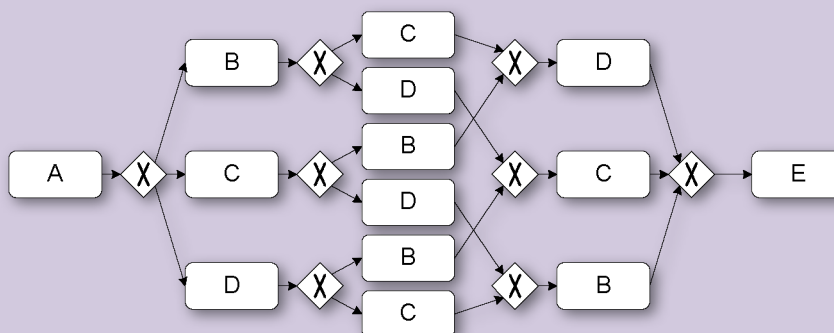
Kao što je naznačeno u Odeljku 2, zabluda je da je rudarenje procesa ograničeno na otkrivanje toka kontrole. Kao što je prikazano na Slici 1, otkriveni model procesa može pokriti različite perspektive (organizacionu, vremensku, perspektivu podataka itd.). Štaviše, otkrivanje je samo jedan od tri tipa rudarenja procesa prikazana na Slici 3. Druga dva tipa (provera usklađenosti i unapređenje) u velikoj meri se oslanjaju na vezu između



(a) B, C, and D can be executed in any order



(b) B, C, and D can be executed in any order but also multiple times



(c) B, C, and D can be executed in any order, but activities need to be duplicated to model all observed sequences.

Slika 6: Primer koji ilustruje probleme kada se konkurentnost (tj. AND-razdvajanja/spajanja) ne može direktno izraziti. U ovom primeru samo su tri aktivnosti (B, C i D) konkurentne. Zamislite rezultujuće modele procesa kada postoji 10 konkurentnih aktivnosti ($2^{10}=1024$ stanja i $10!=3.628.800$ mogućih sekvenci izvršavanja).

elemenata u modelu i događaja u dnevniku. Ova veza se može iskoristiti za ponovno „puštanje“ (*replay*) dnevnika događaja preko modela. Ponovno puštanje može se koristiti za otkrivanje neslaganja između dnevnika događaja i modela, npr. neki događaji u dnevniku nisu mogući prema definisanom modelu. Tehnike provere usklađenosti kvantifikuju i dijagnostikuju takva odstupanja. Vremenske oznake u dnevniku događaja mogu se koristiti za analizu vremenskog ponašanja tokom ponovnog puštanja. Vremenske razlike između uzročno-posledično povezanih aktivnosti mogu se iskoristiti za dodavanje očekivanih vremena čekanja u model. Ovi primeri pokazuju da odnos između događaja u dnevniku i elemenata u modelu služi kao polazna tačka za različite tipove analize.

U nekim slučajevima uspostavljanje ove veze može biti netrivialno. Na primer, događaj se može odnositi na dve različite aktivnosti ili je nejasno na koju se aktivnost odnosi. Takve nejasnoće moraju se otkloniti da bi se rezultati rudarenja procesa pravilno interpretirali. Pored problema povezivanja događaja sa aktivnostima, postoji i problem povezivanja događaja sa instancama procesa. To se uobičajeno naziva korelacija događaja (*event correlation*).

VP5: Modeli se moraju tretirati kao svrsishodne apstrakcije stvarnosti

Model izveden iz podataka o događajima pruža pogled na stvarnost. Takav pogled treba da obezbedi svrsishodnu apstrakciju ponašanja zabeleženog u dnevniku događaja. Za dati dnevnik događaja može postojati više korisnih pogleda. Štaviše, različite zainteresovane strane mogu zahtevati različite poglede. U stvari, modele otkrivene iz dnevnika događaja treba posmatrati kao „mape“ (poput geografskih karata). Ovaj vodeći princip pruža važne uvide, od kojih su dva opisana u nastavku.

Kao prvo, važno je napomenuti da ne postoji takva stvar kao što je „jedina mapa“ za određeno geografsko područje. U zavisnosti od nameravane upotrebe postoje različite karte: auto-karte, planinarske karte, biciklističke karte itd. Sve ove karte prikazuju pogled na istu stvarnost i bilo bi apsurdno pretpostaviti da postoji nešto

poput „savršene karte“. Isto važi i za modele procesa: model treba da naglasi stvari relevantne za određenog korisnika. Otkriveni modeli mogu se fokusirati na različite perspektive (tok kontrole, tok podataka, vreme, resursi, troškovi itd.) i prikazati ih na različitim nivoima granularnosti i preciznosti, npr. menadžer možda želi da vidi grubi, neformalni model procesa fokusiran na troškove, dok analitičar procesa želi detaljan model procesa fokusiran na odstupanja od uobičajenog toka. Takođe treba primetiti da različite zainteresovane strane žele da posmatraju proces na različitim nivoima: strateški nivo (odluke na ovom nivou imaju dugoročne efekte i zasnovane su na agregiranim podacima o događajima tokom dužeg perioda), taktički nivo (odluke na ovom nivou imaju srednjoročne efekte i uglavnom se zasnivaju na nedavnim podacima) i operativni nivo (odluke na ovom nivou imaju neposredne efekte i zasnovane su na podacima o događajima koji se odnose na tekuće slučajeve).

Drugo, korisno je usvojiti ideje iz kartografije kada je u pitanju izrada razumljivih mapa. Na primer, auto-karte apstrahuju manje važne puteve i gradove. Manje važne stvari se ili izostavljaju ili se dinamički grupišu u zbirne oblike (npr. ulice i predgrađa se spajaju u gradove). Kartografi ne samo da eliminišu nebitne detalje, već koriste i boje da istaknu važne karakteristike. Pored toga, grafički elementi imaju određenu veličinu koja ukazuje na njihov značaj (npr. veličine linija i tačaka mogu varirati). Geografske karte takođe imaju jasno tumačenje x-ose i y-ose, tj. raspored elemenata na mapi nije proizvoljan jer koordinate elemenata imaju značenje. Sve je to u oštrom kontrastu sa preovlađujućim modelima procesa koji obično ne koriste boju, veličinu i prostorne karakteristike kako bi modele učinili razumljivijim. Međutim, ideje iz kartografije mogu se lako ugraditi u konstrukciju otkrivenih procesnih mapa. Na primer, veličina aktivnosti može se koristiti za odraz njene učestalosti ili nekog drugog svojstva koje ukazuje na njen značaj (npr. troškovi ili korišćenje resursa). Širina luka može odražavati važnost odgovarajuće uzročne zavisnosti, a bojenje lukova može se koristiti za isticanje uskih grla.

Navedena zapažanja pokazuju da je važno odabrati pravu reprezentaciju i fino

je podesiti za ciljnu publiku. Ovo je važno kako za vizuelizaciju rezultata krajnjim korisnicima, tako i za vođenje algoritama otkrivanja ka odgovarajućim modelima (videti takođe Izazov I5).

VP6: Rudarenje procesa treba da bude kontinuiran proces

Rudarenje procesa može pomoći u pružanju smislenih „mapa“ koje su direktno povezane sa podacima o događajima. Na takve modele mogu se projektovati i istorijski podaci o događajima i trenutni podaci. Štaviše, procesi se menjaju dok se analiziraju. S obzirom na dinamičnu prirodu procesa, nije preporučljivo posmatrati rudarenje procesa kao jednokratnu aktivnost. Cilj ne bi trebalo da bude kreiranje fiksnog modela, već udisanje života u modele procesa tako da se korisnici i analitičari podstaknu da ih posmatraju na dnevnom nivou.

Uporedite ovo sa korišćenjem hibridnih aplikacija (*mashups*) koje koriste geografsko označavanje (*geo-tagging*). Postoje hiljade aplikacija koje koriste Google Maps (npr. aplikacije koje projektuju informacije o stanju u saobraćaju, nekretninama, restoranima brze hrane ili terminima bioskopskih projekcija na izabranu kartu). Ljudi mogu neprimetno da zumiraju i komuniciraju sa takvim mapama (npr. saobraćajne gužve se projektuju na mapu, a korisnik može da izabere određeni problem da vidi detalje). Takođe bi trebalo da bude moguće sprovoditi rudarenje procesa na osnovu podataka o događajima u realnom vremenu. Koristeći „metaforu mape“, možemo zamisliti događaje koji imaju GPS koordinate koje se mogu projektovati na mape u realnom vremenu. Analogno navigacionim sistemima u automobilima, alati za rudarenje procesa mogu pomoći krajnjim korisnicima: (a) navigacijom kroz procese, (b) projektovanjem dinamičkih informacija na procesne mape (npr. prikazivanjem „zastoja“ u poslovnim procesima) i (c) pružanjem predviđanja u vezi sa tekućim slučajevima (npr. procena „vremena dolaska“ slučaja koji kasni). Ovi primeri pokazuju da je šteta ne koristiti modele procesa aktivnije. Zbog toga rudarenje procesa treba posmatrati kao kontinuiran proces koji pruža operativne informacije na različitim vremenskim skalama (minuti, sati, dani, nedelje i meseci).

4. Izazovi (*Challenges*)

Rudarenje procesa je važan alat za savremene organizacije koje moraju da upravljaju netrivialnim operativnim procesima. S jedne strane, postoji neverovatan rast podataka o događajima. S druge strane, procesi i informacije moraju biti savršeno usklađeni kako bi se ispunili zahtevi koji se odnose na usklađenost, efikasnost i korisničku podršku. Uprkos primenljivosti rudarenja procesa, još uvek postoje važni izazovi kojima se treba pozabaviti; oni ilustruju da je rudarenje procesa disciplina u nastajanju. U nastavku navodimo neke od ovih izazova. Ova lista ne pretenduje da bude konačna i tokom vremena se mogu pojaviti novi izazovi ili postojeći mogu nestati usled napretka u oblasti rudarenja procesa.

I1: Pronalaženje, spajanje i čišćenje podataka o događajima

I dalje su potrebni znatni napor da se izdvoje podaci o događajima pogodni za rudarenje procesa. Obično je potrebno prevazići nekoliko prepreka:

- Podaci mogu biti distribuirani na različitim izvorima. Ove informacije je potrebno spojiti. To obično biva problematično kada se u različitim izvorima podataka koriste različiti identifikatori. Na primer, jedan sistem koristi ime i datum rođenja za identifikaciju osobe, dok drugi sistem koristi njen jedinstveni matični broj građana.
- Podaci o događajima su često „orijentisani na objekte“ umesto „orijentisani na procese“. Na primer, pojedinačni proizvodi, palete i kontejneri mogu imati RFID oznake i zabeleženi događaji se odnose na ove oznake. Međutim, da bi se pratila određena porudžbina kupca, takvi događaji orijentisani na objekte moraju se spojiti i prethodno obraditi.
- Podaci o događajima mogu biti nepotpuni. Uobičajeni problem je što događaji ne ukazuju eksplicitno na instance procesa. Često je moguće izvesti ove informacije, ali to može zahtevati znatne napore. Takođe, vremenske informacije mogu nedostajati za neke događaje. Možda će biti potrebno interpolirati vremenske

oznake kako bi se i dalje koristile dostupne vremenske informacije.

- Dnevnik događaja može sadržati ekstremne vrednosti (*outliers*), tj. izuzetno ponašanje koje se takođe naziva šumom (*noise*). Kako definisati ekstremne vrednosti? Kako otkriti takve vrednosti? Na ova pitanja treba odgovoriti da bi se podaci o događajima očistili.
- Dnevnici mogu sadržati događaje na različitim nivoima granularnosti. U dnevniku događaja bolničkog informacionog sistema događaji se mogu odnositi na jednostavne analize krvi ili na složene hirurške zahvate. Takođe, vremenske oznake mogu imati različite nivoe granularnosti u rasponu od milisekundne preciznosti (28-9-2011:h11m28s32ms342) do grubih informacija o datumu (28-9-2011).
- Događaji se odvijaju u određenom kontekstu (vreme, radno opterećenje, dan u nedelji itd.). Ovaj kontekst može objasniti određene fenomene, npr. vreme odziva je duže nego obično zbog nagomilanog rada u toku ili praznika. Za analizu je poželjno uključiti ovaj kontekst. To podrazumeva spajanje podataka o događajima sa kontekstualnim podacima. Ovde stupa na scenu „prokletstvo dimenzionalnosti“ (*curse of dimensionality*) jer analiza postaje neizvodljiva kada se doda previše varijabli.

Potrebni su bolji alati i metodologije za rešavanje navedenih problema. Štaviše, kao što je ranije navedeno, organizacije moraju tretirati dnevnike događaja kao građane prvog reda, a ne kao sporedni proizvod. Cilj je dobiti dnevnike događaja sa ★★★★★ zvezdica (videti Tabelu 1). Ovde su lekcije naučene u kontekstu skladištenja podataka (*data warehousing*) korisne za obezbeđivanje visokokvalitetnih dnevnika događaja. Na primer, jednostavne provere prilikom unosa podataka mogu pomoći da se značajno smanji udeo netačnih podataka o događajima.

I2: Rad sa složenim dnevnicima događaja različitih karakteristika

Dnevnici događaja mogu imati veoma različite karakteristike. Neki dnevnici mogu

Izazovi:

I1: Pronalaženje, spajanje i čišćenje podataka o događajima

I2: Rad sa složenim dnevnicima događaja različitih karakteristika

I3: Kreiranje reprezentativnih benčmarka

I4: Upravljanje pomeranjem koncepta (*Concept Drift*)

I5: Poboljšanje reprezentacione pristrasnosti (*Representational Bias*) korišćene za otkrivanje procesa

I6: Balansiranje između kriterijuma kvaliteta kao što su prilagođenost (*fitness*), jednostavnost (*simplicity*), preciznost (*precision*) i generalizacija (*generalization*)

I7: Međurudarenje između različitih organizacija (*Cross-Organizational Mining*)

I8: Pružanje operativne podrške

I9: Kombinovanje rudarenja procesa sa drugim vrstama analize

I10: Poboljšanje upotrebljivosti za nestručnjake

I11: Poboljšanje razumljivosti za nestručnjake

biti izuzetno veliki, što otežava rukovanje njima, dok su drugi toliko mali da nema dovoljno podataka za donošenje pouzdanih zaključaka. U nekim domenima beleže se zapanjujuće količine događaja. Zbog toga su potrebni dodatni napor za poboljšanje performansi i skalabilnosti. Na primer, ASML kontinuirano prati sve svoje skenere poluprovodničkih pločica (*wafer scanners*). Ove skenere koriste različite organizacije (npr. Samsung i Texas Instruments) za proizvodnju čipova (približno 70% čipova se proizvodi pomoću ASML skenera). Postojeći alati imaju poteškoća u radu sa petabajtima podataka prikupljenih u takvim domenima. Pored broja zabeleženih događaja, postoje i druge karakteristike kao što su prosečan broj događaja po slučaju, sličnost među slučajevima, broj jedinstvenih događaja i broj jedinstvenih putanja. Razmotrimo dnevnik događaja *L1* sa sledećim

karakteristikama: 1000 slučajeva, u proseku 10 događaja po slučaju i male varijacije (npr. nekoliko slučajeva prati iste ili veoma slične putanje). Dnevnik događaja L2 sadrži samo 100 slučajeva, ali u proseku ima 100 događaja po slučaju i svi slučajevi prate jedinstvenu putanju. Jasno je da je L2 mnogo teže analizirati nego L1, iako oba dnevnika imaju sličnu ukupnu veličinu (približno 10.000 događaja).

Pošto dnevnici događaja sadrže samo uzorak ponašanja, ne treba pretpostaviti da su potpuni. Tehnike rudarenja procesa moraju da se nose sa nepotpunošću korišćenjem „pretpostavke otvorenog sveta“ (*open world assumption*): činjenica da se nešto nije dogodilo ne znači da se ne može dogoditi. Ovo čini rad sa malim dnevnicima događaja sa velikom varijabilnošću pravim izazovom.

Kao što je ranije pomenuto, neki dnevnici sadrže događaje na veoma niskom nivou apstrakcije. Ovi dnevnici obično bivaju izuzetno veliki, a pojedinačni događaji niskog nivoa malo interesuju zainteresovane strane. Zbog toga bi bilo poželjno agregirati događaje niskog nivoa u događaje visokog nivoa. Na primer, pri analizi procesa dijagnostike i lečenja određene grupe pacijenata, istraživača možda ne zanimaju pojedinačni testovi zabeleženi u informacionom sistemu laboratorije.

U ovom trenutku, organizacije moraju da koriste pristup pokušaja i grešaka da bi videle da li je dnevnik događaja pogodan za rudarenje procesa. Zbog toga bi alati trebalo da omoguće brzi test izvodljivosti za dati skup podataka. Takav test bi trebalo da ukaže na potencijalne probleme sa performansama i upozori na dnevnike koji su daleko od potpunih ili su previše detaljni.

I3: Kreiranje reprezentativnih benčmarka

Rudarenje procesa je tehnologija u razvoju. To objašnjava zašto dobri benčmark standardi još uvek nedostaju. Na primer, dostupne su desetine tehnika otkrivanja procesa i različiti proizvođači nude različite proizvode, ali ne postoji konsenzus o kvalitetu ovih tehnika. Iako postoje ogromne razlike u funkcionalnosti i performansama, teško je uporediti različite tehnike i alate. Zbog toga je potrebno razviti dobre benčmarke koji se sastoje od primernih skupova podataka i reprezentativnih kriterijuma kvaliteta.

Za klasične tehnike rudarenja podataka dostupni su mnogi dobri benčmarci. Ovi benčmarci su stimulisali proizvođače alata i istraživače da poboljšaju performanse svojih tehnika. U slučaju rudarenja procesa to je veći izazov. Na primer, relacioni model koji je uveo Kod 1969. godine jednostavan je i široko podržan. Kao rezultat toga, potrebno je malo truda za konvertovanje podataka iz jedne baze podataka u drugu i nema problema sa tumačenjem. Za procese takav jednostavan model nedostaje. Standardi predloženi za modelovanje procesa su mnogo komplikovaniji i malo proizvođača podržava potpuno isti skup koncepata. Procesi su jednostavno složeniji od tabelarnih podataka.

Ipak, važno je kreirati reprezentativne benčmarke za rudarenje procesa. Neki početni radovi su već dostupni. Na primer, postoje različite metrike za merenje kvaliteta rezultata rudarenja procesa (prilagođenost, jednostavnost, preciznost i generalizacija). Štaviše, nekoliko dnevnika događaja je javno dostupno (v. www.processmining.org). Pogledajte, na primer, dnevnik događaja korišćen za prvi *Business Process Intelligence Challenge (BPIC'11)* koji je organizovala radna grupa (doi: 10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffcf54).

S jedne strane, trebalo bi da postoje benčmarci zasnovani na skupovima podataka iz stvarnog života. S druge strane, postoji potreba za kreiranjem sintetičkih skupova podataka koji beleže određene karakteristike. Takvi sintetički skupovi podataka pomažu u razvoju tehnika rudarenja procesa koje su prilagođene nepotpunim dnevnicima, bučnim dnevnicima (*noisy event logs*) ili specifičnim populacijama procesa.

Pored kreiranja reprezentativnih benčmarka, potrebno je postići i veći konsenzus o kriterijumima koji se koriste za ocenu kvaliteta rezultata rudarenja procesa (videti takođe Izazov I6). Štaviše, tehnike unakrsne validacije (*cross-validation*) iz rudarenja podataka mogu se prilagoditi za procenu rezultata. Razmotrimo, na primer, *k*-slojnu proveru (*k-fold checking*). Dnevnik događaja se može podeliti na *k* delova. *k*−1 delova se može koristiti za učenje modela procesa, a tehnike provere usklađenosti mogu se koristiti za procenu rezultata u odnosu na preostali deo. Ovo se može ponoviti *k* puta, pružajući time uvide u kvalitet modela.

I4: Upravljanje pomeranjem koncepta (*Concept Drift*)

Termin pomeranje koncepta (*concept drift*) odnosi se na situaciju u kojoj se proces menja dok se analizira. Na primer, na početku dnevnika događaja dve aktivnosti mogu biti konkurentne, dok kasnije u dnevniku ove aktivnosti postaju sekvencijalne. Procesi se mogu menjati usled periodičnih/sezonskih promena (npr. „u decembru je veća potražnja“ ili „petkom popodne ima manje raspoloživih radnika“) ili usled promenjenih tržišnih uslova (npr. „tržište postaje konkurentnije“). Takve promene utiču na procese i od vitalnog je značaja otkriti ih i analizirati. Pomeranje koncepta u procesu može se otkriti deljenjem dnevnika događaja na manje dnevnike i analizom „otisaka stopa“ (*footprints*) manjih dnevnika. Takva analiza „drugog reda“ zahteva znatno više podataka o događajima. Ipak, malo je procesa u stacionarnom stanju i razumevanje pomeranja koncepta je od primarne važnosti za upravljanje procesima. Zbog toga su potrebna dodatna istraživanja i podrška u alatima kako bi se adekvatno analiziralo pomeranje koncepta.

I5: Poboljšanje reprezentacione pristrasnosti (*Representational Bias*) korišćene za otkrivanje procesa

Tehnika otkrivanja procesa proizvodi model koristeći određeni jezik (npr. BPMN ili Petrijeve mreže). Međutim, važno je odvojiti vizuelizaciju rezultata od reprezentacije koja se koristi tokom samog procesa otkrivanja. Izbor ciljnog jezika često obuhvata nekoliko implicitnih pretpostavki. On ograničava prostor pretrage; procesi koji se ne mogu predstaviti ciljnim jezikom ne mogu se ni otkriti. Ova takozvana „reprezentaciona pristrasnost“ koja se koristi tokom procesa otkrivanja treba da bude svestan izbor i ne bi trebalo da bude (samo) vođena željenom grafičkom reprezentacijom.

Razmotrimo, na primer, Sliku 6: da li ciljni jezik dozvoljava konkurentnost ili ne, može uticati i na vizuelizaciju otkrivenog modela i na klasu modela koje algoritam uzima u obzir. Ako reprezentaciona pristrasnost ne dozvoljava konkurentnost (Slika 6(a) nije moguća) i ne dozvoljava

višestrukih aktivnosti koje imaju istu oznaku (Slika 6(c) nije moguća), onda su mogući samo problematični modeli kao što je onaj prikazan na Slici 6(b). Ovaj primer pokazuje da je potreban pažljiviji i prefinjeniji izbor reprezentacione pristranosti.

I6: Balansiranje između kriterijuma kvaliteta kao što su prilagođenost, jednostavnost, preciznost i generalizacija

Dnevnik događaja su često daleko od potpunih, tj. dat je samo primer ponašanja. Modeli procesa obično dozvoljavaju eksponencijalni ili čak beskonačan broj različitih tragova (u slučaju petlji). Štaviše, neki tragovi mogu imati mnogo manju verovatnoću od drugih. Zbog toga je nerealno pretpostaviti da je svaki mogući trag prisutan u dnevniku događaja. Da bismo ilustrovali da je nepraktično uzimati potpune dnevnik zdravo za gotovo, razmotrimo proces koji se sastoji od 10 aktivnosti koje se mogu izvršavati paralelno i odgovarajući dnevnik koji sadrži informacije o 10.000 slučajeva. Ukupan broj mogućih preplitanja u modelu sa 10 konkurentnih aktivnosti je $10! = 3.628.800$. Prema tome, nemoguće je da svako preplitanje bude prisutno u dnevniku jer ima manje slučajeva (10.000) od potencijalnih tragova (3.628.800). Čak i ako u dnevniku postoje milioni slučajeva, izuzetno je malo verovatno da će sve moguće varijacije biti prisutne. Dodatna komplikacija je što su neke alternative ređe od drugih. One se mogu smatrati „šumom“. Nemoguće je izgraditi razuman model za takva ponašanja sa šumom. Otkriveni model mora da apstrahuje od toga; bolje je istražiti ponašanje niske frekvencije korišćenjem provere usklađenosti.

Šum i nepotpunost čine otkrivanje procesa izazovnim problemom. U stvari, postoje četiri konkurentne dimenzije kvaliteta: (a) prilagođenost (*fitness*), (b) jednostavnost (*simplicity*), (c) preciznost (*precision*) i (d) generalizacija (*generalization*). Model sa dobrom prilagođenošću dozvoljava većinu ponašanja viđenih u dnevniku događaja. Model ima savršenu prilagođenost ako svi tragovi u dnevniku mogu biti ponovo reprodukovani (*replayed*) na modelu od početka do kraja. Najjednostavniji model koji može da objasni ponašanje viđeno u dnevniku jeste najbolji model. Ovaj princip je poznat kao Okamova

britva. Prilagođenost i jednostavnost same po sebi nisu dovoljne za procenu kvaliteta otkrivenog modela procesa. Na primer, vrlo je lako konstruisati izuzetno jednostavnu Petrijevu mrežu („model cveta“ / *flower model*) koja je u stanju da reprodukuje sve tragove u dnevniku događaja (ali i bilo koji drugi dnevnik događaja koji se odnosi na isti skup aktivnosti). Slično tome, nepoželjno je imati model koji dozvoljava samo tačno ono ponašanje koje je viđeno u dnevniku događaja. Zapamtite da dnevnik sadrži samo primer ponašanja i da mnogi tragovi koji su mogući možda još nisu viđeni. Model je precizan ako ne dozvoljava „previše“ ponašanja. Jasno je da „modelu cveta“ nedostaje preciznost. Model koji nije precizan pati od nedovoljnog prilagođavanja (*underfitting*). Nedovoljno prilagođavanje je problem kada model preterano generalizuje ponašanje iz uzorka u dnevniku (tj. model dozvoljava ponašanja veoma različita od onoga što je viđeno u dnevniku). Model treba da generalizuje, a ne da ograniči ponašanje samo na primere viđene u dnevniku. Model koji ne generalizuje pati od preteranog prilagođavanja (*overfitting*). Preterano prilagođavanje je problem kada se generiše vrlo specifičan model, dok je očigledno da dnevnik sadrži samo primer ponašanja (tj. model objašnjava konkretan uzorak dnevnika, ali sledeći uzorak dnevnika istog procesa može proizvesti potpuno drugačiji model procesa).

Balansiranje prilagođenosti, jednostavnosti, preciznosti i generalizacije je izazovno. To je razlog zašto većina moćnijih tehnika otkrivanja procesa pruža različite parametre. Potrebno je razviti poboljšane algoritme kako bi se bolje balansirale ove četiri konkurentne dimenzije kvaliteta. Štaviše, svi korišćeni parametri moraju biti razumljivi krajnjim korisnicima.

I7: Međurudarenje između različitih organizacija (*Cross-Organizational Mining*)

Tradicionalno, rudarenje procesa se primenjuje unutar jedne organizacije. Međutim, kako servisne tehnologije, integracija lanca snabdevanja i računarstvo u oblaku (*cloud computing*) postaju sve rasprostranjeniji, postoje scenariji u kojima su dnevnik događaja više organizacija dostupni za analizu. U principu, postoje dva scenarija za međurudarenje između organizacija.

Kao prvo, možemo razmotriti kolaborativno okruženje gde različite organizacije rade zajedno na obradi instanci procesa. O takvom međuorganizacionom procesu može se razmišljati kao o „slagalici“ (*jigsaw puzzle*), tj. ceo proces je isečen na delove i raspoređen po organizacijama koje moraju da sarađuju kako bi uspešno završile slučajeve. Analiza dnevnika događaja unutar samo jedne od ovih organizacija je nedovoljna. Da bi se otkrili procesi od početka do kraja (*end-to-end*), dnevnik događaja različitih organizacija moraju se spojiti. Ovo je netrivialan zadatak jer događaji moraju biti korelisani preko organizacionih granica.

Drugo, možemo razmotriti okruženje u kojem različite organizacije u suštini izvršavaju isti proces deleći iskustva, znanje ili zajedničku infrastrukturu. Razmotrimo, na primer, platformu Salesforce.com. Procese prodaje mnogih organizacija upravlja i podržava Salesforce. S jedne strane, ove organizacije dele infrastrukturu (proces, baze podataka itd.). S druge strane, one nisu prinuđene da prate striktan model procesa jer se sistem može konfigurisati tako da podržava varijante istog procesa. Kao drugi primer, razmotrimo osnovne procese koji se izvršavaju unutar bilo koje opštine (npr. izdavanje građevinskih dozvola). Iako sve opštine u jednoj zemlji moraju da podržavaju isti osnovni skup procesa, mogu postojati i razlike. Očigledno je zanimljivo analizirati takve varijacije među različitim organizacijama. Ove organizacije mogu učiti jedne od drugih, a pružaoci usluga mogu poboljšati svoje usluge i ponuditi usluge sa dodatnom vrednošću zasnovane na rezultatima međuorganizacionog rudarenja procesa.

Nove tehnike analize moraju se razviti za oba tipa međuorganizacionog rudarenja procesa. Ove tehnike takođe treba da uzmu u obzir pitanja privatnosti i bezbednosti. Organizacije možda ne žele da dele informacije iz konkurentskih razloga ili zbog nedostatka poverenja. Zbog toga je važno razviti tehnike rudarenja procesa koje čuvaju privatnost (*privacy-preserving process mining*).

I8: Pružanje operativne podrške

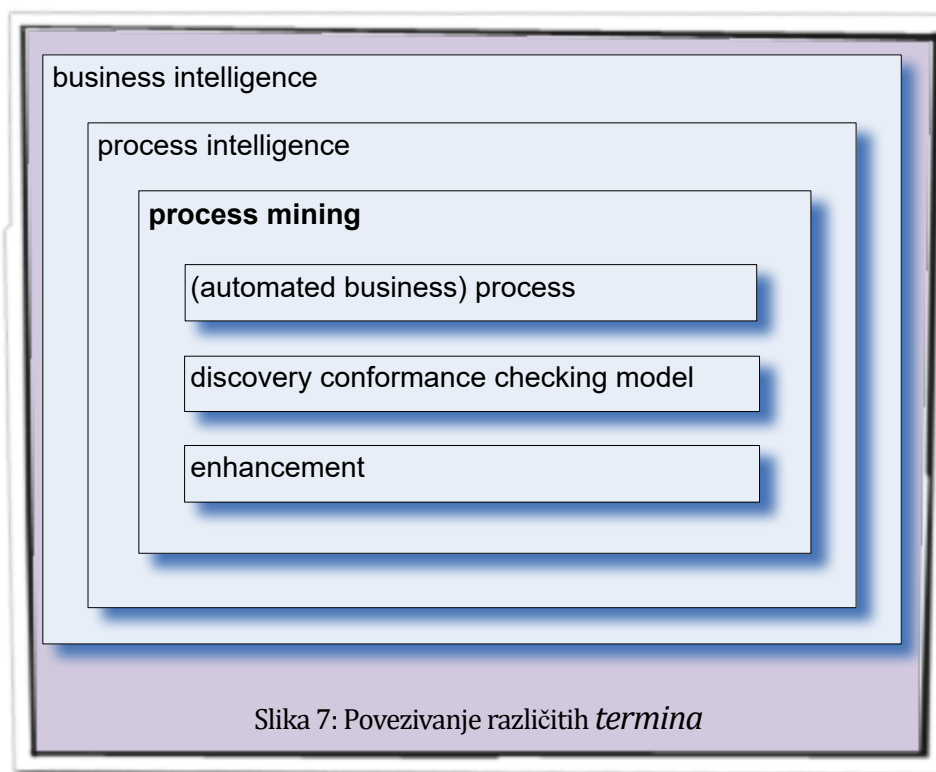
U početku, fokus rudarenja procesa bio je na analizi istorijskih podataka. Danas se, međutim, mnogi izvori podataka ažuriraju u (skoro) realnom vremenu i dostupna

je dovoljna računarska snaga za analizu događaja čim se oni dogode. Stoga, rudarenje procesa ne bi trebalo da bude ograničeno na oflajn analizu i može se koristiti za onlajn operativnu podršku. Mogu se identifikovati tri aktivnosti operativne podrške: otkrivanje odstupanja (detect), predviđanje (predict) i preporučivanje (recommend). Onog trenutka kada slučaj odstupa od unapred definisanog procesa, to se može otkriti i sistem može generisati upozorenje. Često se želi da se takva obaveštenja generišu odmah (kako bi se još uvek moglo uticati na stvari), a ne na oflajn način. Istorijski podaci se mogu koristiti za izgradnju prediktivnih modela. Ovi modeli se mogu koristiti za vođenje tekućih instanci procesa. Na primer, moguće je predvideti preostalo vreme obrade slučaja. Na osnovu takvih predviđanja mogu se izgraditi i sistemi za preporuke koji predlažu određene akcije za smanjenje troškova ili skraćanje vremena protoka. Primena tehnika rudarenja procesa u takvom onlajn okruženju stvara dodatne izazove u pogledu računarske snage i kvaliteta podataka.

I9: Kombinovanje rudarenja procesa sa drugim vrstama analize

Operacioni menadžment, a posebno operaciona istraživanja, grana je nauke o menadžmentu koja se u velikoj meri oslanja na modelovanje. Ovde se koristi širok spektar matematičkih modela, od linearnog programiranja i planiranja projekata do modela redova čekanja, Markovljevih lanaca i simulacije. Rudarenje podataka može se definisati kao „analiza (često velikih) skupova podataka radi pronalaženja neočekivanih veza i sumiranja podataka na nove načine koji su i razumljivi i korisni vlasniku podataka“. Razvijen je širok spektar tehnika: klasifikacija (npr. učenje stabla odlučivanja), regresija, klasterovanje (npr. *k*-means klasterovanje) i otkrivanje obrazaca (npr. učenje asocijativnih pravila).

Oba polja (operacioni menadžment i rudarenje podataka) pružaju vredne tehnike analize. Izazov je kombinovati tehnike iz ovih oblasti sa rudarenjem procesa. Uzmimo za primer simulaciju. Tehnike rudarenja procesa mogu se koristiti za učenje simulacionog modela na osnovu istorijskih podataka. Zatim se simulacioni model može koristiti za pružanje operativne podrške. Zbog bliske veze između dnevnika



Slika 7: Povezivanje različitih termina

događaja i modela, model se može koristiti za ponovno puštanje istorije, a simulacije se mogu pokrenuti iz trenutnog stanja, čime se obezbeđuje taster za „premotavanje unapred“ u budućnost na osnovu podataka uživo.

Slično tome, poželjno je kombinovati rudarenje procesa sa vizuelnom analitikom. Vizuelna analitika kombinuje automatizovanu analizu sa interaktivnim vizuelizacijama radi boljeg razumevanja velikih i složenih skupova podataka. Vizuelna analitika koristi zadivljujuće sposobnosti ljudi da uoče obrasce u nestrukturisanim podacima. Kombinovanjem automatizovanih tehnika rudarenja procesa sa interaktivnom vizuelnom analitikom moguće je izvući više uvida iz podataka o događajima.

I10: Poboljšanje upotrebljivosti za nestručnjake

Jedan od ciljeva rudarenja procesa jeste kreiranje „živih modela procesa“, tj. modela koji se koriste svakodnevno, a ne statičnih modela koji završavaju u nekoj arhivi. Novi podaci o događajima mogu se koristiti za otkrivanje novonastalog ponašanja. Veza između podataka o događajima i modela procesa omogućava projekciju trenutnog stanja i nedavnih aktivnosti na ažurne modele. Otuda krajnji korisnici mogu svakodnevno komunicirati sa rezultatima

rudarenja procesa. Takve interakcije su veoma vredne, ali zahtevaju i intuitivne korisničke interfejsse. Izazov je sakriti sofisticirane algoritme rudarenja procesa iza korisničkih interfejsa jednostavnih za upotrebu, koji automatski podešavaju parametre i predlažu odgovarajuće tipove analize.

I11: Poboljšanje razumljivosti za nestručnjake

Čak i ako je lako generisati rezultate rudarenja procesa, to ne znači da su ti rezultati zaista korisni. Korisnik može imati problema sa razumevanjem izlaza ili može biti u iskušenju da donese pogrešne zaključke. Da bi se izbegli takvi problemi, rezultate treba predstaviti korišćenjem odgovarajuće reprezentacije (videti takođe VP5). Štaviše, pouzdanost rezultata uvek mora biti jasno naznačena. Možda ima premalo podataka da bi se opravdali određeni zaključci. U stvari, postojeće tehnike otkrivanja procesa obično ne upozoravaju na nisku prilagođenost ili na preterano prilagođavanje. One uvek prikazuju model, čak i kada je jasno da ima premalo podataka da bi se doneli bilo kakvi zaključci.

Epilog

IEEE Radna grupa za rudarenje procesa ima za cilj da (a) promoviše primenu rudarenja procesa, (b) vodi programere softvera,

konsultante, poslovne menadžere i krajnje korisnike pri korišćenju najsavremenijih tehnika i (c) stimuliše istraživanja o rudarenju procesa. Ovaj manifest navodi glavne principe i namere radne grupe. Nakon uvođenja teme rudarenja procesa, manifest katalogizuje neke vodeće principe (Odeljak 3) i izazove (Odeljak 4). Vodeći principi se mogu koristiti kako bi se izbegle očigledne greške. Lista izazova je namenjena usmeravanju istraživačkih i razvojnih napora. I jedni i drugi imaju za cilj povećanje nivoa zrelosti rudarenja procesa.

Za kraj, nekoliko reči o terminologiji. U prostoru rudarenja procesa koriste se sledeći termini: rudarenje toka posla (*workflow mining*), rudarenje (poslovnih) procesa (*[business] process mining*), automatizovano otkrivanje (poslovnih) procesa (*automated [business] process discovery*) i inteligencija (poslovnih) procesa (*[business] process intelligence*). Čini se da različite organizacije koriste različite termine za koncepte koji se preklapaju. Na primer, Gartner promovise termin „Automated Business Process Discovery“ (ABPD), a Software AG koristi „Process Intelligence“ za svoju kontrolnu platformu. Termin „rudarenje toka posla“ čini se manje prikladnim jer je kreiranje modela toka posla samo jedna od mnogih mogućih primena rudarenja procesa. Slično tome, dodavanje termina „poslovni“ sužava obuhvat na određene primene rudarenja procesa. Postoje brojne primene rudarenja procesa (npr. analiza korišćenja visokotehnoloških sistema ili analiza veb-sajtova) gde se ovaj dodatak čini neprikladnim. Iako je otkrivanje procesa važan deo spektra rudarenja procesa, to je samo jedan od mnogih slučajeva upotrebe. Provera usklađenosti, predviđanje, organizaciono rudarenje, analiza društvenih mreža itd. jesu drugi slučajevi upotrebe koji se protežu izvan otkrivanja procesa.

Slika 7 povezuje neke od upravo pomenutih termina. Sve tehnologije i metode koje imaju za cilj pružanje operativnih informacija koje se mogu koristiti za podršku donošenju odluka mogu se pozicionirati pod okriljem poslovne inteligencije (Business Intelligence – BI). Inteligencija (poslovnih) procesa može se posmatrati kao kombinacija BI i BPM-a, tj. BI tehnike se koriste za analizu i poboljšanje procesa i njihovog upravljanja. Rudarenje procesa se može posmatrati kao konkretizacija inteligencije procesa uzimajući dnevnik događaja kao polaznu

tačku. (Automatizovano) otkrivanje poslovnih procesa je samo jedan od tri osnovna tipa rudarenja procesa. Slika 7 može biti donekle obmanjujuća u smislu da većina BI alata ne pruža funkcionalnosti rudarenja procesa kako je opisano u ovom dokumentu. Termin BI se često praktično sužava ka određenom alatu ili metodi koji pokrivaju samo mali deo šireg BI spektra.

Mogu postojati komercijalni razlozi za korišćenje alternativnih termina. Neki proizvođači takođe žele da naglase određeni aspekt (npr. otkrivanje ili inteligenciju). Međutim, da bi se izbegla konfuzija, bolje je koristiti termin „rudarenje procesa“ (process mining) za disciplinu pokrivenu ovim manifestom.

Rečnik pojmova

Aktivnost (Activity): dobro definisan korak u procesu. Događaji se mogu odnositi na početak, završetak, otkazivanje itd. aktivnosti za određenu instancu procesa.

Automatizovano otkrivanje poslovnih procesa (Automated Business Process Discovery): videti *Otkrivanje procesa*.

Poslovna inteligencija (Business Intelligence – BI): širok skup alata i metoda koji koriste podatke za podršku donošenju odluka.

Inteligencija poslovnih procesa (Business Process Intelligence): videti *Inteligencija procesa*.

Upravljanje poslovnim procesima (Business Process Management – BPM): disciplina koja kombinuje znanje iz informacionih tehnologija i znanje iz menadžerskih nauka i primenjuje oba na operativne poslovne procese.

Slučaj / Instanca procesa (Case): videti *Instanca procesa*.

Pomeranje koncepta (Concept Drift): fenomen da se procesi često menjaju tokom vremena. Posmatrani proces se može postepeno (ili iznenada) promeniti usled sezonskih promena ili povećane konkurencije, što otežava analizu.

Provera usklađenosti (Conformance Checking): analiza da li stvarnost, zabeležena u dnevniku događaja, odgovara modelu i obrnuto. Cilj je otkriti odstupanja i izmeriti njihovu ozbiljnost. Provera usklađenosti je jedan od tri osnovna tipa rudarenja procesa.

Medurudarenje između organizacija (Cross-Organizational Process Mining): primena tehnika rudarenja procesa na dnevnik događaja koji potiču iz različitih

Ovaj manifest je izvorno objavljen u: „Business Process Management Workshops 2011, Lecture Notes in Business Information Processing, Vol. 99, Springer-Verlag, 2011“, i preveden je na različite jezike. Posetite zvaničnu stranicu IEEE Radne grupe za rudarenje procesa: <http://www.win.tue.nl/ieeetfpm/> radi dodatnih informacija.

organizacija.

Rudarenje podataka (Data Mining): analiza (često velikih) skupova podataka radi pronalazjenja neočekivanih veza i sažimanja podataka na nove načine koji su razumljivi i korisni vlasniku podataka.

Događaj (Event): akcija zabeležena u dnevniku, npr. početak, završetak ili otkazivanje aktivnosti za određenu instancu procesa.

Dnevnik događaja (Event Log): skup događaja koji se koristi kao ulaz za rudarenje procesa. Događaji ne moraju biti smešteni u posebnu datoteku dnevnika (npr. događaji mogu biti raspoređeni po različitim tabelama baze podataka).

Prilagođenost (Fitness): mera koja određuje koliko dobro dati model dopušta ponašanje zabeleženo u dnevniku događaja. Model ima savršenu prilagođenost ako svi tragovi u dnevniku mogu biti u potpunosti reprodukovani na modelu od početka do kraja.

Generalizacija (Generalization): mera koja određuje koliko je model u stanju da obuhvati neviđeno ponašanje. Model koji pati od „preteranog prilagođavanja“ (*overfitting*) nije u stanju da dovoljno generalizuje.

Unapređenje modela (Model Enhancement): jedan od tri osnovna tipa rudarenja procesa. Model procesa se proširuje ili poboljšava korišćenjem informacija izvučenih iz nekog dnevnika događaja. Na primer, uska grla se mogu identifikovati ponovnim puštanjem dnevnika događaja preko modela procesa uz analizu vremenskih oznaka.

MXML: format zasnovan na XML-u za razmenu dnevnika događaja. XES zamenjuje MXML kao novi format za rudarenje procesa nezavisan od konkretnih alata.

Operativna podrška (Operational Support): onlajn analiza podataka o događajima sa ciljem praćenja i uticaja

na tekuće instance procesa. Mogu se identifikovati tri aktivnosti operativne podrške: *otkrivanje odstupanja* (generisanje upozorenja ako uočeno ponašanje odstupa od modelovanog), *predviđanje* (predviđanje budućeg ponašanja na osnovu prošlosti, npr. predviđanje preostalog vremena obrade) i *preporučivanje* (predlaganje odgovarajućih akcija za postizanje određenog cilja, npr. minimiziranje troškova).

Preciznost (*Precision*): mera koja određuje da li model zabranjuje ponašanje koje je znatno drugačije od ponašanja viđenog u dnevniku događaja. Model sa niskom preciznošću pati od „nedovoljnog prilagođavanja“ (*underfitting*).

Otkrivanje procesa (*Process Discovery*): jedan od tri osnovna tipa rudarenja procesa. Na osnovu dnevnika događaja uči se model procesa. Na primer, α -algoritam je sposoban da otkrije Petrijevu mrežu identifikovanjem procesnih obrazaca u skupovima događaja.

Instanca procesa / Slučaj (*Process Instance / Case*): entitet koji se obrađuje u procesu koji se analizira. Događaji se odnose na instance procesa. Primeri instanci procesa su porudžbine kupaca, odštetni zahtevi u osiguranju, zahtevi za kredit itd.

Inteligencija procesa (*Process Intelligence*): grana poslovne inteligencije (BI) fokusirana na upravljanje poslovnim procesima (BPM).

Rudarenje procesa (*Process Mining*): tehnike, alati i metode za otkrivanje, praćenje i poboljšanje stvarnih procesa (tj. procesa koji nisu pretpostavljeni)

izdvajanjem znanja iz dnevnika događaja koji su uobičajeno dostupni u današnjim (informacionim) sistemima.

Reprezentaciona pristrasnost

(*Representational Bias*): izabrani ciljni jezik za predstavljanje i konstrukciju rezultata rudarenja procesa.

Jednostavnost (*Simplicity*): mera koja operacionalizuje princip Okamove britve, tj. najjednostavniji model koji može da objasni ponašanje viđeno u dnevniku predstavlja najbolji model. Jednostavnost se može kvantifikovati na različite načine, npr. brojem čvorova i lukova u modelu.

XES: standard zasnovan na XML-u za dnevnike događaja. Ovaj standard je IEEE Radna grupa za rudarenje procesa usvojila kao podrazumevani format za razmenu dnevnika događaja (v. www.xes-standard.org).

Authors

Wil van der Aalst
Arya Adriansyah
Ana Karla Alves de
Medeiros
Franco Arcieri
Thomas Baier
Tobias Blickle
Jagadeesh Chandra Bose
Peter van den Brand
Ronald Brandtjen
Joos Buijs
Andrea Burattin
Josep Carmona
Malu Castellanos
Jan Claes
Jonathan Cook
Nicola Costantini
Francisco Curbera
Ernesto Damiani
Massimiliano de Leoni

Pavlos Delias
Boudewijn van Dongen
Marlon Dumas
Schahram Dustdar
Dirk Fahland
Diogo R. Ferreira
Walid Gaaloul
Frank van Geffen
Sukriti Goel
Christian Günther
Antonella Guzzo
Paul Harmon
Arthur ter Hofstede
John Hoogland
Jon Espen Ingvaldsen
Koki Kato
Rudolf Kuhn
Akhil Kumar
Marcello La Rosa
Fabrizio Maggi

Donato Malerba
Ronny Mans
Alberto Manuel
Martin McCreesh
Paola Mello
Jan Mendling
Marco Montali
Hamid Motahari
Nezhad
Michael zur Muehlen
Jorge Munoz-Gama
Luigi Pontieri
Joel Ribeiro
Anne Rozinat
Hugo Seguel Pérez
Ricardo Seguel Pérez
Marcos Sepúlveda
Jim Sinur
Pnina Soffer
Minseok Song

Alessandro Sperduti
Giovanni Stilo
Casper Stoel
Keith Swenson
Maurizio Talamo
Wei Tan
Chris Turner
Jan Vanthienen
George Varvaressos
Eric Verbeek
Marc Verdonk
Roberto Vigo
Jianmin Wang
Barbara Weber
Matthias Weidlich
Ton Weijters
Lijie Wen
Michael Westergaard
Moe Wynn