بیانیه فرایند کاوی

Wil van der Aalst, Arya Adriansyah, Ana Karla Alves de Medeiros,Franco Arcieri, ThomasBaier, 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 F. van Dongen, Marlon Dumas, Schahram Dustdar, Dirk Fahland, Diogo R. Ferreira, Walid Gaaloul, Frank van Geffen, Sukriti Goel, ChristianG¨unther, 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 S. Mans, Alberto Manuel, Martin McCreesh, Paola Mello, Jan Mendling, Marco Montali, Hamid R. Motahari-Nezhad, Michael zur Muehlen, Jorge Munoz-Gama, LuigiPontieri, JoelRibeiro, Anne Rozinat, Hugo Seguel P´erez, Ricardo Seguel P´erez, Marcos Sep´ulveda, Jim Sinur, Pnina Soffer, Minseok Song, Alessandro Sperduti, Giovanni Stilo, CasperStoel, Keith Swenson, Maurizio Talamo, WeiTan, ChrisTurner, Jan Vanthienen, George Varvaressos, Eric Verbeek, Marc Verdonk, Roberto Vigo, Jianmin Wang, BarbaraWeber, Matthias Weidlich, TonWeijters, Lijie Wen, Michael Westergaard, and MoeWynn

IEEE Task Force on Process Mining[1]

http://www.win.tue.nl/ieeetfpm

(نسخه ۱)

ترجمه: آصف پورمعصومی

asef.pourmasoumi@stu.um.ac.ir

چکیده: تکنیک‌های فرایندکاوی قادر هستند تا دانش را با استفاده از نگاره‌های رویداد که بطور معمول در سیستم‌های اطلاعاتی امروزه موجود می‌باشند، استخراج نمایند. این تکنیک‌ها ابزارهای جدیدی را برای کشف، نظارت و بهبود فرایندها در حوزه‌های کاربردی گوناگون، فراهم می‌کنند. دو دلیل عمده برای علاقه فزاینده و روز افزون در حوزه فرایند کاوی وجود دارد. از یک طرف، هر روز تعداد بیشتری رویداد در پایگاه داده‌ها ثبت می‌شود که بنابراین جزییات تاریخچه‌ای از فرایندهای کسب و کار را فراهم می‌آورد. از طرف دیگر با توجه به فضای رقابتی کسب و کار و محیط در حال تغییر، نیاز به بهبود و پشتیبانی فرایندهای کسب و کار می‌باشد. این بیانیه[2] توسط IEEE Task Force on Process Mining و با هدف ترویج و توسعه فرایند کاوی ایجاد شده است. بعلاوه با تعریف مجموعه‌ای از اصول راهنما و لیست چالش‌های مهم، امید است که این بیانیه یک راهنما برای توسعه دهندگان نرم‌افزار، دانشمندان، مشاوران، مدیران کسب و کار و کاربران نهایی فراهم آورد. هدف این است که با افزایش بلوغ فرایند کاوی به عنوان یک ابزار جدید، طراحی (مجدد)، کنترل و پشتیبانی از فرآیندهای کسب و کار عملیاتی ارتقا داده شود.

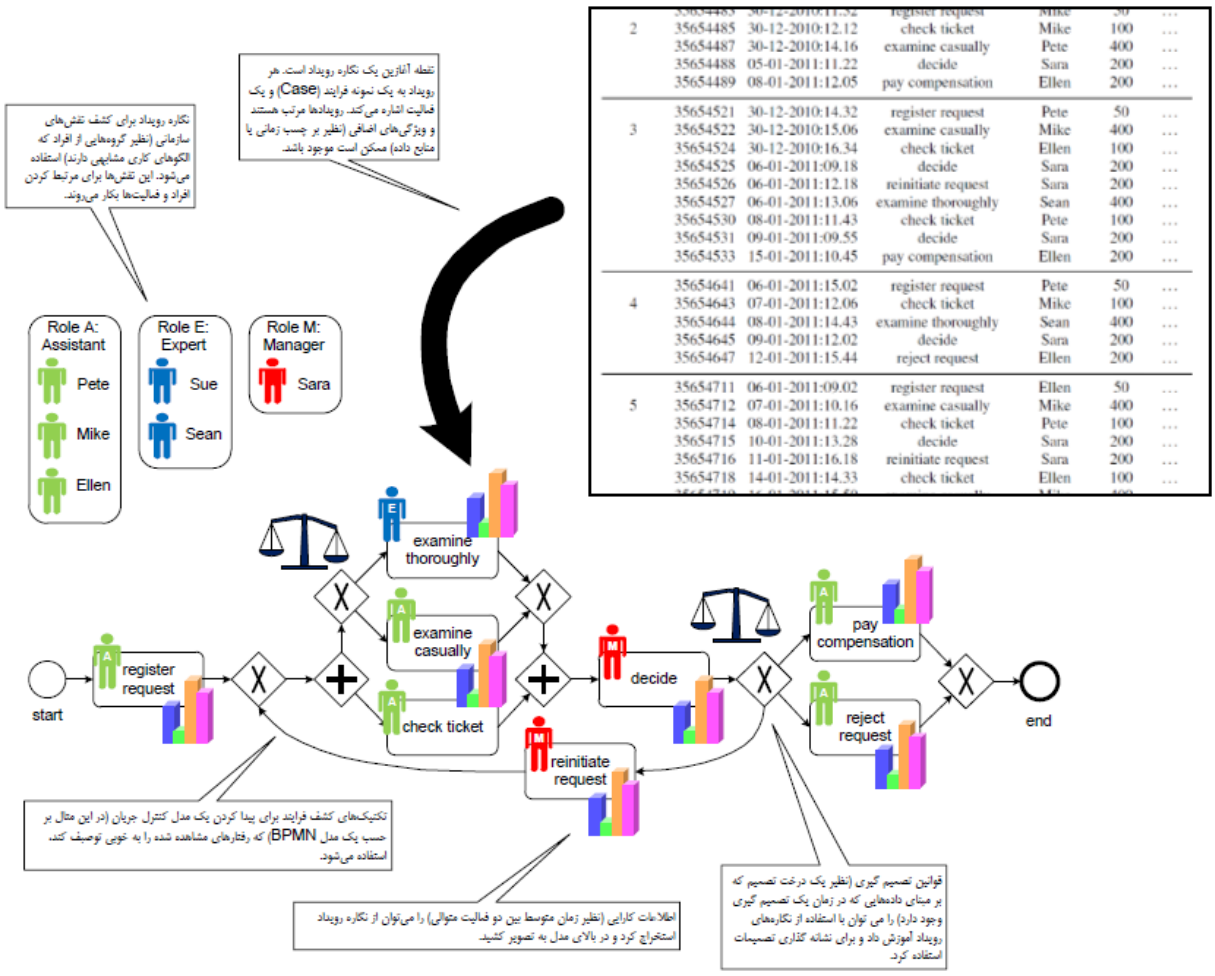
---

[1] - نسخه اصلی در کارگاه BPM2011، Part I, LNBIP 99, pp. 169–194, 2012 چاپ شده است.

[2] - manifesto

۱- برنامه ویژه IEEE برای فرایند کاوی

یک مانیفست یک "توصیف عمومی از اصول و اهداف" است که توسط یک گروه از افراد تولید می‌شود. این بیانیه توسط اعضا و حامیان برنامه ویژه IEEE برای فرایند کاوی نوشته شده است. هدف از این برنامه ویژه، گسترش تحقیق، توسعه، آموزش، پیاده‌سازی، تکامل و فهم فرایند کاوی می‌باشد.



2	35654485	30-12-2010:12.12	check ticket	Mike	100
	35654487	30-12-2010:14.16	examine casually	Pete	400
	35654488	05-01-2011:11.22	decide	Sara	200
	35654489	08-01-2011:12.05	pay compensation	Ellen	200
3	35654521	30-12-2010:14.32	register request	Pete	50
	35654522	30-12-2010:15.06	examine casually	Mike	400
	35654524	30-12-2010:16.34	check ticket	Ellen	100
	35654525	06-01-2011:09.18	decide	Sara	200
	35654526	06-01-2011:12.18	reinitiate request	Sara	200
	35654527	06-01-2011:13.06	examine thoroughly	Sean	400
	35654530	08-01-2011:11.43	check ticket	Pete	100
	35654531	09-01-2011:09.55	decide	Sara	200
	35654533	15-01-2011:10.45	pay compensation	Ellen	200
4	35654641	06-01-2011:15.02	register request	Pete	50
	35654643	07-01-2011:12.06	check ticket	Mike	100
	35654644	08-01-2011:14.43	examine thoroughly	Sean	400
	35654645	09-01-2011:12.02	decide	Sara	200
	35654647	12-01-2011:15.44	reject request	Ellen	200
5	35654711	06-01-2011:09.02	register request	Ellen	50
	35654712	07-01-2011:10.16	examine casually	Mike	400
	35654714	08-01-2011:11.22	check ticket	Pete	100
	35654715	10-01-2011:13.28	decide	Sara	200
	35654716	11-01-2011:16.18	reinitiate request	Sara	200
	35654718	14-01-2011:14.33	check ticket	Ellen	100

شکل ۱- تکنیک‌های فرایند کاوی دانش را از نگاره‌های رویداد با هدف کشف، نظارت و بهبود فرآیندها، استخراج می‌نمایند[۱].

فرایند کاوی، موضوع تحقیقاتی نسبتا جدیدی می‌باشد که بین هوش محاسباتی و داده کاوی از یک سمت و مدل کردن فرایند از سمت دیگر قرار می‌گیرد. ایده فرایند کاوی کشف، نظارت و بهبود فرآیندهای واقعی (و نه فرآیندهای فرضی) با استخراج دانش از نگاره‌های رویداد[1] موجود در سیستم‌های اطلاعاتی امروزی می‌باشد (*شکل ۱* را مشاهده نمایید). فرایند کاوی شامل کشف (به صورت خودکار) فرایند (به عنوان مثال استخراج مدل‌های فرایندی از یک نگاره رویداد)، چک کردن انطباق (به عنوان مثال ردیابی انحرافات از طریق مقایسه مدل و نگاره)، کاوش سازمانی/شبکه اجتماعی، ساخت خودکار مدل‌های شبیه‌سازی، توسعه مدل، اصلاح مدل، پیش بینی وقایع و توصیه‌های مبتنی بر تاریخچه است.

[1] - event log

فرایند کاوی یک پل ارتباطی مهم بین داده کاوی و آنالیز و مدل‌کردن فرآیندهای کسب و کار می‌باشد. در زیر چتر هوش‌سازمانی تعداد زیادی اصطلاح[1] معرفی شده است که برای ارجاع به ابزارهای گزارشگیری و داشبوردهای مدیریتی بکار می‌رود. نظارت فعالیت‌های کسب و کار (BAM[2]) به تکنولوژی‌هایی ارجاع می‌کند که نظارت بلادرنگ بر فعالیت‌های کسب و کار را فراهم می‌سازند. پردازش رویدادهای پیچیده (CEP[3])، به تکنولوژی‌هایی که به پردازش حجم زیادی از رویدادها می‌پردازد، اشاره دارد. از این پردازش‌ها برای نظارت، بهینه‌سازی و هدایت فرآیندهای کسب و کار بصورت بلادرنگ استفاده می‌شود. مدیریت بهره‌وری شرکتها(CPM[4])،اصطلاح دیگری است که برای اندازه‌گیری کارایی یک سازمان یا یک شرکت استفاده می شود. همچنین روش‌های مدیریتی دیگری نظیر بهبود مستمر فرآیند (CPI[5])، بهبود فرآیندهای کسب و کار (BPI[6])، مدیریت کیفیت کلی (TQM[7]) و روش six sigma نیز مرتبط با این بحث می‌باشند. وجه مشترک این روش‌ها این است که در تمامی آنها، فرآیند "زیر میکروسکوب قرار داده می‌شود" تا در صورت امکان بهبودهایی شناسایی گردد. فرایند کاوی تکنولوژی جدیدی است که اهداف CPM، BPI، TQM، Six Sigma و روش‌های مشابه را امکان پذیر می‌کند.

در حالیکه ابزارهای BI و روش‌های مدیریت نظیر six sigma و TQM به منظور افزایش عملکرد عملیاتی نظیر کاهش زمان گردش و کاهش نقص‌ها، بکار می‌روند، سازمان‌ها تاکید ویژه‌ای برمدیریت شرکت‌ها، ریسک‌ها و تطابق فرآیندهای اجرایی و فرآیندهای مدل شده دارند. قوانینی نظیر مصوبه Sarbanes-Oxley کنگره آمریکا و قانون Basel II Accord، تمرکز بر بررسی انحرافات و تخلفات را نشان می‌دهد. تکنیک‌های فرآیند کاوی ابزاری را به منظور چک کردن دقیق‌تر انطباق و همچنین بررسی اعتبار و قابلیت اطمینان اطلاعات فرآیندهای اصلی یک سازمان، فراهم می‌کنند.

در طول دهه گذشته، با فراهم شدن دسترسی به داده‌های رویداد، روش‌های فرآیند کاوی به بلوغ رسیدند. علاوه بر این، همانطور که پیشتر نیز به آن اشاره شد، روندهای مدیریتی مرتبط با بهبود فرآیندها (نظیر Six sigma، TQM، CPI و CPM) و بررسی انطباق (نظیر SOX و BAM) نیز از مزایا و قابلیت‌های فرایند کاوی بهره‌مند شده‌اند. خوشبختانه الگوریتم‌های فرآیند کاوی در سیستم‌های آکادمیک و تجاری بسیاری پیاده‌سازی شده است. در حال حاضر یک گروه فعال از محققین مشغول تحقیق برروی فرآیند کاوی می‌باشند و این موضوع در حال تبدیل شدن به یک "موضوع داغ" در حوزه مدیریت فرآیندهای کسب و کار است. علاوه براین، در صنعت نیز علاقه بسیار فراوانی برای استفاده از فرآیند کاوی وجود دارد. تعداد بسیاری از تولیدکنندگان نرم‌افزار در حال افزودن قابلیت‌های فرآیند کاوی در ابزارهای خود هستند. نمونه‌ای از محصولات نرم‌افزاری با قابلیت‌های فرآیند کاوی عبارتند از: سیستم مدیریت بهره‌وری فرآیند ARIS (شرکت نرم‌افزاری AG)، سامانه Comprehend (Open Connect)، نرم افزار Discovery Analyst (Stereo LOGIC)، نرم افزار Flow (Fourspark)، نرم افزار Futura Reflect (Futura Process Intelligence)، نرم‌افزار Interstage Automated Process Discovery (Fujitsu)، ابزار فرآیند کاوی OKT (Exeura)، نرم‌افزار Process Discovery Focus (Iontas/Verint)، تحلیلگر فرآیند (QPR)، ProM(TU/e)، Rbminer/Dbminer

[1] - buzzword
[2] - business activity monitoring
[3] - complex event processing
[4] - corporate performance management
[5] - continuous process improvement
[6] - business process improvement
[7] - total quality management

(UPC) و Reflect|one (Pallas Athena). انگیزه اصلی ایجاد یک بخش ویژه در IEEE برای فرآیند کاوی علاقه روز افزون به آنالیز فرآیندهای مبتنی بر نگاره، بود.

این بخش ویژه در سال ۲۰۰۹ در قالب کمیته فنی داده کاوی (DMTC) جامعه هوش محاسباتی (CIS) موسسه مهندسین برق و الکترونیک (IEEE) تاسیس گردید. اعضای فعلی این بخش عبارتند از: تولید کنندگان نرم‌افزار (نظیر Pallas Athena، SoftwareAG، Futura Process Intelligence، HP، IBM، Infosys، Fluxicon، Businesscape، Iontas/Verint، Fujitsu، Fujitsu Laboratories، Business Process Mining، Stereologic)، شرکت‌های مشاور/کاربران نهایی (نظیر ProcessGold، Business Process Trends، Gartner، Deloitte، Process Sphere، Siav SpA، BPM Chili، BWI Systeme GmbH، Excellentia BPM، Rabobank) و موسسات تحقیقاتی (نظیر TU/e، دانشگاه Padua، Universitat Polit`ecnica de Catalunya، دانشگاه نیومکزیکو، دانشگاه فنی لیسبون، University of Calabria، Penn State University، University of Bari، Humboldt-Universit¨at zu Berlin، Queensland University of Technology، Vienna University of Economics and Business، دانشگاه هیفا، University of Bologna، Ulsan National Institute of Science and Technology، دانشگاه Cranfield، K.U. Leuven، دانشگاه Tsinghua، دانشگاه Innsbruck، دانشگاه Tartu).

اهداف دقیق این بخش ویژه عبارتست از:

- آگاهی رسانی به کاربران نهایی، توسعه دهندگان نرم افزار، مشاوران، مدیران کسب و کار و محققان در مورد آخرین وضعیت فرآیند کاوی،
- توسعه استفاده از تکنیک‌های فرآیند کاوی و ابزارها و ایجاد انگیزه برای کاربردهای جدید،
- ایفای نقش در راستای استاندارد سازی برای ثبت داده‌های رویداد،
- برای سازمان دهی آموزش‌ها، نشست‌های ویژه، کارگاههای آموزشی، پنل‌ها و
- انتشار مقالات، کتب، ویدئوها و نشریات ویژه در مجلات.

از زمان تاسیس این انجمن در سال ۲۰۰۹ تا به اکنون فعالیت‌های زیادی در راستای اهداف فوق الذکر انجام شده است. به عنوان مثال، چندین کارگاه آموزشی نظیر کارگاه‌هایی در ارتباط با هوش فرآیندهای سازمانی (BPI'09, BPI'10 و BPI'11) و همچنین بخش‌های ویژه در کنفرانس اصلی IEEE (نظیر CIDM'11) توسط این انجمن برگزار شده است. دانش این حوزه از طریق آموزش (نظیر WCCI'10 و PMPM'09)، مدارس تابستانی (ESSCaSS'09، ACPN'10، CICH'10 و ...)، ویدئوها (در سایت www.processmining.org موجود است) و چندین نشریه شامل اولین کتاب فرآیند کاوی که اخیرا توسط اشپرینگر به چاپ رسیده[۱]، منتشر شده است. انجمن همچنین اولین رقابت هوش فرآیندهای کسب و کار (BPIC'11)، (اولین رقابت که در آن شرکت کنندگان باید دانش بامعنایی را از یک مجموعه بزرگ و پیچیده از نگاره‌های رویداد استخراج نمایند) را سازمان دهی کرده است. همچنین انجمن در سال ۲۰۱۰، XES را استانداردسازی نمود (www.xes-standard.org)، یک استاندارد برای قالب ثبت کردن داده‌های رویداد که قابل توسعه بوده و توسط کتابخانه *OpenXES* (www.openxes.org) پشتیبانی می‌شود و در ابزارهایی نظیر ProM،

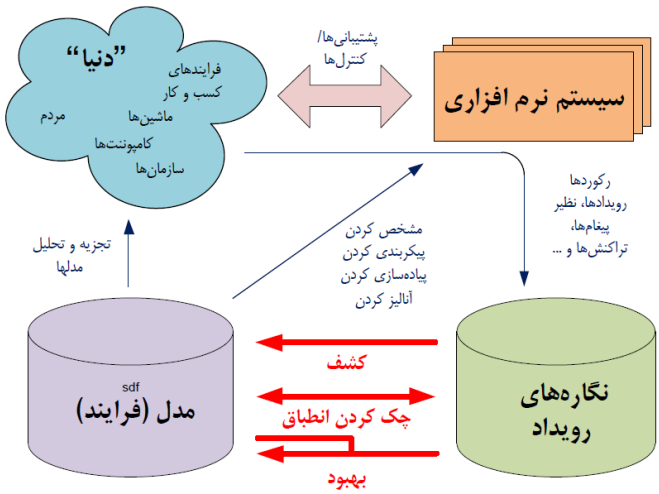
XESame، Nitro و ... بکار رفته است. از خوانندگان محترم دعوت می‌شود تا برای کسب اطلاعات بیشتر در مورد فعالیت‌های این انجمن به سایت http://www.win.tue.nl/ieeetfpm/ مراجعه نمایند.

۲- فرآیند کاوی: وضعیت کنونی

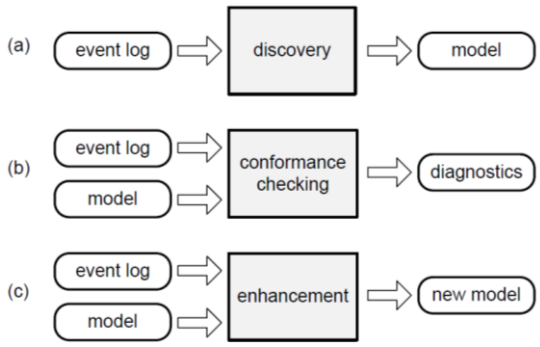
قانون Moore، توسعه قابلیت‌های سیستم‌های اطلاعاتی و سایر سیستم‌ها که وابسته به محاسبات است، را به خوبی توصیف کرده است. گوردن موری از بنیانگذاران شرکت اینتل، در سال ۱۹۶۵ پیش‌بینی نمود که تعداد اجزای موجود در مدارهای مجتمع هر سال دو برابر خواهد شد. طی پنجاه سال گذشته این رشد به درستی نمایی بوده ولو اینکه سرعت آن کمی کندتر بوده است. این پیشرفت‌ها منجر به رشد دیدنی "دنیای دیجیتال" شده است (مثلا تمامی داده‌هایی به صورت الکترونیکی ذخیره و/یا تبادل می‌شوند). علاوه براین، جهان حقیقی و جهان دیجیتال بیشتر و بیشتر به سمت همدیگر سوق می‌یابند.

رشد دنیای دیجیتال که به خوبی منطبق با فرآیندهای سازمان‌ها می‌باشد، ثبت و آنالیز رویدادها را امکان پذیر ساخته است. رویدادها می‌تواند طیف گسترده‌ای از اتفاقات نظیر نقدکردن پول از دستگاه‌های خودپرداز، تنظیم دستگاه اشعه ایکس توسط یک پزشک، درخواست گواهینامه توسط یک شهروند، ارسال اظهارنامه مالیاتی و یا دریافت یک شماره بلیط الکترونیکی توسط یک مسافر را شامل شود. چالش مهم، استخراج رویدادها در قالب‌های بامعنایی است که منجر به ارزش افزوده گردد، به عنوان مثال شناخت دقیقتر فرآیند، شناسایی گلوگاه‌ها، پیش‌بینی مشکلات، ثبت و ضبط نقض قوانین، توصیه‌های مفید و ساده کردن فرآیندها. فرآیند کاوی دقیقا به منظور رسیدن به این اهداف ارائه شده است.

نقطه شروع برای فرآیند کاوی، یک فایل نگاره رویداد است. در تمامی تکنیک‌های فرآیند کاوی فرض می‌شود که ثبت رویدادها به صورت *متوالی* بطوریکه هر رویداد به یک فعالیت (یک گام خوش تعریف در فرآیندها) اشاره داشته و مرتبط با یک case (نمونه فرآیند) خاص باشد، امکان پذیر است. نگاره‌های رویداد ممکن است که حاوی اطلاعات اضافی پیرامون رویدادها نیز باشد. در حقیقت، تکنیک‌های فرآیند کاوی هر زمان که ممکن باشد از اطلاعات اضافی نظیر *منابعی* (مثل شخص یا وسیله) که در یک فعالیت مصرف یا مقداردهی اولیه می‌شوند، *برچسب زمانی* رویدادها، یا عناصر داده‌ای که با یک رویداد ثبت می‌شوند (نظیر مقدار یک سفارش) و ...، استفاده می‌نمایند.



شکل ۲ – جایگاه سه بخش اصلی فرایند کاوی الف) کشف، ب) چک کردن انطباق، و ج)بهبود [۱]



شکل ۳– سه بخش اصلی فرایند کاوی بر حسب ورودی و خروجی توضیح داده شده است: الف) کشف، ب) چک کردن انطباق، و ج) بهبود

همانطور که در *شکل ۲* نشان داده شده است، نگاره‌های رویداد در هر سه شاخه فرآیند کاوی استفاده می‌شوند. اولین نوع فرآیند کاوی، کشف فرآیند می‌باشد. یک تکنیک کشف فرآیند، به عنوان ورودی یک نگاره رویداد دریافت کرده و یک مدل خروجی بدون داشتن هیچ گونه اطلاعات قبلی تولید می‌کند. کشف فرآیند رایج‌ترین تکنیک فرآیند کاوی است. برای بسیاری از سازمان‌ها جالب و جذاب است که تکنیک‌های موجود، محققا می‌توانند فرآیندهای واقعی را براساس نمونه نگاره‌های رویداد اجرا شده کشف نمایند. دومین شاخه فرآیند کاوی، چک کردن انطباق است. در این دسته از روش‌ها، یک مدل موجود با نگاره‌های رویداد همان فرآیند مقایسه می‌شود. چک‌کردن انطباق برای بررسی اینکه آیا واقعیتی که در نگاره‌ها ذخیره شده با مدل فرآیند موجود منطبق است و برعکس، استفاده می‌شود. باید توجه شود که در این تعریف، انواع مدل‌های فرآیند می‌تواند استفاده شود. می‌توان چک‌کردن انطباق را برروی مدل‌های رویه‌ای، مدل‌های سازمانی، مدل‌های فرآیند توصیفی، قوانین/سیاست‌های کسب و کار و ... اعمال نمود. سومین شاخه از فرآیند کاوی، بهبود می‌باشد. ایده اصلی این است که با استفاده از اطلاعات موجود در مورد فرآیندهای واقعی در حال اجرا در سازمان‌ها که در نگاره‌های رویداد ثبت شده، فرآیندهای جاری را بهبود یا توسعه داد. درحالیکه چک کردن انطباق، میزان انطباق مدل با واقعیت را اندازه‌گیری می‌کند، سومین شاخه فرآیند کاوی به توسعه یا تغییر یک مدل پیشین می‌پردازد. به عنوان مثال با استفاده از برچسب زمان در نگاره‌های رویداد می‌شود مدلی تولید کرد که گلوگاه‌ها، سطوح سرویس دهی، زمان انجام کار و فرکانس رخدادها را نشان دهد.

شکل ۳ سه دسته فرایند کاوی را بر اساس ورودی و خروجی نشان می‌دهد. تکنیک‌های کشف فرآیند یک نگاره رویداد را به عنوان ورودی دریافت کرده و یک مدل به عنوان خروجی تولید می‌کنند. مدل شناسایی شده غالبا یک مدل فرآیند (نظیر Petri Net، BPMN، EPC و یا نمودار فعالیت UML) می‌باشد، اما به هرحال ممکن است که مدل، سایر جنبه ها را نیز توصیف نماید (نظیر شبکه‌های اجتماعی). تکنیک‌های چک‌کردن انطباق به یک نگاره رویداد و یک مدل به عنوان ورودی نیاز دارند. خروجی شامل اطلاعات تشخیصی است که تفاوت‌ها و اشتراکات بین مدل و نگاره را نشان می‌دهد. تکنیک‌های بهبود مدل (اصلاح یا توسعه) نیز به یک مدل و یک نگاره رویداد به عنوان ورودی نیاز دارند. خروجی این دسته از روش‌ها، یک مدل اصلاح شده یا توسعه یافته است.

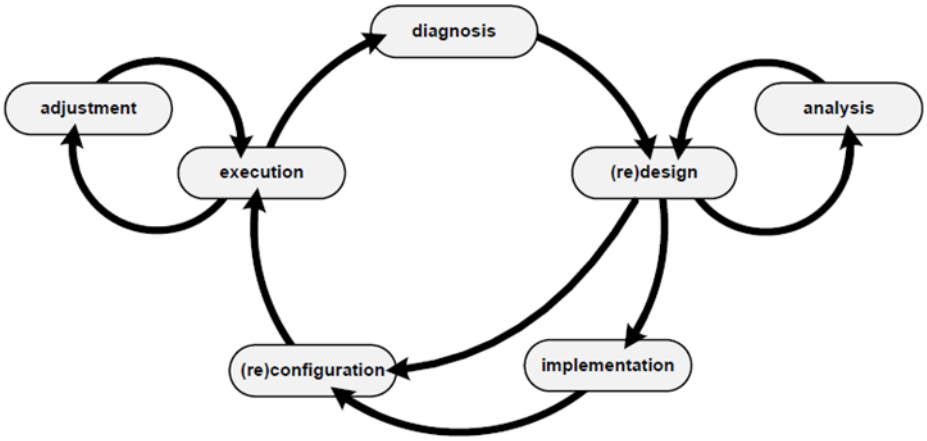
فرایند کاوی جنبه‌های متفاوتی را پوشش می‌دهد. از دیدگاه کنترل جریان، تمرکز اصلی برروی کنترل جریان است، به عنوان مثال ترتیب فعالیت‌ها. هدف از کاوش این جنبه، یافتن خصوصیات خوب همه مسیرهای ممکن است. به طور معمول نتایج برحسب یک Petri net و یا سایر نمادهای نمایش فرآیند (EPC، BPMN، و یا نمودار فعالیت UML) نشان داده می‌شود. جنبه کاوش سازمانی بر اطلاعات پیرامون منابع پنهان شده در نگاره تمرکز دارد، مثلا چه نقش‌هایی(نظیر مردم،

سیستم‌ها، نقش‌ها یا واحدها) دخیل هستند و چطور با یکدیگر مرتبط می‌باشند. هدف، ساختاردهی سازمان با طبقه بندی مردم برحسب نقش‌ها یا واحدهای سازمانی و یا نمایش شبکه اجتماعی است. جنبه case، بر مشخصات و خواص نمونه فرایندها تمرکز دارد. بدیهی است که یک نمونه فرایند با استفاده از مسیرش یا با نقشی که در آن فعالیت دارد، مشخص می‌شود. اما به هر حال، یک نمونه فرایند می‌تواند با استفاده از عناصر داده‌ای متناظر نیز مشخص گردد. به عنوان مثال، اگر یک نمونه فرایند نشان‌دهنده پرکردن یک سفارش باشد، دراینصورت دانستن تامین‌کننده یا تعداد سفارش می‌تواند جذاب باشد. جنبه زمانی، مرتبط با زمانبندی و فرکانس رویدادها است. زمانیکه رویدادها دارای برچسب زمانی باشند، شناسایی گلوگاه‌ها، اندازه‌گیری سطوح سرویس‌دهی، نظارت بر بکارگیری منابع، و پیش‌بینی زمان پردازش باقی‌مانده نمونه‌های جاری امکان‌پذیر خواهد بود.

در ارتباط با بحث فرآیند کاوی تصورات غلطی وجود دارد. برخی از تولید کنندگان نرم‌افزار، آنالیزورها، و محققین، حوزه فرایند کاوی را محدود به تکنیک‌های خاص داده‌کاوی برای کشف فرآیند که فقط به صورت برون‌خط قابل استفاده است، می‌دانند. این موضوع صحیح نیست، بنابراین ما برروی سه ویژگی زیر تاکید می‌کنیم.

- فرایند کاوی محدود به کشف جریان–کنترل نیست. کشف فرآیند از روی نگاره‌های رویداد مهمترین بخش فرایند کاوی است. اما به هر حال، فرآیند کاوی محدود به کشف جریان–کنترل نیست. از یک طرف، کشف فرآیند یکی از سه شاخه اصلی فرآیند کاوی است (کشف، چک کردن انطباق و بهبود). از طرف دیگر حوزه، محدود به جریان–کنترل نیست و جنبه‌های سازمانی و زمان هم مهم بوده و نقش مهمی بازی می‌کند.
- فرآیند کاوی فقط یک نوع خاص از داده‌کاوی نیست. فرآیند کاوی می‌تواند به عنوان یک لینک مفقوده بین داده کاوی و BPM سنتی مبتنی بر مدل می‌باشد. اکثر روش‌های داده‌کاوی به هیچ عنوان فرآیند محور نیستند. مدل‌های فرآیند که بحث همزمانی را هم پوشش می‌دهند، قابل مقایسه با ساختارهای ساده داده‌کاوی نظیر درخت تصمیم و قوانین انجمنی نیستند. بنابراین، الگوریتم‌ها و شیوه‌های کاملا جدید نمایش مورد نیاز است.
- فرآیند کاوی محدود به پردازش‌های برون‌خطی[1] نیست. تکنیک‌های فرآیند کاوی، دانش را از داده‌های تاریخچه‌ای که در سیستم‌های اطلاعاتی امروزی موجود است، استخراج می‌نمایند. اگرچه که "داده‌های پس‌از واقعه[2]" مورد استفاده قرار می‌گیرد، اما نتایج قابل اعمال بر نمونه‌های جاری نیز هست. به عنوان مثال، زمان پایان کار سفارش یک مشتری که بخشی از آن انجام شده، توسط مدل کشف‌شده قابل به پیش‌بینی است.

[1] - offline
[2] - post mortem data



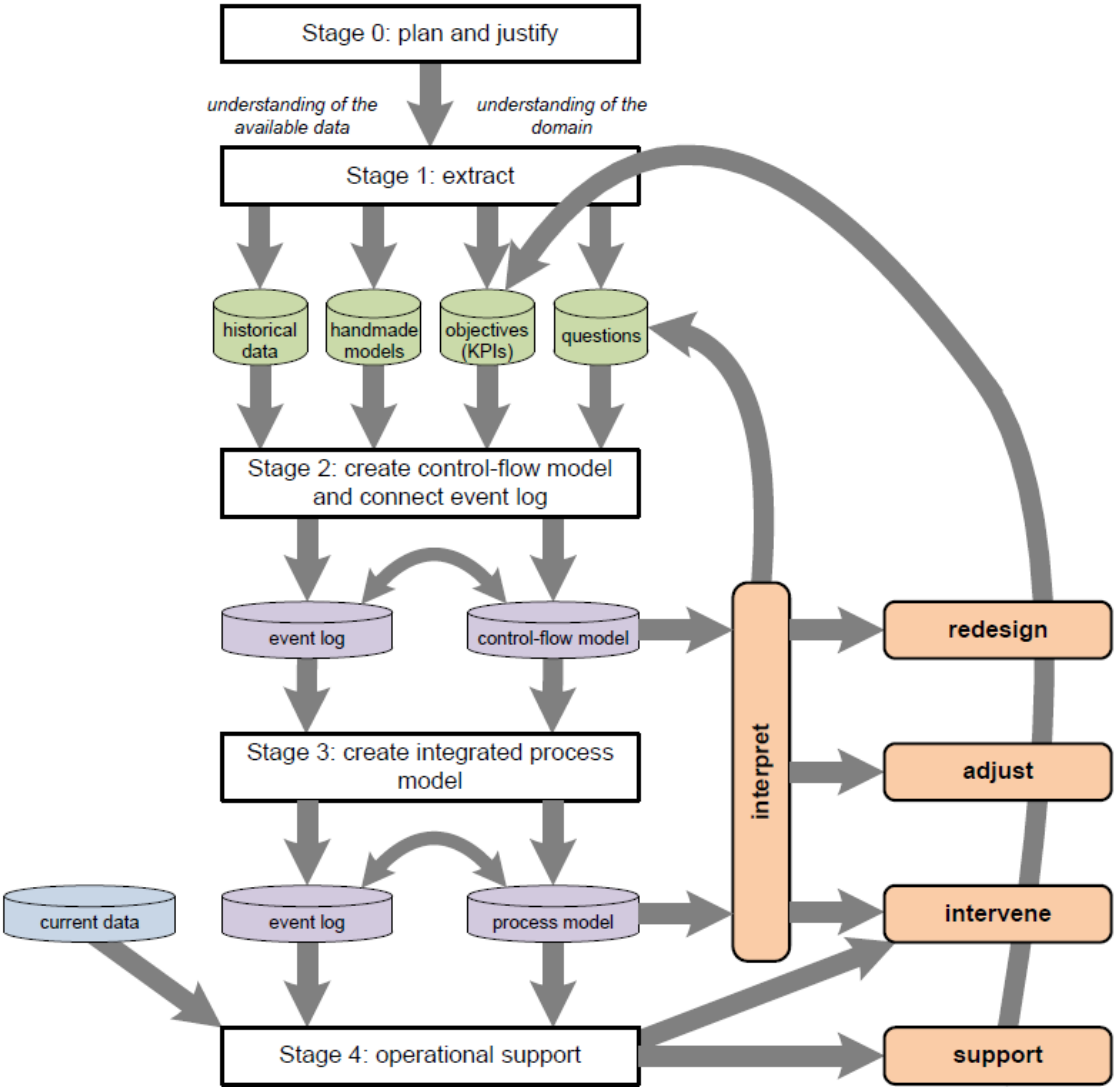
شکل ٤– چرخه عمر BPM فازهای مختلف یک فرایند کسب و کار و سیستم‌های اطلاعاتی مرتبط آن نشان می‌دهد؛ فرایند کاوی (بالقوه) در تمامی فازها نقش ایفا می‌کند (بجز فاز پیاده‌سازی)

برای مشخص کردن جایگاه فرآیند کاوی، از چرخه عمر مدیریت فرآیندهای کسب و کار(BPM) نشان داده شده در *شکل ۴* استفاده کرده‌ایم. چرخه زندگی BPM هفت فاز یک فرآیند کسب و کار و سیستم های اطلاعاتی مرتبط با آن را نشان می‌دهد. در فاز طراحی (مجدد)[1] یک مدل فرآیند جدید ایجاد می‌شود و یا اینکه یک مدل موجود برروز می‌گردد. در فاز آنالیز، یک مدل کاندید و جایگزین‌های آن آنالیز می‌شوند. بعد از فاز طراحی (مجدد) مدل پیاده‌سازی می‌شود (فاز پیاده سازی) و یا سیستم موجود پیکره‌بندی(مجدد) می‌شود. در فاز اجرا، مدل طراحی شده تصویب می‌شود. در حین فاز اجرا فرآیند مانیتور می‌شود. ضمن اینکه ممکن است تنظیمات کوچکی بدون طراحی مجدد فرآیند اعمال گردد (فاز تنظیم[2]). در فاز تشخیص[3]، فرآیند تصویب شده آنالیز شده و خروجی این فاز ممکن است که منجر به یک فاز طراحی مجدد فرآیند گردد. فرآیند کاوی ابزار ارزشمندی برای اکثر فازهای نشان داده شده در *شکل ۴* می‌باشد. پرواضح است که فاز تشخیصی می‌تواند از فرآیند کاوی بهره‌مند گردد. به هر حال، فرآیند کاوی محدود به فاز تشخیصی نیست. به عنوان مثال در فاز اجرا، تکنیک‌های فرآیند کاوی می‌توانند برای پشتیبانی عملیاتی بکار روند. برای نمونه‌های جاری، براساس مدل‌های یاد گرفته شده از اطلاعات قدیمی، می‌توان پیش‌بینی و یا توصیه نمود. روش‌های مشابه پشتیبانی تصمیم برای تنظیم فرآیندها و یا کمک در پیکربندی فرآیندها می‌تواند استفاده شود.

درحالیکه *شکل ۴* چرخه زندگی کلی BPM را نشان می‌دهد، *شکل ۵* برروی فعالیت‌های دقیق فرآیند کاوی و اجزای آن تمرکز دارد. *شکل ۵*، مراحل ممکن در یک پروژه فرآیند کاوی را نشان می‌دهد. تمامی پروژه‌های فرآیند کاوی با یک برنامه‌ریزی و توجیه برای این برنامه‌ریزی آغاز می‌گردند (فاز ۰). بعد از شروع پروژه، باید داده‌های رویداد، مدل‌ها، اشیا و سوالات، از سیستم‌ها، مدیریت و خبره‌های حوزه استخراج گردد (فاز۱). این امر نیازمند فهم مناسب داده‌های موجود (چه چیزهایی برای آنالیز می‌تواند استفاده شود؟) و درک صحیح دامنه (سوالات مهم چه چیزهایی هستند؟) و نتایج اجزای نشان داده شده در *شکل ۵* (نظیر داده‌های قدیمی، مدل‌های دستی‌ساز، اشیا و سوالات) است. در مرحله ۲، مدل جریان–کنترل ساخته شده و به نگاره‌های رویداد متصل می‌گردد. در اینجا تکنیک‌های خودکار کشف فرآیند می‌تواند استفاده شود. مدل‌های فرآیند کشف شده ممکن است که برای برخی از سوالات پاسخ‌هایی ارائه کند و عملیات طراحی مجدد و یا تنظیم را آغاز کند. علاوه براین

[1] - design phase
[2] - adjustment
[3] - diagnosis

ممکن است که نگاره‌های رویداد با استفاده از مدل، فیلتر و یا سازگار شوند (به عنوان مثال فعالیت‌هایی که کم رخ می دهند و یا نمونه‌های پرت، حذف شوند و یا رویدادهای گم‌شده اضافه شوند). گاهی اوقات برای جمع آوری رویدادهای مرتبط با یک نمونه فرآیند، تلاش بسیاری باید انجام داد. رویدادهای باقی مانده به عناصر مدل فرآیند مرتبط هستند. زمانی که فرآیند نسبتا ساخت‌یافته باشد، ممکن است که طی مرحله ۳ مدل جریان–کنترل با جنبه‌های دیگر (نظیر زمان، داده و منابع) توسعه داده شود. رابطه بین نگاره رویداد و مدلی که در مرحله ۲ ایجاد شد، برای توسعه مدل استفاده می‌شود (به عنوان مثال برچسب زمانی رویدادهای مرتبط برای تخمین زمان انتظار فعالیت‌ها استفاده می‌شود). از این موضوع برای پاسخ‌گویی به سوالات دیگر ممکن است استفاده شود و منجر به فعال شدن اقدامات دیگر شود. در نهایت، مدل‌هایی که در مرحله ۳ تولید شدند، ممکن است در مرحله ۴ برای پشتیبانی عملیاتی استفاده شوند. دانشی که از داده‌های رویداد تاریخی استخراج می‌گردد با اطلاعات مربوط به نمونه‌های جاری ترکیب می‌گردد و از آن برای پیش‌گویی و توصیه استفاده می‌شود. مراحل ۳ و ۴ فقط درصورتی دست‌یافتنی است که فرآیند به اندازه کافی ساخت‌یافته و پایدار باشد.



شکل ۵– مدل چرخه زندگی *L یک پروژه فرایند کاوی شامل ۵ فاز را توصیف می‌کند: برنامه ریزی تنظیم (فاز ۰)، استخراج (فاز ۱)، ایجاد یک مدل کنترل–جریان و اتصال آن به نگاره رویداد (فاز۲)، ایجاد یک مدل فرایند یکپارچه (فاز ۳)، و ایجاد پشتیبانی عملیاتی (فاز ۴) [۱]

در حال حاضر تکنیک‌ها و ابزارهایی وجود دارند که تمامی مراحل *شکل ۵* را پشتیبانی می‌کنند. به هر حال، فرآیند کاوی نسبتا شاخه جدیدی است و اکثر ابزارهای موجود در این زمینه هم نابالغ هستند. علاوه بر این، کاربران آینده‌نگر غالبا از پتانسیل و محدودیت‌های فرآیند کاوی آگاه نیستند. بنابراین، این مانیفست برخی از اصول راهنما (بخش ۳) و چالش‌های (بخش ۴) پیش روی کاربران تکنیک‌های فرآیند کاوی و همینطور محققان و توسعه‌دهندگان این حوزه را فهرست کرده است.

۳- اصول راهنما

در زمان استفاده از فرآیند کاوی در کاربردهای واقعی همانند هر تکنولوژی جدیدی، اشتباهات آشکاری ممکن است رخ دهد. بهمین دلیل، در بخش ۶ اصل راهنما برای جلوگیری از بروز این اشتباهات آورده شده است.

۳-۱- اصل اول: داده‌های رویداد باید بمانند شهروندان درجه یک نگریسته شوند

نقطه شروع هر فعالیت فرآیند کاوی، رویدادهای ثبت شده است. ما مجموعه رویدادها را نگاره رویداد می‌نامیم، اما این بدان معنا نیست که رویدادها الزاما باید در یک فایل نگاره اختصاصی ذخیره شوند. رویدادها ممکن است که در جداول پایگاه داده‌ها، نگاره‌های پیغام، آرشیو پستی، نگارههای تراکنش، و سایر منابع داده ذخیره شده باشند. کیفیت داده‌های رویداد از قالب ذخیره‌سازی آنها مهمتر است. کیفیت نتایج فرآیند کاوی به شدت وابسته به کیفیت داده‌های ورودی است. بهمین دلیل نگاره‌های رویداد در آنالیز فرآیندهای سیستم‌های اطلاعاتی باید به مانند شهروندان درجه یک نگریسته شوند. متاسفانه نگاره‌های رویداد غالبا فقط یک محصول فرعی هستند که برای عیب‌یابی یا پروفایل‌گیری استفاده می‌شوند. به عنوان مثال، "ابزارهای پزشکی فیلیپس" رویدادها را خیلی ساده ذخیره می‌کند. توسعه دهندگان این محصولات فقط دستورات چاپ را در کد وارد می‌کنند. اگرچه که برخی راهنماهای غیر رسمی برای اضافه کردن اینچنین دستوراتی در کد وجود دارد، اما برای افزایش کیفیت داده‌های رویداد نیاز به روش‌های سیستماتیک‌تری است.

تاکنون چندین معیار برای ارزیابی کیفیت داده‌های رویداد ارائه شده است. رویدادها باید *قابل اعتماد*[1] باشند، یعنی باید بتوان اعتماد کرد که داده‌های ذخیره شده واقعا رخ داده‌اند و صفات مربوط به داده‌ها درست هستند. داده‌های رویداد باید *کامل*[2] باشند، مثلا برای داده‌های مربوط به یک حوزه خاص نباید رویدادها گم شوند. هر داده ذخیره شده باید معنای خوش تعریفی[3] داشته باشد. علاوه براین موارد، داده‌های رویداد باید امن بوده و موارد مربوط به امنیت و حریم شخصی[4] در زمان ذخیره کردن آنها رعایت شده باشد. به عنوان مثال، کنشگران[5] باید از نوع رویدادهایی که ذخیره می‌شوند و شیوه استفاده از آنها آگاه باشند.

جدول ۱ پنج سطح بلوغ نگاره رویداد را که از کیفیت عالی (♥♥♥♥♥) تا کیفیت ضعیف (♥) می‌تواند باشد، تعریف کرده است. به عنوان مثال، نگاره‌های رویداد "ابزارهای پزشکی فیلیپس" در سطح ♥♥♥ می‌باشند؛ رویدادها به صورت خودکار ذخیره می‌شوند و رفتارهای ذخیره شده منطبق با واقعیت است، اما از روش سیستماتیک مشخصی برای انتساب معنا به رویدادها

[1] - trustworthy
[2] - complete
[3] - well-defined semantics
[4] - privacy concerns
[5] - actors

و اطمینان از پوشش یک سطح خاص، استفاده نشده است. تکنیک‌های فرآیند کاوی برروی داده‌های با کیفیت ♥♥♥♥♥، ♥♥♥♥ و ♥♥♥ قابل اعمال هستند. در اصل امکان اعمال روش‌های فرآیند کاوی برروی داده‌های با سطح ♥♥ و ♥ هم وجود دارد. اما به هرحال، آنالیز داده‌های اینچنینی با دشواری‌هایی روبرو است و نتایج قابل اعتماد نمی‌باشد. درحقیقت، استفاده از داده‌های باسطح ♥ چندان جالب نیست. سازمان‌ها باید سعی کنند برای بهره مند شدن از فرآیند کاوی داده‌ها را با بالاترین کیفیت ممکن ذخیره نمایند.

جدول ۱- سطح بلوغ نگاره‌های رویداد

سطح	خصوصیات
♥♥♥♥♥	بالاترین سطح: نگاره رویداد بالاترین کیفیت را در این سطح دارد (قابل اعتماد و کامل) و رویدادها خوش-تعریف هستند. رویدادها به صورت خودکار، سیستماتیک، قابل اعتماد و امن ذخیره می‌شوند. ملاحضات امنیتی و مسائل مرتبط با حریم شخصی به میزان کافی در این داده‌ها لحاظ شده است. علاوه براین، رویدادهای ذخیره شده (و تمامی صفات آنها) دارای معنای واضحی هستند و این نشانه وجود یک یا چند هستان‌شناسی[1] است. رویدادها و صفاتشان به این هستان‌شناسی‌ها اشاره می‌کنند. *مثال*: نگاره‌هایی که در سیستم‌های BPM به صورت معنایی نشانه گذاری شده‌اند.
♥♥♥♥	رویدادها به صورت خودکار، سیستماتیک و قابل‌اعتماد ذخیره شده‌اند و کامل و قابل اعتماد هستند. برخلاف سیستم‌هایی که در سطح ♥♥♥ هستند، مفاهیمی چون نمونه فرآیند (case) و فعالیت به صورت صریح پشتیبانی می‌شوند. *مثال*: نگاره‌های رویداد سیستم‌های سنتی BPM/workflow.
♥♥♥	رویدادها به صورت خودکار ذخیره می‌شوند، اما روش سیستماتیکی پشت آن وجود ندارد. به هر حال، برخلاف نگاره‌های سطح ♥♥، تاحدودی تضمین می‌کند که رویدادهای ذخیره شده منطبق بر واقعیت باشند (مثلا نگاره‌های رویداد قابل اعتماد هستند ولی لزوما کامل نباشند). به عنوان مثال رویدادهای ذخیره شده در یک سیستم ERP را در نظر بگیرید. اگرچه که رویدادها باید از گستره وسیعی از جداول استخراج گردند، میتوان فرض کرد که اطلاعات درست هستند (مثلا اگر که فرض کنیم یک پرداخت ذخیره شده توسط یک سیستم ERP، در واقعیت هم وجود دارد و بلعکس، مشکلی ایجاد نمی‌کند). *مثال*: جداول در سیستم‌های ERP، نگاره‌های رویداد سیستم‌های مدیریت ارتباط با مشتری، نگاره‌های تراکنش سیستم‌های تبادل پیام، نگاره‌های رویدادهای سیستم‌های با تکنولوژی بالا و ...
♥♥	رویدادها به صورت خودکار ذخیره می‌شوند، به عنوان مثال در بعضی از سیستم های اطلاعاتی به عنوان محصول فرعی تولید می‌شود. روش سیستماتیکی برای تصمیم گیری در مورد رکوردهای ذخیره شده وجود ندارد. علاوه بر آن، امکان دور زدن سیستم‌های اطلاعاتی هم هست. از اینرو ممکن است که رویدادها گم شوند و یا درست ذخیره نشوند. *مثال*: نگاره‌های رویداد سیستم‌های مدیریت اسناد و محصول، نگاره‌های خطاهای سیستم‌های تعبیه شده، برگه‌های مهندسین خدمات و ...
♥	پایین‌ترین سطح: نگاره‌های رویداد کیفیت پایینی دارند. ممکن است که رویدادهای ذخیره شده ارتباطی با واقعیت نداشته باشند و ممکن است رویدادها گم شوند. نگاره‌های رویدادی که به صورت دستی ذخیره شده‌اند، غالبا این خصوصیات را دارند. *مثال*: نشانه‌ها و علامت‌هایی که در اسناد کاغذی غالبا گذاشته می‌شود، رکوردهای پزشکی مبتنی بر کاغذ و ...

[1] - ontology

۳-۲- اصل دوم: استخراج نگاره‌ها باید با سوال انجام شود

همانطور که در شکل ۵ نشان داده شده است، فعالیت‌های فرآیند کاوی باید مبتنی بر سوال باشد. بدون پرسیدن سوالات دقیق، استخراج داده‌های رویداد بامعنی بسیار سخت است. به عنوان مثال، یک سیستم ERP نظیر SAP که شامل هزاران جدول است را در نظر بگیرید. بدون پرسش سوالات دقیق، نمی توان جداول مرتبط برای استخراج داده‌ها را پیدا کرد.

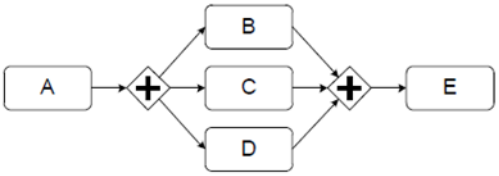
یک مدل فرآیندی نظیر آنچه در شکل ۱ نشان داده شده است، چرخه عمر caseهای (نمونه‌های فرآیند) یک نوع خاص را نشان می‌دهد. از اینرو، پیش از اعمال هر تکنیک فرایند کاوی، نوع نمونه‌ها با استفاده از پرسش و پاسخ باید تجزیه و تحلیل شود که کار ساده‌ای نیست. به عنوان مثال، بررسی سفارشات مشتری را در نظر بگیرید. هر سفارش مشتری ممکن است شامل چند خط سفارش باشد، چراکه غالبا مشتری چند خرید را در یک لیست انجام می‌دهد. یک سفارش مشتری ممکن است منجر به چندین کالا شود. هر تحویل کالا هم ممکن است مرتبط با خطوط سفارش چندین درخواست باشد. بنابراین ارتباط چند-چند[1] بین سفارش و تحویل کالا و ارتباط یک-چند[2] بین سفارشات و خطوط سفارش وجود دارد. با داشتن پایگاه داده‌ایی با داده‌های رویداد مرتبط با سفارشات، خطوط سفارش و تحویل کالاها، مدل‌های فرآیند مختلفی می‌تواند کشف و استخراج گردد. مثلا ممکن است فرآیند با هدف توصیف چرخه عمر تک تک سفارشات استخراج شود، و یا اینکه با هدف کشف چرخه عمر تک تک خطوط سفارش و یا چرخه عمر تحویل کالا، آنالیز و استخراج گردد.

۳-۳- اصل سوم: همزمانی[3]، انتخاب[4] و سایر سازه‌های کنترل-جریان باید پشتیبانی شود

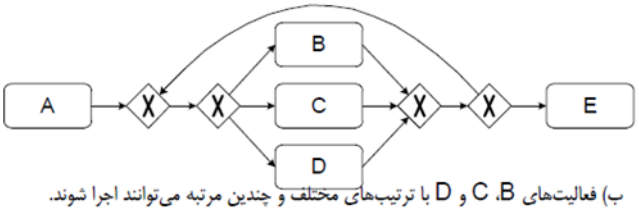
تعداد بسیاری زبان‌های مدل‌سازی فرآیند وجود دارد (مانند BPMN، EPCs، Petri nets، BPEL، و نمودار فعالیت UML). برخی از این زبان‌ها از عناصر بسیاری استفاده می‌کنند (زبان BPMN بیش از ۵۰ عنصر گرافیکی دارد)، در حالیکه برخی دیگر بسیار پایه‌ای هستند (Petri nets تنها شامل سه دسته عنصر است: مکان‌ها[5]، انتقال‌ها[6] و کمان‌ها[7]). توصیف کنترل-جریان، پایه و ستون اصلی هر مدل فرآیند است. سازه‌های گردش کار پایه‌ای (که با نام الگو نیز شناخته می‌شوند) که توسط تمامی زبان‌های اصلی مدل‌سازی فرآیند پشتیبانی می‌شوند عبارتند از: دنباله‌ها[8]، مسیردهی‌های موازی(AND-split/join)، انتخاب (XOR-split/join) و حلقه‌ها. پرواضح است که این الگوها باید توسط تکنیک‌های فرآیندکاوی نیز پشتیبانی شوند. به هرحال، بعضی از تکنیک‌ها از همزمانی پشتیبانی نمی‌کنند و تنها سیستم‌های انتقال/زنجیره‌های Markov را پوشش می‌دهند.

---

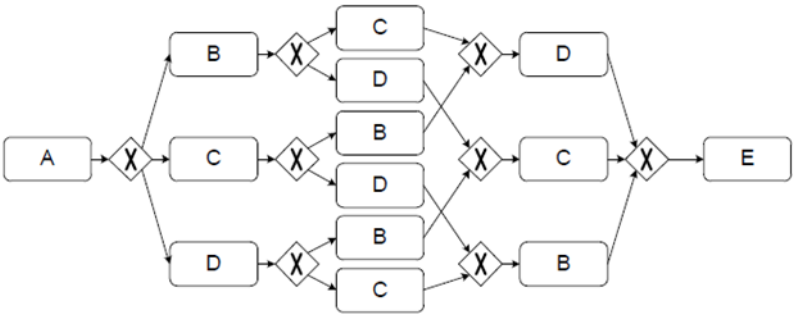
[1] -many to many
[2] - one to many
[3] - concurrency
[4] - choice
[5] - places
[6] - transitions
[7] - arcs
[8] - sequence



الف) فعالیت‌های B، C و D با ترتیب‌های مختلف می‌توانند اجرا شوند.



ب) فعالیت‌های B، C و D با ترتیب‌های مختلف و چندین مرتبه می‌توانند اجرا شوند.



الف) فعالیت‌های B، C و D با ترتیب‌های مختلف می‌توانند اجرا شوند، اما فعالیتها باید تکراری باشند تا همه

شکل ٦– این مثال مشکلاتی که مربوط به زمانیست که همزمانی (نظیر AND-splits/joins) نمی‌تواند به صورت مستقیم بیان شود، را نشان می‌دهد. در این مثال، تنها سه فعالیت (B،C و D) همزمان هستند. نتیجه مدل‌های فرایند در زمانی که تعداد فعالیتهای همزمان ۱۰ عدد است را تصور کنید (۲^۱۰=۱۰۲۴ حالت، ۳،۶۲۸،۶۰۰ = !۱۰ رشته اجرای ممکن).

، اثر استفاده از تکنیک‌های فرآیند کاوی که قادر به کشف همزمانی (no AND-split/joins) نیستند را نشان می‌دهد. نگاره رویداد $L = \{\langle A,B,C,D,E\rangle,\langle A,B,D,C,E\rangle,\langle A,C,B,D,E\rangle,\langle A,C,D,B,E\rangle,\langle A,D,B,C,E\rangle,\langle A,D,C,B,E\rangle\}$ را در نظر بگیرید. L شامل نمونه‌هایی است که با A شروع می‌شوند و به E ختم می‌شوند. فعالیت‌های B، C و D با ترتیب‌های مختلف بین A و E رخ داده‌اند. مدل BPMN نشان داده شده در شکل ۶–الف یک نمایش فشرده از یک فرآیند با استفاده از دو سازه AND است. فرض کنید که یک تکنیک فرآیند کاوی از سازه‌های AND پشتیبانی نکند. نظیر دو فرآیند دیگری که در نشان داده شده است. مدل‌های BPMN شکل ۶–b فشرده بوده اما رفتارهای خیلی زیادی (مثلا نمونه‌هایی نظیر $\langle A,B,B,B,E\rangle$ طبق مدل امکان پذیر بوده اما منطبق با نگاره‌های رویداد نیستند) را نشان می‌دهد. مدل BPMN شکل ۶–c نمونه‌های موجود در L را پوشش می‌دهد، اما صراحتا تمامی دنباله‌ها را کد می‌کند و به همین دلیل، دیگر یک نمایش فشرده از نگاره‌ها نیست. این مثال نشان می‌دهد که برای مدل‌های دنیای واقعی که مستعد داشتن ده‌ها فعالیت همزمان هستند، مدل‌های نتیجه، دارای underfitting بسیار بالا بوده (اجازه تولید رفتارهای متعدد را می‌دهد) و/یا درصورتیکه از همزمانی پشتیبانی نکنند، دارای پیچیدگی بسیاری هستند.

همانطور که در نشان داده شده است، پشتیبانی از الگوهای پایه ای جریان کار مهم است. در کنار الگوهای اساسی مطرح شده، پشتیبانی از OR-splits/joins نیز بسیار مطلوب می‌باشد، چراکه باعث می‌شود تا نمایش فشرده از تصمیمات فراگیر و همگام سازی‌های جزئی، امکان پذیر شود.

۳-٤- اصل چهارم: رویدادها باید مرتبط با المان‌های مدل باشند

همانطور که در بخش ۲ تاکید شد، یک تصور غلط که وجود دارد این است که فرآیند کاوی محدود به کشف کنترل-جریان است. همانطور که در *شکل ۱* نشان داده شد، مدل فرآیند کشف شده ممکن است که جنبه‌های گوناگونی را پوشش دهد (چشم‌انداز سازمانی، بعد زمان، جنبه چشم انداز داده‌ای و ...). علاوه بر این، کشف فرآیند تنها یکی از سه انواع فرآیند کاوی نشان داده شده در *شکل ۳* است. دو نوع دیگر فرآیند کاوی (چک‌کردن انطباق و بهبود) بسیار مبتنی بر ارتباط بین *المان‌های موجود در مدل* و *رویدادهای موجود در نگاره* است. ممکن است این ارتباط برای "بازپخش[1]" نگاره رویداد برروی مدل استفاده شود. بازپخش، برای شناسایی اختلافات موجود بین مدل و نگاره رویداد استفاده می‌شود. به عنوان مثال، برخی از رویدادها در نگاره بر طبق مدل امکان پذیر نیستند. تکنیک‌های چک‌کردن انطباق، این اختلافات را شناسایی و تعیین می‌کنند. برچسب‌های زمانی موجود در نگاره‌های رویداد برای آنالیز رفتارهای موقتی در حین بازپخش می‌تواند استفاده شود. به منظور اضافه کردن زمان انتظار[2] به مدل، می‌توان از تفاوت‌های زمان بین فعالیت‌هایی که به صورت سببی به یکدیگر مرتبط اند، استفاده کرد. این مثال‌ها نشان می‌دهد که ارتباط بین رویدادها در نگاره و المان‌ها در مدل، به عنوان نقطه آغازین انواع مختلفی از آنالیزها بکار می‌رود.

در بعضی از موارد، ایجاد این ارتباط کار آسانی نیست. به عنوان مثال ممکن است که یک رویداد به دو فعالیت مختلف ارجاع کند و یا اینکه مشخص نباشد به چه فعالیتی اشاره می‌کند. به منظور تفسیر صحیح‌تر نتایج فرآیند کاوی، لازم است که این ابهامات از بین برده شوند. در کنار مشکل برقراری ارتباط بین رویدادها و فعالیت‌ها، چالش مشخص‌کردن ارتباط بین رویدادها و نمونه فرآیندها نیز وجود دارد. این مشکل عموما با نام *ارتباط رویداد*[3] شناخته می‌شود.

۳-۵- اصل پنجم: مدل‌ها باید به عنوان انتزاعی هدفمند از واقعیت تلقی شوند

مدل‌های استخراج شده از داده‌های رویداد *دیدهایی از واقعیت* ارائه می‌کنند. اینچنین دیدی باید طوری باشد که یک انتزاع هدفمند از رفتار نهفته در نگاره‌های رویداد ایجاد کند. برروی یک نگاره رویداد ممکن است چندین دید مفید وجود داشته باشد. علاوه بر این، ممکن است که ذینفعان مختلف نیاز به دیدهای متفاوتی داشته باشند. در حقیقت باید به مدل‌های استخراج شده از نگاره‌های رویداد به مانند "*نقشه‌ها*" (مانند نقشه‌های جغرافیایی) نگریست. این اصل راهنما، بینش‌های مهمی را ایجاد می‌کند، که دو نمونه از آنها در ادامه ذکر شده است.

اول از همه، باید توجه کرد که چیز مشابه نقشه برای یک ناحیه جغرافیایی خاص وجود ندارد. بر اساس کاربرد مورد نظر، نقشه‌های مختلفی موجود است: نقشه راه‌ها، نقشه پیاده‌روی، نقشه دوچرخه سواری و .... همه این نقشه‌ها یک دید از یک

---

[1] - replay
[2] - waiting time
[3] - event correlation

واقعیت یکسان به ما می‌دهند و تصور اینکه یک "*نقشه کامل*" داشته باشیم، بیهوده و اشتباه است. همین مطلب در مورد مدل‌های فرآیندی نیز صدق می‌کند. مدل باید بر چیزهایی که مرتبط با یک نوع خاص از کاربر هست، تاکید کند. مدل‌های کشف شده ممکن است برروی دیدگاه‌های مختلفی (کنترل-جریان، جریان داده، زمان، منابع، هزینه و ...) تمرکز داشته باشند و این دیدها را با دقت و درشت دانگی[1] متفاوتی نشان دهند. مثلا ممکن است که یک مدیر بخواهد یک مدل فرآیند غیر رسمی دانه درشت[2] متمرکز بر هزینه‌ها را ببیند، درحالیکه، یک تحلیلگر فرآیند تمایل داشته باشد تا بر جزییات مدل فرآیندی که دارای انحرافاتی از حالت گردش طبیعی است، متمرکز شود. همچنین ممکن است که ذی‌نفعان مختلف بخواهند تا یک فرآیند را با درشت دانگی متفاوتی مشاهده کنند: *سطح استراتژیک* (تصمیمات در این سطح تاثیر طولانی مدت دارند و براساس داده‌های رویداد جمع آوری شده در بازه زمانی طولانی‌تری هستند)، *سطح تاکتیکی* (تصمیمات در این سطح اثر متوسط داشته و غالبا براساس داده‌های رویداد اخیر هستند)، و *سطح عملیاتی* (تصمیمات در این سطح اثر سریع داشته و براساس داده‌های رویداد مرتبط با نمونه‌های در حال اجرا می باشند).

ثانیا، استفاده از ایده‌های نقشه کشی در زمانی که قرار است یک نقشه قابل فهم تولید شود، مفید است. به عنوان مثال، نقشه‌های راه، مسیرها و شهرهایی که کمتر مورد توجه هستند را خلاصه می‌کند. چیزهای کم اهمیت یا حذف شده و یا اینکه به صورت خودکار در قالب اشکال کلی خوشه بندی می‌شوند (مثلا خیابان‌ها و حومه شهرها با شهرها ادغام می‌شوند). نقشه‌کش‌ها نه تنها جزییات بی ربط را حذف نموده، بلکه از رنگ‌ها نیز به منظور مشخص کردن ویژگی‌های مهم استفاده می‌کنند. علاوه بر آن عناصر گرافیکی دارای ابعاد هستند که برای مشخص کردن اهمیتشان استفاده می‌شود (اندازه نقاط و قطر خطوط ممکن است متغییر باشد). همچنین نقشه‌های جغرافیایی دارای یک تفسیر شفاف از محور x و y هستند، به عنوان مثال قالب‌بندی نقشه اختیاری نبوده، چراکه مختصات عناصر نقشه دارای معنی است. همه این موارد در تضاد کامل با مدل‌های فرآیند جریان اصلی[3] هستند که معمولا از معیارهای رنگ، اندازه و مکان برای قابل فهم‌تر شدن مدل‌ها استفاده نمی‌کنند. به هر حال، ایده‌های حوزه نقشه‌کشی به راحتی در ساخت نقشه‌های فرآیندهای کشف شده می‌تواند استفاده شود. به عنوان مثال، اندازه یک فعالیت برای منعکس کردن فرکانس رخداد آن یا برخی ویژگی‌های دیگر که نشان دهنده اهمیت آن هستند، می‌تواند استفاده شود. پهنای یک لینک اهمیت وابستگی سببی مربوطه را می‌تواند نشان دهد، و از رنگ آن برای مشخص کردن گلوگاه‌ها می‌توان استفاده کرد.

مشاهدات بالا نشان می‌دهد که انتخاب نمایش صحیح و تنظیم دقیق آن برای مخاطبان مدنظر حائز اهمیت است. این مشاهدات برای تصویرسازی نتایج برای کاربران نهایی و برای هدایت الگوریتم‌های کشف فرآیند به سمت مدل‌های مناسب، مهم می‌باشد(چالش ۵ را هم ببینید).

۳-۶- اصل ششم: فرآیند کاوی باید یک فرآیند پیوسته باشد

فرآیند کاوی می‌تواند کمک کند تا "نقشه‌های" با معنایی که مستقیما به نگاره‌های رویداد متصل باشند، تهیه شود. هم نگاره‌های رویداد قدیمی و هم نگاره‌های رویداد کنونی می‌توانند به مدل‌های اینچنینی نگاشت شوند. علاوه براین، فرآیندها در حین آنالیز نیز ممکن است تغییر کنند. بنابراین با توجه به ماهیت پویای داده‌ها، توصیه نمی‌شود که به فرآیند کاوی با دید یک

[1] - granularity
[2] - coarse
[3] - mainstream process models

فعالیت یکبار اجرا نگریسته شود. نباید هدف این باشد که یک مدل ثابت از داده ها ایجاد شود، بلکه باید نفس زندگی و پویایی را در فرآیندها ایجاد کرد، طوری که کاربران و تحلیلگران تشویق شوند که به صورت روزانه فرآیندها را بررسی نمایند.

این کاربرد شبیه به استفاده از مشاب‌ها[1] در برچسب گذاری جغرافیایی[2] است. امروزه هزاران مشاب وجود دارند که از نقشه‌های گوگل (به عنوان مثال نرم افزارهایی که وضعیت ترافیک ، بنگاه های املاک، رستوران‌ها و یا زمان پخش فیلم‌ها را برروی یک نقشه انتخاب شده به تصویر می‌کشند) استفاده می‌کنند. کاربران می‌توانند به صورت یکپارچه از امکان بزرگنمایی[3]/کوچک نمایی[4] این نقشه ها استفاده کرده و با آنها تعامل داشته باشند (مثلا شلوغی‌های ترافیکی برروی نقشه ترسیم می شود و کاربر با زوم کردن برروی یک مشکل ترافیکی خاص می‌تواند جزییات بیشتری را مشاهده نماید). همچنین استفاده از فرآیند کاوی برای داده‌های زمان–واقعی[5] نیز باید امکان پذیر باشد. با کمک گرفتن از استعاره نقشه، می‌توانیم فکر کنیم که رویدادها مختصات GPS دارند و می‌توانند برروی یک نقشه در یک زمان حقیقی به تصویر کشیده شوند. ابزارهای فرآیند کاوی مشابه سیستم‌های ناوبری خودرو می‌توانند به کاربران نهایی کمک کنند : الف) با حرکت برروی فرآیند ب) با نگاشت اطلاعات پویا برروی نقشه‌های فرآیند (نظیر نمایش ازدحام ترافیکی در فرآیندهای کسب و کار) و ج) با انجام پیش بینی‌هایی در ارتباط با نمونه‌های در حال اجرا (نظیر تخمین زمان ورود نمونه‌ای که تاخیر دارد). این مثال‌ها ثابت می‌کند که بسیار حیف است که از مدل‌های فرآیند به شکل بهتر و فعالانه‌تری استفاده نشود. بنابراین، به فرآیند کاوی باید به صورت یک روند پیوسته‌ای نگریسته شود که متناسب با بازه‌های زمانی مختلف (دقیقه، ساعت، روز، هفته و ماه) اطلاعات قابل اجرایی را استخراج می‌کند.

٤- چالش‌ها

فرآیند کاوی ابزار مهمی برای سازمان‌های مدرنیست که نیازمند مدیریت فرآیندهای عملیاتی غیر جزئی خود هستند. از یک سو، با رشد باورنکردنی داده‌های رویداد مواجه هستیم. از طرف دیگر، فرآیندها و اطلاعات باید کاملا با هم تراز شوند تا نیازمندی‌های مربوط به انطباق، کارایی و خدمات مشتری برآورده شود. علی رغم کاربردهای بسیار فرآیند کاوی، هنوز چالش‌های بسیاری وجود دارد که نیاز به تحقیق و بررسی دارند. وجود این چالش‌ها ثابت می‌کند که فرآیند کاوی یک شاخه در حال ظهور است. در ادامه، ما برخی از این چالش‌ها را لیست کرده‌ایم. انتظار نمی‌رود که این لیست کامل باشد و با گذشت زمان چالش‌های جدیدی ممکن است بوجود آید و یا اینکه برخی از چالش‌های فعلی با پیشرفت فرآیند کاوی رفع گردد.

٤-١- چالش اول: پیدا کردن، ادغام کردن و تمیز سازی داده های رویداد

علی رغم تکنولوژی‌های موجود، هنوز باید تلاش بسیاری انجام داد تا بتوان داده‌های مناسب برای فرآیند کاوی را استخراج کرد. معمولا چندین مانع وجود دارد که باید بر طرف گردد:

— ممکن است داده‌ها بین چندین منبع توزیع شده باشند. در این صورت باید این اطلاعات ادغام گردند. این موضوع زمانی مشکل ساز می‌شود که در منابع مختلف از شناسه‌های متفاوتی استفاده شود. به عنوان مثال، در یک سیستم

[1] - mashups
[2] - geo-tagging
[3] - zoom in
[4] - zoom out
[5] - real time

ممکن است که از نام و تاریخ تولد برای تشخیص یک فرد استفاده شود در حالی که در سیستم دیگر از کد ملی فرد برای شناسایی او استفاده گردد.

- داده‌های رویداد غالبا ‘‘شی محور[1]’’ هستند تا اینکه ‘‘فرآیند محور’’ باشند. به عنوان مثال، محصولات خاص، ظروف چوپی و یا کانتینرها ممکن است که برچسب های RFID داشته باشند و رویدادهای مربوط به این برچسب ها را ذخیره کنند. در هر حال، برای بررسی سفارش یک مشتری مشخص، نیاز است که این رویدادهای شی–محور ادغام و پیش پردازش شوند.
- نگاره رویداد ممکن است که ناکامل[2] باشد. مشکل رایجی که وجود دارد این است که رویدادها به صورت صریح به نمونه‌های فرآیند اشاره نمی‌کنند. غالبا می‌توان این اطلاعات را از رویدادها استخراج نمود، اما نیاز به تلاش و هزینه قابل توجهی است. همچنین ممکن است که اطلاعات زمانی برای برخی از رویدادها در دسترس نباشد. بنابراین ممکن است نیاز باشد که برچسب‌های زمانی به نوعی در رویدادها وارد و ثبت شود.
- یک نگاره رویداد ممکن است شامل داده های پرت[3] باشد، به عنوان مثال، رفتارهای استثنا هم به صورت نویز در نظر گرفته می شود. چطور باید داده های پرت را تعریف کرد؟ چطور باید این داده های پرت را شناسایی کرد؟ برای اینکه داده ها تمیز و قابل استفاده شوند، باید به این سوالات پاسخ داده شود.
- نگاره‌ها ممکن است که رویدادهایی با سطوح مختلفی از درشت دانگی[4] داشته باشند. در نگاره‌های رویداد سیستم اطلاعاتی یک بیمارستان، ممکن است که رویدادها به یک تست خون ساده و یا یک فرآیند پیچیده جراحی اشاره کنند. حتی برچسب‌های زمانی نیز ممکن است که سطح درشت دانگی متفاوتی داشته باشند، از دقت بر حسب میلی ثانیه (28-9-2011:h11m28s32ms342) گرفته تا اطلاعات درشت فقط مربوط به تاریخ (28-9-2011).
- رویدادها در یک زمینه خاص (نظیر آب و هوا، حجم کاری، روز هفته و غیره) معمولا اتفاق می‌افتند. این زمینه ممکن است که پدیده مسلم و قطعی‌ایی را توضیح دهد؛ مثلا زمان پاسخ به دلیل تعطیلی و یا کارهای در دست انجام از حالت معمول بیشتر باشد. برای تجزیه و تحلیل بهتر، مطلوبست که زمینه را نیز در نظر بگیریم. این بدان معناست که داده‌های رویداد را با داده‌های زمینه‌ای ادغام کنیم.

برای حل مشکلات فوق، نیاز به متدولوژی‌ها و ابزارهای جدیدی است. علاوه بر این، همانطور که پیش از این نیز اشاره شد، سازمان‌ها باید با داده‌های رویداد به دید شهروندان درجه یک برخورد کنند و نه با دید یک محصول جانبی. هدف، رسیدن به نگاره‌های رویداد ۵ ستاره (جدول ۱ را ملاحظه فرمایید) است. در اینجا، درس‌هایی که درباره ایجاد کردن انباره داده آموختیم برای اطمینان از کیفیت بالای نگاره‌های رویداد مفید می‌باشد. به عنوان مثال، چک کردن‌های ساده در حین ثبت داده کمک می‌کند که سهم داده‌های رویداد ناصحیح به میزان قابل توجهی کاهش یابد.

---

[1] - object centric
[2] - incomplete
[3] - outliers
[4] - granularity

٤-٢- چالش دوم: تعامل با نگاره‌های رویداد پیچیده که ویژگی‌های متفاوت دارند

نگاره‌های رویداد می‌توانند ویژگی‌های بسیار متفاوتی داشته باشند. برخی از نگاره‌های رویداد ممکن است به شدت بزرگ باشند طوری که نگهداری و بررسی آنها را مشکل سازد، در حالی که نگاره‌های رویداد دیگری ممکن است آنقدر کوچک باشد که داده‌های کافی برای نتیجه‌گیری قابل اطمینان موجود نباشد.

در برخی از حوزه‌ها، مقادیر بسیار عظیمی از رویدادها ثبت شده است. بنابراین، تلاش مضاعفی برای بهبود کارایی و مقیاس پذیری نیاز است. به عنوان مثال، شرکت ASML تمامی اسکنرهای ویفرش[1] را به صورت پیوسته رصد می‌کند. سازمان‌های مختلفی (نظیر سامسونگ[2] یا شرکت ابزارهای تگزاس[3]) از این اسکنرها برای تولید تراشه‌ها استفاده می‌کنند (تقریبا ۷۰ درصد تراشه‌ها توسط اسکنرهای ویفر ASML تولید می‌شوند). ابزارهای موجود در مواجه با حجم پتابایتی داده‌ها در حوزه‌های اینچنینی مشکل دارند. علاوه بر تعداد رویدادهای ذخیره شده، ویژگی‌های دیگری نیز هستند که باید محاسبه شوند، نظیر متوسط تعداد رویدادها در هر نمونه، شباهت بین نمونه‌ها، تعداد رویدادهای منحصر به فرد و تعداد مسیرهای منحصر به فرد. نگاره رویداد L_1 با ویژگی‌های ذیل را در نظر بگیرید: ۱۰۰۰ نمونه فرآیند، به طور متوسط ۱۰ رویداد به ازای هر نمونه و مقدار کمی هم تفاوت در نمونه‌ها (به عنوان مثال، تعداد زیادی از نمونه‌ها مسیرهای یکسان و یا بسیار مشابه را طی می‌کنند). نگاره‌های رویداد L_2 شامل ۱۰۰ نمونه، و هر نمونه به طور متوسط حاوی ۱۰۰ رویداد و تمامی نمونه‌ها هم یک مسیر واحد را دارند. پر واضح است که علی رغم آنکه سایز هر دو نگاره برابر است (به طور تقریبی ۱۰،۰۰۰ رویداد)، اما آنالیز نگاره L_2 به مراتب سخت از L_1 می‌باشد.

همانطور که نگاره‌های رویداد فقط شامل رویدادهای ساده هستند، اما ممکن است که کامل نباشند. تکنیک‌های فرآیند کاوی باید با بهره گیری از "فرض جهان باز" (این حقیقت که اتفاق نیفتادن یک یک چیز به معنای عدم امکان رخداد آن چیز نیست) با ناکامل بودن داده‌ها تعامل کنند. ناکامل بودن داده‌ها باعث می‌شود که تعامل با نگاره‌های رویداد کوچک که تفاوت زیادی دارند پر چالش باشد.

همانطور که پیش تر نیز اشاره شد، برخی از نگاره‌ها حاوی رویدادهای با سطح انتزاع بسیار پایین هستند. این نگاره ها غالبا بسیار بزرگ هستند و معمولا ذی‌نفعان علاقه کمتری به رویدادهای سطح پایین دارند. بنابراین ممکن است که نیاز باشد تا رویدادهای سطح پایین را باهم تجمیع کرده و تبدیل به رویدادهای سطح بالا نماییم. به عنوان مثال، زمانی که فرآیند تشخیص و درمان گروه مشخصی از بیماران را تجزیه و تحلیل می‌کنیم، ممکن است که علاقه‌ای به تستهای انفرادی افراد که در سیستم اطلاعاتی آزمایشگاه بیمارستان ذخیره شده است، نداشته باشیم.

در این نقطه از زمان، سازمان‌ها از روش‌های سعی و خطا برای تعیین اینکه نگاره‌های رویدادشان مناسب فرآیند کاوی هست یا خیر، استفاده می‌کنند. بهمین دلیل، ابزارها باید اجازه آزمایش‌های امکان‌سنجی سریع برای یک مجموعه داده مشخص را فراهم کنند. اینچنین آزمایش‌هایی باید مشکلات عملکردی بالقوه را شناسایی کرده و در مورد نگاره‌هایی که خیلی ناقص هستند یا خیلی اطلاعات جزئی دارند، هشدار دهند.

---

[1] - Wafer Scanner
[2] - Samsung
[3] - texas Instruments

٤-٣- چالش سوم: ایجاد Representative Benchmarks

با توجه به اینکه فرآیند کاوی یک تکنولوژی نوظهور است، بهمین دلیل هنوز نیاز به محک‌های[1] خوب می‌باشد. به عنوان مثال، ده ها روش و نرم افزار کشف فرآیند وجود دارد و فروشندگان مختلف محصولات متفاوتی را توصیه می‌کنند، اما هیچ اجماعی در مورد کیفیت این تکنیک‌ها وجود ندارد. علی رغم اینکه تفاوت‌های بسیاری در عملکرد و کارایی این ابزارها و تکنیک‌ها وجود دارد، اما مقایسه آنها کار دشواری است. بنابراین، محک‌های خوب شامل مجموعه داده‌های نمونه و معیارهای نشان دهنده کیفیت مورد نیاز است.

برای تکنیک‌های داده کاوی کلاسیک، تعداد زیادی محک وجود دارد. این محک‌ها، باعث شده‌اند تا تولیدکنندگان ابزار و محققان انگیزه بیشتری برای افزایش کیفیت کارایی تکنیک‌هایشان داشته باشند. در مورد فرآیند کاوی این موضوع پرچالش‌تر می باشد. به عنوان مثال، مدل رابطه‌ایی که توسط Codd در ۱۹۶۹ ارائه شد، ساده است و به صورت گسترده پشتیبانی می‌شود. در نتیجه، برای تبدیل از یک پایگاه داده به پایگاه داده دیگری انرژی کمی نیاز است و مشکلات تفسیرهای گوناگون وجود ندارد. اما در مورد فرآیندها، این چنین مدل ساده‌ای وجود ندارد. استانداردهایی که برای مدل کردن فرآیند ارائه شده است بسیار پیچیده‌تر بوده و تعداد کمی از تولید کنندگان هستند که از مفاهیم یکسانی استفاده می‌کنند. فرآیندها، به مراتب پیچیده تر از داده‌های جدولی هستند.

با این اوصاف، ایجاد محک‌های نماینده برای فرآیند کاوی ضروری است. چندین کار ابتدایی در این زمینه انجام شده است. به عنوان مثال، چندین معیار گوناگون (تناسب[2]، سادگی[3]، دقت[4] و تعمیم[5]) برای ارزیابی کیفیت نتایج فرآیند کاوی ارائه شده است. همچنین، چندین نمونه نگاره رویداد به صورت عمومی در دسترس است (رجوع به www.processmining.org). به عنوان مثال می‌توانید نگاره‌های رویدادی که در اولین مسابقه چالش هوش فرآیند کسب و کار که توسط task force سازمان دهی شده، را مشاهده نمایید (رجوع به doi:10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffcf54).

از یک طرف باید benchmarkهایی بر مبنای مجموعه داده‌های واقعی در دسترس باشد. از طرف دیگر نیاز به تولید داده‌های مصنوعی است که ویژگی‌های خاصی داشته باشند. این داده‌های مصنوعی در تولید تکنیک‌های فرآیندکاویی که در خور نگاره‌های رویداد ناکامل، نگاره‌های رویداد نویزی، یا جمعیت مشخصی از فرآیندها هستند، کمک می‌کنند.

در کنار ایجاد representative benchmarks، همچنین نیاز به توافق بیشتری بر سر معیارهایی که برای ارزیابی کیفیت نتایج فرآیند کاوی (چالش ۶ را هم مطالعه کنید) بکار می‌روند، می‌باشد. علاوه بر این، می‌توان از تکنیک‌های ارزیابی متقابل در داده کاوی نیز برای ارزیابی نتایج اقتباس کرد. به عنوان مثال روش چک کردن k-fold را در نظر بگیرید. یک راهکار این است که نگاره‌های رویداد را به k قسمت شکست. سپس k-1 قسمت را برای یادگیری یک مدل فرآیند استفاده کرد و از

---

[1] - benchmarks
[2] - fitness
[3] - simplicity
[4] - precision
[5] - generalization

تکنیک‌های چک کردن انطباق برای ارزیابی نتایج با توجه به قسمت‌های باقی مانده استفاده نمود. این فرآیند را می‌توان k بار تکرار کرد، و بنابراین می‌توان به بینش‌هایی در مورد کیفیت مدل دست یافت.

٤-٤- مواجه با شیفت معنایی [1]

اصطلاح شیفت یا جابجایی معنایی به مواردی اشاره دارد که فرآیند در حالیکه تحت آنالیز است، تغییر کند. به عنوان مثال، ممکن است که در ابتدای نگاره‌های رویداد، دو فعالیت همزمان باشند، در حالیکه در ادامه این دو فعالیت پی‌درپی شوند. فرآیندها ممکن است که به علت تغییرات دوره ای/ فصلی (به عنوان مثال، "در ماه اسفند که نزدیک به سال جدید است، تقاضاهای بیشتری وجود دارد" یا "در عصر روزهای جمعه کارمندان کمتری در شرکت هستند") و یا به دلیل تغییر شرایط (به عنوان مثال، بازار رقابتی‌تر شود) تغییر کنند. این تغییرات برروی فرآیندها تاثیر می‌گذارند و شناسایی و آنالیز آنها حیاتی است. برای کشف شیفت معنایی در یک فرآیند، باید نگاره‌های رویداد را به قسمت‌های کوچکتر شکست و سپس ردپای نگاره‌های کوچکتر را آنالیز کرد. این آنالیزهای مرتبه دوم، نیاز به داده‌های رویداد بسیار بیشتری دارد. با این حال، تعداد فرآیندهایی که ثابت هستند کم است و فهم و درک شیفت معنایی دارای اهمیت نخست برای مدیریت فرآیندها می‌باشد. بنابراین، تحقیقات و ابزارهای پشتیبانی بیشتری برای آنالیز مناسب شیفت معنایی مورد نیاز می‌باشد.

٤-٥- چالش پنجم: ارتقای Representational Bias ایی که برای کشف فرآیند استفاده می شوند

روش‌های کشف فرآیند با استفاده از یک زبان مشخص (مثلا BPMN یا شبکه پتری) یک مدل تولید می‌کنند. به هر حال، ذکر این نکته مهم است که باید بین شیوه نمایش نتایج با شیوه نمایشی که در حین کشف فرآیند استفاده می‌شود، تمایز قائل شد. انتخاب زبان هدف غالبا شامل چندین فرض ضمنی است. فرآیندهایی که نمی‌توانند با استفاده از زبان هدف نشان داده شوند، نمی‌توانند کشف هم شوند و این امر فضای جستجو را محدودتر می‌کند. این موضوع که representational bias یا تعصب بازنمودی نامیده می‌شود و در فرآیند کشف استفاده می‌شود، باید یک انتخاب آگاهانه باشد و نباید فقط مبتنی بر شیوه نمایش گرافیکی ایی که مورد ترجیح است، انتخاب شود.

به عنوان مثال، را در نظر بگیرید. در این شکل، امکان پشتیبانی از همزمانی یا عدم امکان آن، می‌تواند بر تصویرسازی مدل کشف شده و همچنین کلاس مدل‌هایی که توسط الگوریتم در نظر گرفته شده، تاثیر داشته باشد. اگر representational bias اجازه همزمانی را ندهد (شکل ۶-الف امکان پذیر نیست) و اجازه وجود چندین فعالیت با یک برچسب یکسان را ندهد (شکل ۶-ج هم امکان پذیر نیست)، در این صورت فقط مدل‌های مشکل‌سازی نظیر مدلی که در شکل ۶-ب نشان داده شده امکان پذیر است. این مثال نشان می‌دهد که یک انتخاب دقیق و تصفیه شده از representational bias مورد نیاز است.

٤-٦- چالش ششم: ایجاد توازن بین معیارهای کیفی نظیر تناسب، سادگی، دقت و تعمیم

[1] - concept drift

غالبا نگاره‌های رویداد فاصله زیادی تا کامل بودن دارند، و غالبا تنها رفتارهای نمونه در دسترس می‌باشند. در مدل‌های فرایند معمولا امکان وجود تعداد نمایی یا حتی نامحدود (در حالتی که حلقه وجود داشته باشد) دنباله[1] وجود دارد. علاوه بر این، برخی از دنباله‌ها ممکن است احتمال بسیار کمتری نسب به سایر دنباله‌ها داشته باشند. بنابراین، این تصور که همه دنباله‌های ممکن در نگاره رویداد موجود هستند، فرض غیر واقعی است. برای نشان دادن این موضوع که دسترسی به نگاره‌های کامل غیر عملی است، فرآیندی را در نظر بگیرید که شامل ۱۰ فعالیت بوده که می‌توانند به صورت موازی اجرا شوند و همچنین یک فایل نگاره رویداد مرتبط که حاوی ۱۰،۰۰۰ حالت است. تعداد کل حالت‌های ممکن جایگذاری در یک مدل با ۱۰ فعالیت موازی برابر با ۱۰!=۳،۶۲۸،۸۰۰ می‌باشد. بنابراین، عملی نیست که تمامی این ۱۰! حالت در فایل نگاره رویداد حضور داشته باشد چرا که تعداد آنها (۱۰،۰۰۰) بسیار کمتر از کل حالت ها (۳،۶۲۸،۸۰۰) است. حتی اگر میلیون‌ها حالت هم در نگاره موجود باشد، بسیار بعید است که تمامی گونه‌های ممکن در آن موجود باشد. پیچیدگی مضاعف دیگری که وجود دارد این است که برخی از حالت‌ها بسیار کم رخدادتر از سایر حالت‌ها هستند. این حالت‌ها ممکن است به عنوان "نویز" در نظر گرفته شوند. ساخت یک مدل مناسب برای چنین رفتارهای نویزی ایی امکان پذیر نیست. مدل کشف شده باید انتزاعی از این رفتارها باشد و بهتر است که رفتارهای با فرکانس کم با بهره‌گیری از تکنیک‌های چک کردن انطباق بررسی شوند.

وجود نویز و ناکامل بودن داده‌ها، مساله کشف فرآیند را پر چالش کرده است. در حقیقت، چهار بعد کیفی رقابت وجود دارد: الف) تناسب، ب)سادگی، ج) دقت و د)تعمیم. یک مدل با تناسب خوب اجازه بازپخش بیشتر رفتارهای دیده شده در نگاره رویداد را می‌دهد. یک مدل استخراج شده دارای تناسب کامل است هرگاه تمامی دنباله‌های موجود در نگاره رویداد قابل باز تولید توسط مدل از ابتدا تا به انتها باشند. ساده‌ترین مدلی که بتواند رفتار دیده شده در نگاره رویداد را نشان دهد، بهترین مدل است. این اصل به عنوان قانون Occam's Razor شناخته شده است. تناسب و سادگی به تنهایی برای ارزیابی کیفیت مدل فرآیند کشف شده کافی نیست. به عنوان مثال، به راحتی می‌شود یک مدل شبکه پتری بسیار ساده تولید کرد (مدل گل[2]) که همه دنباله‌های موجود در نگاره رویداد را باز تولید کند (و البته هر نگاره رویداد دیگری که دارای مجموعه فعالیتهای مشابه است). به طریق مشابه، مطلوب نیست که مدلی داشته باشیم که فقط اجازه تولید دقیق رفتارهای نشان داده شده در نگاره رویداد را بدهد. به یاد بیاورید که نگاره رویداد تنها حاوی رفتارهای مثال است و از تعداد بسیار زیاد نگاره‌های قابل رخداد، ممکن است که هنوز بسیاری دیده نشده باشند. یک مدل دقیق است هرگاه اجازه تولید رفتارهای خیلی زیاد ندهد. به وضوح، "مدل گل" فاقد دقت است. یک مدل که دقیق نباشد، کمتر از حد مناسب[3] است. Underfitting مشکلی است که در آن، مدل، رفتارهای نمونه دیده شده در نگاره رویداد را بیش از حد تعمیم می‌دهد (به عنوان مثال، مدل اجازه رفتارهای بسیار متفاوت از رفتارهایی که در نگاره رویداد دیده شده است را می‌دهد). یک مدل باید تعمیم دهد و نباید رفتارها را فقط محدود به آنچه در نگاره است، کند. مدلی که تعمیم نمی‌دهد، overfitting است. بیش از حد مناسب بودن یا overfitting، مشکلی است که در آن یک مدل خیلی خاص تولید شود، در حالیکه واضح است که نگاره‌های رویداد تنها حاوی رفتارهای نمونه هستند (به عنوان مثال، مدل یک نگاره نمونه را توضیح می‌دهد، اما یک نگاره نمونه دیگر از همان فرآیند ممکن است که یک مدل فرآیندی کاملا متفاوت دیگر تولید نماید).

---

[1] - trace
[2] - flower model
[3] - underfitting

ایجاد توازن بین تناسب، سادگی، دقت و تعمیم پرچالش است. به همین دلیل، بیشتر روش‌های قدرتمند کشف فرآیند پارامترهای گوناگونی را ارائه می‌کنند. الگوریتم‌های بهبود یافته بیشتری برای ایجاد توازن بین ۴ بعد کیفی رقابتی مورد نیاز است. علاوه بر این، هر پارامتری که استفاده می‌شود باید قابل فهم توسط کاربر نهایی باشد.

٤-٧- چالش هفتم: کاوش بین سازمانی

فرآیند کاوی به طور سنتی در درون یک سازمان انجام می‌شود. به هر حال، هر چه تکنولوژی خدمت‌رسانی، یکپارچگی زنجیره تامین، و محاسبات ابری گسترده‌تر می‌شود، سناریوهای مختلفی بوجود می‌آید که نگاره‌های رویداد چندین سازمان مورد آنالیز قرار گیرد. در اصل، دو حالت برای کاوش فرآیند چند سازمانی وجود دارد.

در نوع اول، حالتی را در نظر بگیرید که در آن سازمان‌های مختلف باهم دیگر همکاری می‌کنند تا نمونه‌های فرآیند را اجرا کنند. در واقع می‌توان این فرآیندهای چند سازمانی را به صورت قطعات پازل در نظر گرفت، مثلا فرآیند کلی به چند قسمت شکسته می‌شود و بین سازمان‌ها توزیع می‌شود تا با همکاری هم فرآیند را تکمیل کنند. آنالیز نگاره‌های رویداد یکی از این سازمان‌های درگیر، کافی نیست. برای کشف فرآیند انتها به انتها، نگاره‌های رویداد سازمان‌های مختلف باید باهم ادغام شوند. این کار ساده نیست چراکه رویدادها باید مرتبط با سراسر مرزهای سازمانی باشند.

در مدل دوم، حالتی را در نظر بگیرید که سازمان‌های مختلف اساسا فرآیند یکسانی را در حالیکه از دانش، تجربیات یا زیر ساخت مشترکی بهره می‌برند، اجرا می‌کنند. به عنوان مثال سایت Salesforce.com را در نظر بگیرید. فرآیند فروش بسیاری از سازمان‌ها توسط Salesforce مدیریت و پشتیبانی می‌شود. از یک طرف، این سازمان‌ها از یک زیر ساخت استفاده می‌کنند (فرآیندها، پایگاه داده‌ها، و ...). از طرف دیگر، آنها مجبور نیستند که از یک مدل فرآیند دقیق پیروی کنند، چراکه سیستم می‌تواند پیکربندی شود تا گونه‌های مختلفی از یک فرآیند تولید شود. به عنوان مثالی دیگر، فرآیندهای پایه‌ای (نظیر مجوز صدور ساخت) که در هر شهرداری اجرا می‌شود را در نظر بگیرید. اگر چه که همه شهرداری‌های یک کشور باید از یک مجموعه فرآیندهای پایه‌ای مشابه پشتیبانی کنند، ولی بازهم ممکن است تفاوت‌هایی وجود داشته باشد. پرواضح است که آنالیز این گونه‌های موجود در بین سازمان‌های مختلف جذاب است. این سازمان‌ها می‌توانند از همدیگر یاد بگیرند و تامین‌کنندگان سرویس می‌توانند خدماتشان را ارتقا داده و بر اساس نتایج بدست آمده از فرآیند کاوی بین سازمانی، خدمات ارزش افزوده پیشنهاد نمایند.

روش‌های تجزیه و تحلیل جدیدی برای توسعه هر دو نوع فرآیند کاوی بین سازمانی مورد نیاز است. این تکنیک‌ها باید بحث حریم شخصی و مسائل امنیتی را نیز در نظر بگیرند. سازمان‌ها ممکن است که به دلایل رقابتی و یا عدم اطمینان تمایلی به اشتراک‌گذاری اطلاعاتشان نداشته باشند. بنابراین ضروریست که تکنیک‌های فرآیند کاوی که قادر به حفظ حریم شخصی باشند، توسعه داده شود.

٤-٨- چالش هشتم: ارائه پشتیبانی عملیاتی[1]

[1] - Providing Operational Support

در ابتدا، تمرکز فرایند کاوی بر آنالیز داده‌های تاریخچه‌ای متمرکز بود. امروزه، به هرحال، بسیاری از منابع داده تقریبا به صورت بلادرنگ بروز می‌شوند و قدرت محاسباتی کافی برای انالیز رویدادها در زمانیکه رخ می‌دهند، وجود دارد. از اینرو، فرایند کاوی نباید محدود به آنالیزهای برون–خط باشد، بلکه باید قابل استفاده برای عملیات پشتیبانی روی خط نیز باشد. سه نوع عملیات پشتیبانی موجود است: شناسایی، پیش بینی، و توصیه. لحظه‌ای که یک نمونه از فرآیند از پیش تعریف شده منحرف شود، قابل شناسایی است و سیستم تواند یک هشدار اعلام نماید. غالبا در بسیاری موارد نیاز است که این هشدار بصورت فوری و نه به صورت برون–خط تولید شود. داده‌های تاریخچه‌ای برای ساخت مدل پیش‌بینی‌کننده می‌تواند بکار رود و از آن برای هدایت نمونه فرآیندهای در حال اجرا نیز می‌توان استفاده کرد. به عنوان مثال، می‌توان زمان پردازشی باقی مانده اجرای یک نمونه را پیش بینی نمود. بر اساس این پیش بینی‌ها می‌شود یک سیستم توصیه‌گر تولید کرد تا پیشنهادات خاصی برای کاهش هزینه‌ها یا کوتاه کردن زمان اجرا ارائه نماید. استفاده از تکنیک‌های فرایند کاوی در محیط روی خط، چالش‌های مضاعفی از نظر قدرت محاسباتی و کیفیت داده در پی دارد.

٤-٩- چالش نهم: ترکیب فرایند کاوی با سایر انواع آنالیز

مدیریت عملیات و به صورت ویژه تحقیق در عملیات یک شاخه از علم مدیریت است که به شدت متکی بر مدل سازی است. در این حوزه، انواع مدل‌های ریاضی اعم از برنامه‌ریزی خطی و برنامه‌ریزی پروژه تا مدل‌های صف‌بندی، زنجیره‌های مارکو و شبیه‌سازی استفاده می‌شود. یک تعریف از داده کاوی "تجزیه و تحلیل مجموعه‌های داده (غالبا بزرگ) برای پیدا کردن روابط ناپیدا و برای خلاصه‌کردن داده‌ها در یک شیوه جدید که قابل فهم و مفید برای مالک داده باشد" است. طیف گسترده‌ای از تکنیک‌ها توسعه داده شده است: طبقه‌بندی[1] (نظیر یادگیری درخت تصمیم[2])، رگرسیون[3]، خوشه‌بندی[4] (نظیر خوشه بندی k-means) و کشف الگو (نظیر یادگیری قوانین انجمنی[5]).

هر دو زمینه (مدیریت عملیات و داده کاوی) تکنیک‌های تجزیه و تحلیل ارزشمندی را فراهم آورده‌اند. چالشی که وجود دارد این است که چطور این تکنیک‌ها را در حوزه فرآیند کاوی استفاده کنیم. به عنوان مثال شبیه‌سازی را در نظر بگیرید. تکنیک‌های فرآیند کاوی می‌تواند برای یادگیری یک مدل شبیه‌سازی بر اساس داده‌های تاریخچه‌ای استفاده شود. متعاقبا، مدل‌های شبیه‌سازی برای ارائه پشتیبانی عملیاتی می‌تواند بکار رود. از آنجایی که ارتباط تنگاتنگی بین نگاره رویداد و مدل وجود دارد، مدل برای بازتولید تاریخچه می‌تواند استفاده شود.

به طور مشابه، ترکیب فرایند کاوی با تجزیه تحلیل‌های تصویری[6] نیز مطلوب است. آنالیز بصری، به منظور درک بهتر مجموعه داده‌های پیچیده و بزرگ، تجزیه و تحلیل خودکار را با تصویرسازی تعاملی ترکیب می‌کند. آنالیزهای تصویری، از قابلیت‌های شگفت‌انگیز انسان برای دیدن الگوها در داده‌های غیر ساخت‌یافته استفاده می‌کنند. با ترکیب تکنیک‌های خودکار فرایند کاوی با آنالیزهای تصویری تعاملی، می‌توان بینش بیشتری از داده‌های رویداد استخراج نمود.

---

1 - classification
2 - decision tree learning
3 - regression
4 - clustering
5 - association rule learning
6 - visual analytics

٤-١٠- چالش دهم: بهبود قابلیت استفاده برای افراد غیر متخصص

یکی از اهداف فرایند کاوی ایجاد "مدل‌های فرآیند زنده" (نظیر مدل‌های فرایندی که به صورت روزانه استفاده می‌شوند) بجای فرآیندهای ایستایی است که در برخی آرشیوها بایگانی می‌شوند. داده‌های رویداد جدید می‌تواند برای کشف رفتارهای جدید استفاده شود. پیوند بین داده‌های رویداد و مدل‌های فرآیند اجازه می‌دهد تا نگاشتی بین وضعیت فعلی و فعالیت‌های اخیر با مدل‌های بروز ایجاد شود. از این رو، کاربران نهایی می‌توانند به صورت روزانه در تعامل باشند. این تعامل بسیار ارزشمند است و البته نیاز به واسط کاربری قابل حس دارد. در اینجا چالش اصلی این است که الگوریتم‌های پیشرفته فرآیند کاوی را در پشت یک واسط کاربری کاربرپسندی که پارامترها را به صورت خودکار تنظیم کند و انواع آنالیزهای مناسب را پیشنهاد نماید، پیاده‌سازی کنیم.

٤-١١- چالش یازدهم: بهبود قابلیت فهم برای افراد غیر متخصص

در فرایند کاوی، حتی اگر بتوانیم به نتایجی دست یابیم، لزوما به این معنا نیست که این نتایج مفید و قابل استفاده هستند. کاربران ممکن است که در درک نتایج خروجی مشکل داشته باشند و یا اینکه نتیجه گیری‌های نادرستی انجام دهند. برای جلوگیری از چنین مشکلاتی، باید نتایج با استفاده از یک شیوه ارائه مناسب (چالش ۵ را مشاهده کنید)، نمایش داده شوند. علاوه بر این، قابل اعتماد بودن نتایج نیز باید به صورت شفاف مشخص شود. ممکن است که داده‌های اندکی برای تصدیق کردن نتایج وجود داشته باشد. در حقیقت، تکنیک‌های کنونی کشف فرایند برای تناسب پایین و یا تناسب بیش از حد هشداری نمی‌دهند. آنها همیشه یک مدل را به عنوان خروجی نشان می‌دهند، حتی اگر واضح باشد که داده‌ها آنقدر کم هستند که نشود هیچ نتیجه‌ای را تصدیق کرد.

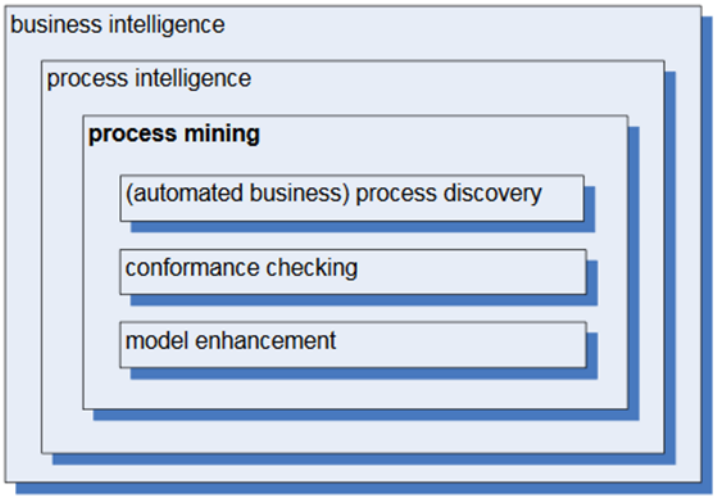
٥- خاتمه

در فرایند کاوی، IEEE Task Force قصد دارد که: الف) کاربردهای فرایند کاوی را توسعه دهد، ب) توسعه دهندگان نرم افزار، مشاورین، مدیران کسب و کار و کاربران نهایی را در زمان استفاده از تکنیک‌های سطح بالا هدایت کند، و ج) محرکی برای تحقیق در حوزه فرایند کاوی باشد. این بیانیه به بیان اصول اصلی و اهداف task force می‌پردازد. این بیانیه، پس از معرفی مبحث فرایند کاوی، فهرستی از اصول راهنما (۳- **اصول راهنما**

) و چالش‌ها (۴- **چالش‌ها**

) را معرفی کرده است. اصول راهنما در جلوگیری از بروز اشتباهات فاحش در مفید است. لیست چالش‌ها هم به منظور هدایت مسیر تحقیقات و تلاش‌های توسعه‌ای ارائه شده است. مقصود کلی از ارائه این دو لیست افزایش بلوغ فرایند کاوی است.

برای جمع‌بندی، چند کلمه و اصطلاح که در حوزه فرایند کاوی استفاده می‌شود آورده شده است: کاوش گردش‌کار (workflow mining)، فرایند (کسب و کار) کاوی، کشف خودکار فرایند (کسب و کار) و هوش فرایند (کسب و کار). به نظر می‌رسد که سازمان‌های مختلف از اصطلاحات متفاوتی برای مفاهیمی که باهم متداخلند، استفاده می‌کنند. به عنوان مثال،



شکل ۷– ارتباط بین اصطلاحات مختلف

گارتنر[1] اصطلاح "کشف خودکار فرایند کسب و کار" و گروه نرم افزاری AG اصطلاح "هوش فرایند" را برای اشاره به پلتفرم کنترلشان استفاده می‌کنند. به نظر می‌رسد که اصطلاح "کاوش گردش کار" خیلی مناسب نباشد، چراکه ایجاد مدل‌های گردش کار تنها یکی از بسیار کاربردهای ممکن فرایند کاوی است. به طریق مشابه اضافه کردن واژه "کسب و کار" حوزه بحث را به کاربردهای خاصی از فرایندکاوی محدود می‌کند. از آنجایی که تعداد بیشماری کاربرد فرایندکاوی موجود است (نظیر آنالیز سیستم‌های با تکنولوژی بالا یا تجزیه و تحلیل وب سایت‌ها)، بهمین دلیل اضافه کردن این واژه به نظر می‌رسد که مناسب نیست. اگر چه کشف فرایند بخش مهمی از حوزه فرایند کاوی است، اما تنها یکی از موارد بسیار زیاد کاربردها است. چک کردن انطباق، پیش بینی، کاوش سازمانی، تجزیه و تحلیل شبکه‌های اجتماعی و ... موارد استفاده‌ای هستند که آنسوی کشف فرایند توسعه داده می‌شوند.

شکل ۷ ارتباط برخی از اصطلاحات ذکر شده را نشان می‌دهد. کلیه تکنولوژی‌ها و روش‌هایی که اطلاعات کاربردی مناسب پشتیبانی تصمیم گیری تولید می‌کنند، را می‌توان در زیر چتر هوش تجاری (یا هوش کسب و کار) قرار داد. هوش فرایند (کسب و کار) ترکیب هوش تجاری و فرایند کاوی است، نظیر تکنیک‌هایی از هوش تجاری که برای تجزیه و تحلیل و بهبود فرایندها و مدیریتشان استفاده می‌شود. فرایند کاوی را می‌توان به عنوان یک نسخه دقیقتر شده هوش فرایندی در نظر گرفت که از نگاره‌های رویداد به عنوان نقطه آغازین استفاده می‌کند. کشف (خودکار) فرایند (کسب و کار) تنها یکی از سه شاخه اصلی فرایند کاوی است. ***شکل ۷*** ممکن است کمی گمراه کننده باشد، زیرا اکثر ابزارهای هوش تجاری عملکردهای فرایند کاوی شرح داده شده در این سند را ارائه نمی‌کنند. اصطلاح هوش تجاری اغلب به سمت دسته‌ای از ابزارهای مشخص یا روش‌هایی که تنها بخش کوچکی از طیف گسترده تر هوش تجاری است، متمایل شده است.

ممکن است که دلایل تجاری هم برای استفاده از اصطلاحات جایگزین وجود داشته باشد. برخی از فروشندگان تمایل دارند که بر یک جنبه خاص تاکید کنند (مثل، کشف یا هوش). به هرحال، برای جلوگیری از سردرگمی، بهتر است که از اصطلاح "فرایند کاوی" برای مفاهیم پوشش داده شده توسط این مانیفست استفاده شود.

[1] - Gartner

دسترسی آزاد. این فصل تحت مجوز لیسانس Creative Commons Attribution Noncommercial توزیع شده است که اجازه هر گونه استفاده، توزیع و تولید مجدد در هر رسانه غیر تجاری را که ارجاع به نویسندگان اصلی و منبع داده شده باشد، می‌دهد.

مرجع

1. van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, Berlin (2011)

واژه نامه

— **فعالیت (Activity):** یک مرحله خوش-تعریف در فرایند است. رویدادها ممکن است به شروع، لغو شدن، پایان یافتن و ... یک فعالیت برای یک نمونه فرایند مشخص اشاره کنند.

— **کشف خودکار فرایند کسب و کار(Automated Business Process Discovery):** بخش کشف فرایند را ببینید.

— **هوش تجاری (Business Intelligence):** مجموعه گسترده‌ای از ابزارها و روش‌ها که از داده برای پشتیبانی عملیات تصمیم گیری استفاده می‌کنند.

— **هوش فرایند کسب وکار (Business Process Intelligence):** بخش هوش فرایند را ببینید.

— **مدیریت فرایند کسب و کار (Business Process Management):** مجموعه اصولی که ترکیبی از دانش تکنولوژی اطلاعات و دانش علم مدیریت بوده و این دانش‌ها را برای فرایندهای کسب و کار عملیاتی بکار می‌برد.

— **نمونه (case):** نمونه فرایند را مشاهده کنید.

— **شیفت معنایی (Concept Drift):** پدیده تغییر فرایندها در طول زمان را گویند. فرایندهای مشاهده شده ممکن است که به تدریج (یا ناگهانی) به دلیل تغییرات فصلی یا افزایش رقابت تغییر کنند و در نتیجه تجزیه و تحلیل آنها پیچیده‌تر خواهد شد.

— **چک کردن انطباق (Conformance Checking):** تجزیه و تحلیل اینکه آیا واقعیت آنطور که ثبت شده است منطبق بر مدل هست و بلعکس. هدف این است که انحرافات شناسایی و شدت آنها اندازه‌گیری شود. چک کردن انطباق یکی از سه شاخه اصلی فرایند کاوی است.

— **فرایند کاوی چند سازمانی متقابل (Cross-Organizational Process Mining):** به کاربردهای تکنیک‌های فرایند کاوی با استفاده از نگاره‌های رویدادی که نشات گرفته از چندین سازمان مختلف هستند، اتلاق می‌شود.

— **داده کاوی (Data Mining):** تجزیه و تحلیل مجموعه داده (غالبا بزرگ) به منظور پیدا کردن روابط غیر منتظره و خلاصه کردن داده‌ها به شیوه‌ای که بینش جدیدی ایجاد شود.

— **رویداد (Event):** یک عمل ثبت شده در نگاره نظیر شروع، تکمیل و یا لغو شدن یک فعالیت برای یک نمونه فرایند مشخص.

— **نگاره رویداد (Event Log):** مجموعه‌ای از رویدادها که به عنوان ورودی فرایند کاوی استفاده می‌شوند. نیازی نیست که رویدادها در فایل نگاره جداگانه‌ای ثبت شوند (مثلا ممکن است که رویدادها در جداول چندین پایگاه داده پراکنده شده باشند).

— **تناسب (Fitness):** یک معیار اندازه گیری برای تعیین اینکه یک مدل به چه میزان می‌تواند رفتارهای دیده شده در نگاره رویداد را بازتولید کند.

— **تعمیم (Generalization):** یک معیار اندازه گیری برای تعیین اینکه یک مدل به چه میزان اجازه می‌دهد که رفتارهای دیده نشده در نگاره رویداد تولید گردد، است. یک مدل که overfitting است به اندازه کافی نمی‌تواند تعمیم داشته باشد.

— **بهبود مدل (Model Enhancement):** یکی از سه شاخه اصلی فرایند کاوی است. یک مدل فرایند با استفاده از اطلاعات استخراج شده از تعدادی نگاره بهبود یا توسعه داده می‌شود. به عنوان مثال، گلوگاه‌ها با استفاده از بازپخش نگاره‌ها برروی مدل و با بهره‌گیری از برچسب‌های زمانی قابل شناسایی هستند.

— **MXML:** یک قالب مبتنی بر XML برای تبادل نگاره‌های رویداد است. XES به عنوان یک قالب جدید مستقل از ابزار فرایند کاوی جایگزین MXML شده است.

— **عملیات پشتیبانی (Operational Support):** تجزیه و تحلیل روی–خط داده‌های رویداد با هدف مانیتور کردن و تاثیر گذاشتن بر نمونه فرایندهای در حال اجرا است. سه نوع عملیات پشتیبانی قابل تعریف است: شناسایی (تولید یک هشدار در مواقعی که رفتار مشاهده شده از رفتار مدل شده منحرف شود)، پیش‌بینی (پیش‌بینی رفتار آینده با استفاده از رفتارهای گذشته، مثلا پیش‌بینی‌کردن زمان باقی مانده اجرای فرایندها)، و توصیه (پیشنهاد اقدام مناسب برای تحقق بخشیدن یک هدف مشخص مثل کم کردن هزینه‌ها).

— **دقت (Precision):** معیاری که تعیین می‌کند که آیا یک مدل مانع از بروز رفتارهایی که تفاوت بسیاری با رفتارهای دیده شده در نگاره رویداد دارند، می‌شود یا خیر. یک مدل با دقت پایین under fitting است.

— **کشف فرایند (Process Discovery):** یکی از سه شاخه اصلی فرایند کاوی است. یک مدل فرایند بر اساس یک نگاره رویداد بدست می‌آید (با استفاده از یادگیری). به عنوان مثال، الگوریتم آلفا قادر است که یک شبکه پتری را با شناسایی الگوهای فرایند در مجموعه‌ای از رویدادها کشف کند.

— **نمونه فرایند (Process Instance):** موجودیت بکار گرفته شده توسط فرایند که تجزیه و تحلیل می‌شود. رویدادها به نمونه‌های فرایند اشاره می‌کنند. مثال نمونه فرایند عبارتست از سفارشات مشتری، درخواست‌های بیمه، درخواست‌های وام و ....

— **هوش فرایند (Process Intelligence):** یک شاخه از هوش تجاری است که بر مدیریت فرایندهای کسب و کار تمرکز دارد.

— **فرایندکاوی (Process Mining):** تکنیک‌ها، ابزارها، و روش‌های کشف، مانیتورینگ و بهبود فرایندهای واقعی (و نه فرایندهای فرضی) با استخراج دانش از نگاره‌های رویداد موجود در سیستم‌های (اطلاعاتی) رایج امروزی است.

— **تعصب نمایشی (Representational Bias):** زبان مقصد انتخاب شده برای نمایش و ساخت نتایج فرایند کاوی.

— **سادگی (Simplicity):** معیاری برای عملیاتی کردن قانون Occam's Razor است، ساده‌ترین مدلی که بتواند رفتار مشاهده شده در نگاره را توضیح دهد، بهترین مدل است. به شیوه‌های مختلفی می‌توان سادگی را تعیین کرد؛ نظیر تعداد گره‌ها و یال‌های موجود در مدل.

— **XES:** یک استاندارد مبتنی بر XML برای نگاره‌های رویداد است. این استاندارد توسط IEEE Task Force on Process Mining به عنوان قالب پیش فرض تبادل نگاره‌های رویداد پذیرفته شده است (ر.ج. http://www.xes-standard.org).