

تحلیل نگاره رویداد با هدف فرآیندکاوی: چالش‌ها و راه‌کارها

شکوفه قالیبافان^۱، محسن کاهانی^۲، بهشید بهکمال^۳

^۱دانشجوی کارشناسی ارشد، دانشگاه فردوسی مشهد، sh.ghalibafan@stu.um.ac.ir

^۲استاد، دانشگاه فردوسی مشهد، kahani@um.ac.ir

^۳استادیار، دانشگاه فردوسی مشهد، behkamal@um.ac.ir

چکیده

فرآیندکاوی یک زمینه تحقیقاتی نسبتاً جدید است که بین هوش محاسباتی و داده‌کاوی در یک سو و مدل‌سازی فرآیند و تحلیل آن در سوی دیگر قرار دارد. هدف اصلی فرآیندکاوی، کشف مدل‌های فرآیند بر اساس نگاره رویداد می‌باشد. نگاره رویداد، نقطه شروع فرآیندکاوی است؛ بنابراین کیفیت نگاره رویداد تاثیر مستقیم بر خروجی تکنیک‌های مختلف فرآیندکاوی خواهد گذاشت. در واقع، تکنیک‌های مختلف فرآیندکاوی در صورتی که نگاره رویداد بی کیفیت را به عنوان ورودی دریافت کنند، نتایج مناسبی را تولید نخواهند کرد. بنابراین مدیریت مسائل و چالش‌های تحلیل نگاره رویداد اهمیت بسیار بالایی برای موفقیت فرآیندکاوی دارد. با توجه به مانیفست فرآیندکاوی، برای نگاره رویداد ۵ سطح کیفی از یک ستاره تا پنج ستاره تعریف می‌گردد. هدف این پژوهش، بررسی چالش‌ها و مشکلات موجود جهت تحلیل نگاره رویداد و بررسی کارها و روش‌های انجام شده در راستای دستیابی به نتایج بهتر برای فرآیندکاوی می‌باشد. در این پژوهش مسائل متفاوتی که در تحلیل نگاره رویداد رخ می‌دهد و روش‌هایی که در راستای پرداختن به آن‌ها انجام شده است به صورت طبقه‌بندی شده ارائه می‌گردد. در واقع، کارهای انجام شده در زمینه بهبود کیفیت نگاره رویداد و بهبود فرآیندکاوی بر اساس نوع روش دسته‌بندی شده است، همچنین مسئله یا مسائلی که روش موردنظر به آن پرداخته است، بر اساس دسته‌بندی موجود، استخراج شده است. در پایان ارزیابی مبتنی بر ویژگی برای آن‌ها انجام می‌شود.

واژه‌های کلیدی

فرآیندکاوی، نگاره رویداد، کیفیت.

۱. مقدمه

فرآیندکاوی یک زمینه تحقیقاتی نسبتاً جدید است که همانند پلی بین داده‌کاوی، هوش تجاری و مدیریت فرآیندهای کسب‌وکار قرار دارد. فرآیندکاوی، تکنیکی برای بازسازی فرآیندهای کسب‌وکار از رویدادهای ثبت شده در سیستم‌های اطلاعاتی است. هدف اصلی فرآیندکاوی، کشف مدل‌های فرآیند بر اساس نگاره رویداد می‌باشد. در واقع نگاره رویداد نقطه شروع فرآیندکاوی است. انواع مختلف فرآیندکاوی شامل اکتشاف، تطبیق و بهبود است [۱]. همان‌طور که گفته شد، نگاره رویداد به عنوان ورودی هر تکنیک فرآیندکاوی استفاده می‌شود. بنابراین برای به دست آوردن نتایج بهتر در فرآیندکاوی نیاز به نگاره رویداد با کیفیت می‌باشد.

جدول ۱، نمونه‌ای از یک نگاره رویداد [۲]

case id	timestamp	activity	resource	cost
654423	30-04-2014:11.02	Register request	John	300
654423	30-04-2014:11.06	Check completeness of documents	Ann	400
654426	30-04-2014:16.10	Register request	John	200
654426	30-04-2014:16.14	Make appointment	Ann	450



۲. بیان مسئله و ضرورت موضوع

654423	30-04-2014:11.12	Ask for second opinion	Pete	100
654423	30-04-2014:11.18	Prepare decision	Pete	450

کیفیت نگاره رویداد در موفقیت تکنیک‌های فرآیندکاوی نقش کلیدی را ایفا می‌کند. به طور کلی، شناخت مسائل و چالش‌های نگاره رویداد و بررسی روش‌های موجود در پرداختن به این مسائل، سهم به سزایی در تولید نتایج مناسب برای تکنیک‌های مختلف فرآیندکاوی دارد.

تکنیک‌های مختلف فرآیندکاوی در صورتی که نگاره رویداد بی‌کیفیت را به عنوان ورودی دریافت کنند، نتایج مناسبی را تولید نخواهند کرد؛ مثلاً ممکن است در حوزه اکتشاف، مدل‌های شبه اسپاگتی تولید شود که تفسیر آن‌ها نیز دشوار است. مورد دیگری که در ارتباط با این مسائل ممکن است رخ دهد، نتایج پیچیده یا غیر قابل اعتماد است که در ارتباط با مواردی از جمله خصیصه مورد از دست‌رفته اتفاق می‌افتد. مشکل کشف جریان کنترلی صحیح توسط الگوریتم‌های فرآیندکاوی در ارتباط با مشکل موقعیت از دست‌رفته هم جزء این دسته از مسائل می‌باشد [۳].

بنابراین یکی از مهم‌ترین مسائل در حوزه فرآیندکاوی، مدیریت مسائل و چالش‌های تحلیل نگاره رویداد است که برای موفقیت فرآیندکاوی دارای اهمیت بالایی می‌باشد. به طور کلی، پیش‌پردازش نگاره رویداد به عنوان یکی از مسائل کیفیت داده مطرح می‌گردد [۳].

در این پژوهش، با بیان مسائل، چالش‌ها و مشکلات نگاره رویداد، یک طبقه‌بندی برای روش‌های موجود در جهت رفع این مشکلات ارائه کرده و در ادامه بیان خواهد شد که هر روش به حل چه مسأله‌ای می‌پردازد. در واقع، نوآوری اصلی این پژوهش، طبقه‌بندی و دسته‌بندی کارهای انجام‌شده در زمینه بهبود کیفیت نگاره رویداد و بهبود فرآیندکاوی و همچنین استخراج مسئله یا مسائلی است که طبق طبقه‌بندی موجود به آن پرداخته شده است که در بخش ۴ شرح داده شده است.

۳. کارهای مرتبط

مسائل و چالش‌های به وجود آمده در تحلیل نگاره رویداد در فرآیندکاوی را می‌توان به دو دسته کلی تقسیم کرد که در مقالات [۳] و [۴] به آن پرداخته شده است. این دو دسته کلی در شکل ۱ آمده است. ما این مسائل را با نام‌های C1 تا C8 نامگذاری کرده‌ایم تا در بخش بعدی و بیان روش‌های مقالات مختلف برای پرداختن به این مسائل، مشخص کنیم هر مقاله در جهت پرداختن به چه کلاسی از این مسائل گام برداشته است.

دسته اول از این مسائل، مربوط به مشخصات فرآیند می‌باشد. این دسته از مسائل در واقع مسائل ناشی از مشخصات مثل فعالیت‌های ریزدانه، تغییرپذیری یا ناهمگونی فرآیندها، تحول فرآیندها و حجم بالای فرآیندها می‌باشد. دسته دوم از این مسائل، مربوط به کیفیت نگاره رویداد می‌باشد.

در واقع، هرچه کیفیت نگاره رویداد بالاتر باشد، نتایج مناسب‌تری برای فرآیندکاوی، به دست خواهد آمد. در جدول ۱ نمونه‌ای از یک نگاره رویداد نشان داده شده است.

برای نگاره رویداد ۵ سطح کیفی از یک ستاره تا پنج ستاره تعریف می‌گردد [۱]. در پایین‌ترین سطح (یعنی یک ستاره)، نگاره رویداد، کیفیت ضعیفی دارد. بین رویدادهای ثبت شده در این نوع فایل، با واقعیت تناظری وجود ندارد. همچنین رویدادهای از دست‌رفته در این نوع فایل وجود دارد.

مثالی از فایل با پایین‌ترین سطح، نگاره رویدادی است که در آن، رویدادها به صورت دستی ثبت شده‌اند. نمونه‌ای از این نوع فایل، در جدول ۲ آمده است. مثلاً در این نوع فایل، شناسه مورد^۱ که یکی از مهم‌ترین موارد در نگاره رویداد است، ثبت نشده است. همچنین برچسب‌های زمانی به درستی ثبت نشده است. بعضی از آن‌ها فقط روز را نشان می‌دهند و برخی روز و ساعت را دارند. برخی موارد در این فایل از دست‌رفته‌اند مانند منبع و برخی فعالیت‌ها. فعالیتی مثل ارسال به مشتری اشتباه ثبت شده و چنین فعالیتی برای فرآیند وجود ندارد. همچنین منابع به درستی ثبت نشده‌اند. به طور کلی اشتباهات بسیاری در این نوع فایل به چشم می‌خورد و چنین فایل‌هایی قابل اعتماد نیست و استفاده از آن برای فرآیندکاوی منجر به نتایج غلط می‌شود.

جدول ۲، نمونه‌ای از نگاره رویداد با پایین‌ترین سطح

timestamp	activity	resource	reqitem
2016-02-03	Submit request	sara	25
2016.02.00	Check request		25
2016-05-02:10.06	Send to customer	elen	14
2016-02-04	Validate	sara	
2016-02-03:10.08		mike	14
2016-02-03	Deliver		14

در بالاترین سطح، نگاره رویداد دارای کیفیت عالی است که قابل اعتماد و خوش‌تعریف بوده، رویدادها و همه خصیصه‌های آن‌ها در آن به صورت اتوماتیک ثبت شده و معنای روشن دارند [۱].

در این پژوهش، ابتدا در بخش ۲، مسئله و ضرورت موضوع بیان خواهد شد. بخش ۳ شامل کارهای انجام‌شده در زمینه بیان مسائل و چالش‌های به وجود آمده در تحلیل نگاره رویداد در فرآیندکاوی می‌باشد. در بخش ۴، نوآوری اصلی کار بیان خواهد شد. در نهایت یک ارزیابی مبتنی بر ویژگی برای روش‌ها انجام می‌شود.

¹ Case Id

این دسته از مسائل در واقع مسائل مرتبط با کیفیت ثبت واقعه‌هاست که در نگاره رویداد ظاهر می‌شود.

در مسائل مربوط به دسته اول یعنی مشخصات فرآیند با ۴ دسته چالش روبه‌رو هستیم که شامل موارد زیر است [۳]:

۱- تعداد زیاد مورد رویداد: امروزه با رشد بی‌سابقه داده از طیف گسترده‌ای از منابع و سیستم‌ها روبه‌رو هستیم. برای مثال سیستم‌های با تکنولوژی بالا مثل سیستم‌های پزشکی، به دلیل این‌که معمولا رویدادهای خیلی سطح پایین را به عنوان رویدادهای اجرا شده توسط مولفه‌های سیستم دریافت می‌کنند، حجم زیادی داده تولید می‌کنند. هر رویداد در چنین محیط‌هایی عمر کمی دارد و صدها رویداد در مدت زمان کوتاهی می‌تواند انجام شود؛ یعنی متوسط تعداد رویدادها و تعداد موارد می‌تواند بسیار بالا باشد.

۲- تعداد زیاد ناهمگونی مورد از دنباله‌های مجزا: بسیاری از فرآیندها به صورت انعطاف‌پذیر طراحی می‌شوند. این نتایج در نگاره رویداد شامل ترکیب ناهمگونی از استفاده از سناریوها با رفتارهای متنوع و غیر ساخت‌یافته است (تعداد دنباله‌های مجزا بسیار بالاست). تکنیک‌های فرآیندکاوی در ارتباط با ناهمگونی در نگاره رویداد دچار مشکل می‌شوند مثلا الگوریتم‌های کشف فرآیند مدل‌های مشابه اسپاگتی تولید می‌کنند.

۳- تعداد زیاد دانه‌دانه بودن رویدادهای فعالیت‌های مجزا: فرآیندهایی که روی تعداد زیادی از فعالیت‌ها تعریف می‌شوند، ریزدانه هستند. تکنیک‌های فرآیندکاوی مشکلاتی را در ارتباط با رویدادهای ریزدانه دارند. برای مثال مدل‌های فرآیند کشف‌شده مشابه اسپاگتی بوده و برای فهم مشکل هستند. در واقع، یکی از چالش‌های اصلی در فرآیندکاوی، پر کردن شکاف بین نمایش مفهومی سطح بالای فرآیند و فایل‌های رویداد سطح پایین است. ۴- انعطاف‌پذیری فرآیند و تغییر مفهوم: اغلب فرآیندهای تجاری در محیط‌های پویا که در ارتباط با بازه وسیعی از متغیرها هستند، اجرا می‌شوند. تحلیل چنین تغییراتی جزء مسائل مهم در رسیدن به فهم دقیق از اجرای فرآیندها در هر لحظه از زمان است.

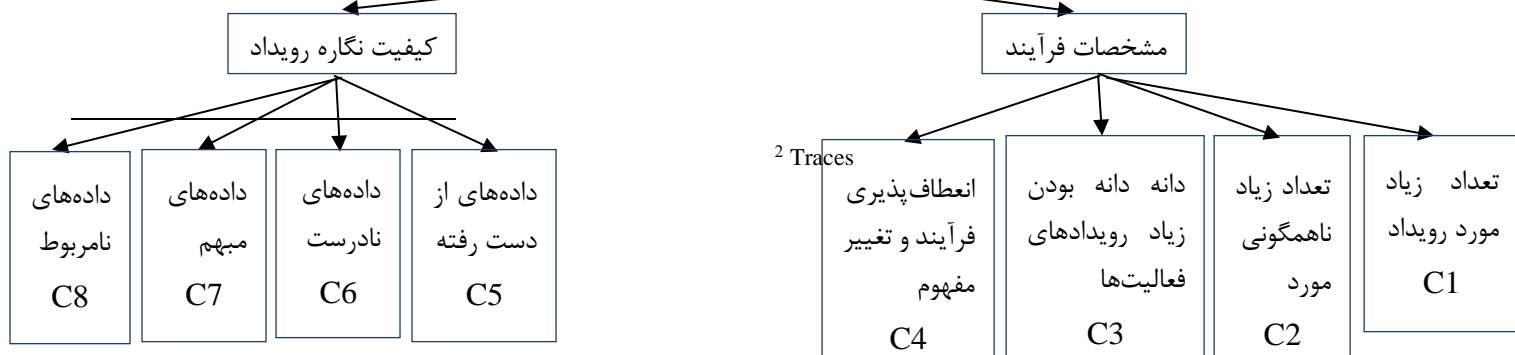
جدول ۳، مسائل مربوط به کیفیت نگاره رویداد [۳]

مورد	رویداد	رابطه	خصیصه مورد	مؤقت	نام فعالیت	برچسب زمانی	منبع	خصیصه رویداد
I1	I2	I3	I4	I5	I6	I7	I8	I9
I10	I11	I12	I13	I14	I15	I16	I17	I18
		I19	I20	I21	I22	I23	I24	I25
I26	I27							

مسائل دسته دوم یعنی مسائل مربوط به کیفیت نگاره رویداد نیز شامل ۴ دسته مسئله به صورت زیر است [۳] [۴]:

- ۱- داده‌های از دست‌رفته: وقتی انواع مختلفی از اطلاعات در نگاره رویداد از دست‌رفته باشد؛ مثلا موجودیت معینی از نگاره رویداد مثل رویداد، خصیصه رویداد و رابطه از دست‌رفته باشد. داده‌های از دست‌رفته غالبا ایجاد مشکل می‌کنند.
- ۲- داده‌های نادرست: متناظر با وضعیتی است که اگر چه داده ممکن است در نگاره رویداد باشد اما ممکن است به صورت نادرست ثبت شده باشد؛ مثلا یک موجودیت، رابطه یا مقدار نادرست باشد.

مسائل و چالش‌های به وجود آمده در تحلیل نگاره رویداد در فرآیندکاوی



در این شکل ابتدا دسته‌های مختلف بیان شده و بعد از ذکر نام شماره مقاله در هر دسته، روش انجام کار بیان گشته است. سپس، مسئله یا مسائلی را که مقاله مورد نظر به آن پرداخته است و در بخش قبل نامگذاری شد برای مقالات مشخص شده است.

انواع روش‌هایی که در زمینه افزایش کیفیت نگاره رویداد و بهبود فرآیند کاوی ارائه شده است را می‌توان به دو دسته روش‌های معنایی و غیر معنایی تقسیم نمود. در دسته روش‌های معنایی که شامل مقالات [۵]، [۶]، [۷]، [۸] و [۹] می‌شود از مواردی مانند قیود فرد خبره و سپس پاک‌سازی نگاره رویداد به صورت معنایی، افزایش بار معنایی و روابط معنایی استفاده شده است. ایده اصلی پاک‌سازی معنایی بهبود کیفیت نگاره رویداد با بررسی موارد برای انطباق با قیود خاص دامنه است. مواردی که با توجه به این قیود نادرست باشند، از نگاره رویداد پاک‌سازی می‌گردند.

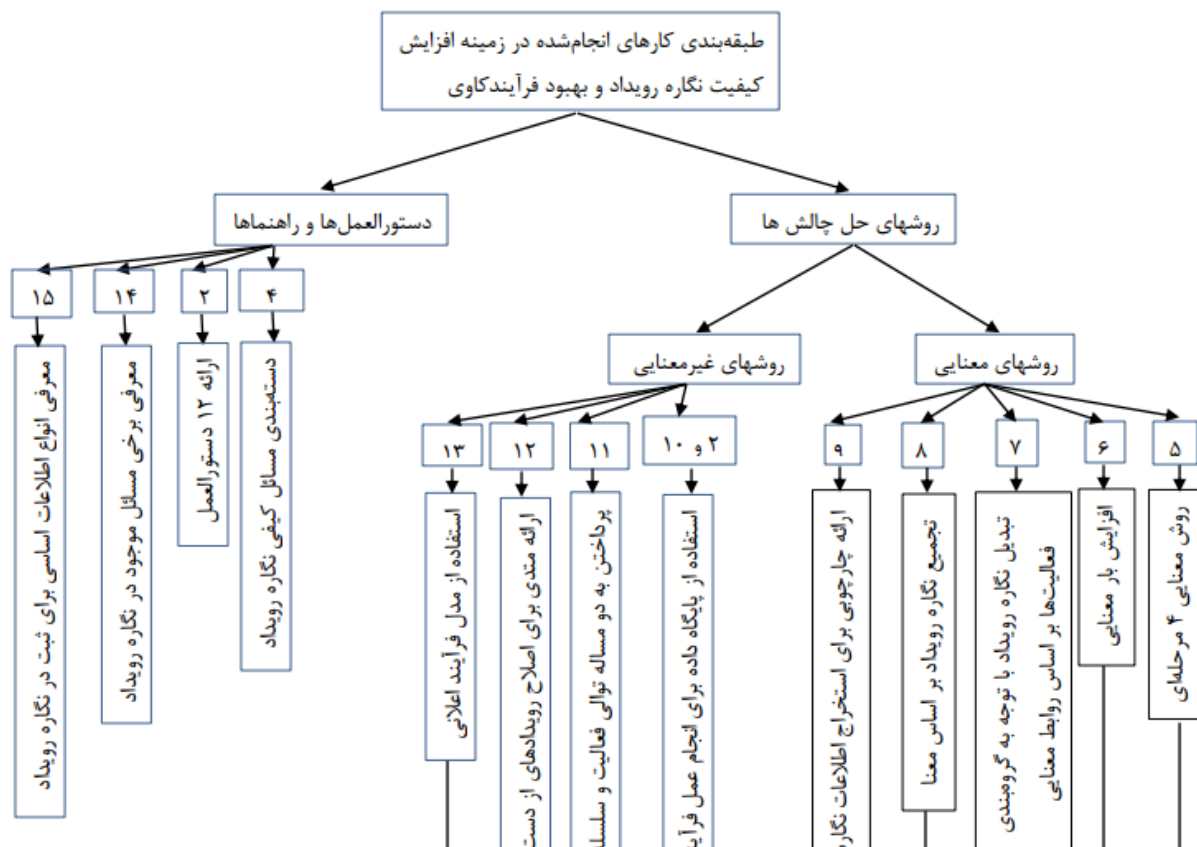
۳- داده‌های مبهم و غیردقیق: وقتی ورودی نگاره رویداد آن قدر درشت باشد که دقت پایین بیاید مثلاً برچسب زمانی می‌تواند خیلی درشت باشد (مثلاً روز)؛ بنابراین نتایج تحلیل قابل اطمینان نیست.

۴- داده‌های نامربوط: ورودی‌های نگاره رویداد ممکن است نامرتب باشد، البته داده‌های مرتبط می‌تواند از موجودیت‌ها به دست آید، اگرچه هنوز هم چالشی برای فرآیند کاوی است.

در مقالات [۳] و [۴] این ۴ دسته مسئله در ۲۷ کلاس دسته‌بندی شده است که در جدول ۳ نشان داده شده است.

۴. دسته‌بندی روش‌ها

در این بخش، کارهای انجام‌شده در زمینه بهبود کیفیت نگاره رویداد و بهبود فرآیند کاوی طبقه‌بندی شده است که در شکل ۲ آمده است.



در مقاله [۱۵] نیز مهم‌ترین اطلاعات شامل رویداد، برچسب زمانی، شناسه و فراداده مربوط به واسطی که رویداد را تولید کرده است، معرفی گشته است.

۵. ارزیابی

یکی از روش‌های ارزیابی موجود در برخی حوزه‌ها، ارزیابی مبتنی بر ویژگی است. برای مثال در مقالات [۱۶] و [۱۷] برای ارزیابی مدل کیفیت از معیارهای در دسترس استفاده شده است. در واقع، برخی معیارها برای ارزیابی مدل کیفیت، جمع‌آوری شده و سپس بر اساس آن‌ها، ارزیابی انجام شده است. همچنین در مقاله [۱۸]، یک روش ارزیابی مبتنی بر ویژگی برای تکنیک‌های باینری کردن به منظور بازیابی کلمه از تصاویر اسناد استفاده شده است. در واقع، برخی ویژگی‌های تصویر باینری شده استخراج شده و با یکدیگر مقایسه شده است.

در این بخش نیز، کارهای انجام شده در زمینه افزایش کیفیت نگاره رویداد و بهبود فرآیندکاوی در جدول ۴ بر اساس پارامترهایی نظیر نوع روش انجام کار، خودکار بودن روش و محدودیت روش ارزیابی شده‌اند.

جدول شماره ۴، ارزیابی مبتنی بر ویژگی کارهای انجام شده

مقاله	روش انجام کار	مشکلات و محدودیت‌ها	خودکار	نیمه خودکار
۵	روش معنایی ۴ مرحله‌ای	امکان حذف برخی موارد لازم		✓
۶	افزایش بار معنایی و اعمال آن به نگاره رویداد	عدم اصلاح مستقیم نگاره رویداد		✓
۷	تبدیل نگاره رویداد با توجه به گروه‌بندی فعالیت‌ها بر اساس روابط معنایی و انجام فرآیندکاوی بر روی نگاره رویداد مجرد	استفاده به منظور خاص و تنها برای کشف مدل‌های فرآیند اعلانی	✓	
۸	تجمیع نگاره رویداد بر اساس معنا	استفاده برای رویکردهای فرآیندکاوی مبتنی بر خوشه‌بندی	✓	
۹	ارائه چارچوبی برای استخراج اطلاعات نگاره رویداد	تنها استخراج یک سری اطلاعات معنایی و عدم اصلاح مشکلات		✓

شده در زمینه افزایش کیفیت نگاره رویداد و بهبود فرآیندکاوی [۶]، [۱۰]، [۱۱]، [۱۲] و

[۱۳] می‌باشد، وارد سطح معنایی نشده بلکه از مواردی مانند دانش مدل فرآیند استفاده می‌گردد؛ مثلاً در مقاله [۱۲] برای اصلاح رویدادهای از دست‌رفته در نگاره رویداد، ابتدا از احتمالات مسیر و سپس از شبکه بیزین استفاده می‌شود. ابتدا و در گام نخست، احتمالات مسیر استفاده می‌شود تا مشخص گردد رویدادهای از دست‌رفته محتمل‌تر چه چیزهایی هستند. بعد از آن با استفاده از شبکه بیزین، محتمل‌ترین برچسب زمانی برای هر ورودی ثبت می‌گردد. در این صورت مشکل رویداد و برچسب زمانی از دست‌رفته و یا نادرست حل می‌گردد.

دسته دیگری که در شکل ۲ ارائه شده مربوط به کارهایی است که راهنماها و دستورالعمل‌هایی را در زمینه واقعه‌نگاری، مسائل کیفی نگاره رویداد و مشکلات مرتبط ارائه داده‌اند که شامل مقالات [۲]، [۴]، [۱۴] و [۱۵] است. هم‌چنین در دسته‌بندی شکل ۲ مشخص شده است هر روش به چه مسئله‌ای از دسته مسائل موجود در تحلیل نگاره رویداد پرداخته است. هدف این راهنماها و دستورالعمل‌ها، ایجاد یک نقطه شروع خوب برای فرآیندکاوی است.

مسائل کیفی دسته‌بندی شده و در نهایت درجه‌بندی شده نگاره رویداد، معرفی برخی مسائل موجود در نگاره رویداد و چگونگی پرداختن به آن‌ها و معرفی انواع اطلاعات اساسی برای ثبت در نگاره رویداد از این دست مسائل است.

به عنوان مثال در مقاله [۱۴] منابع رایج خطا به صورت زیر معرفی شده است:

- رویدادهای از دست‌رفته
- داده‌های حذف شده
- معنای نابجا^۳

³ Misplaced Semantics

✓	احتمال در نظر نگرفتن برخی موارد با توجه به استفاده از محتمل ترین موارد	ارائه متدی برای اصلاح رویدادهای از دسترفته	۱۲	✓	استخراج مستقیم داده‌های مربوط به رویداد، محدود بودن رویدادها و در نظر نگرفتن برخی موارد	استفاده از پایگاه داده برای انجام عمل فرآیندکاوی	۲ و ۱۰
✓	ارائه متد تنها بر اساس قیود	استفاده از مدل فرآیند اعلانی	۱۳	✓	در نظر گرفتن برخی موارد خاص و در نظر نگرفتن مشکلات مهم‌تر	پرداختن به دو مساله توالی فعالیت و سلسله مراتب فعالیت	۱۱

[6] Azzini, Antonia, Chiara Braghin, Ernesto Damiani, and Francesco Zavatarelli. "Using Semantic Lifting for improving Process Mining: a Data Loss Prevention System case study." In *SIMPDA*, pp. 62-73. Springer International Publishing, 2013.

در این جدول، بعد از بیان شماره مقاله و روش انجام کار مقاله مورد نظر محدودیت‌ها یا مشکلات خاص آن روش استخراج شده و روش کار به صورت مورد نظر خودکار یا نیمه خودکار بودن ارزیابی شده است.

[7] Richetti, Pedro H. Piccoli, Fernanda Araujo Baião, and Flávia Maria Santoro. "Declarative process mining: reducing discovered models complexity by pre-processing event logs." In *Business Process Management*, pp. 400-407. Springer International Publishing, 2014.

[8] Deokar, Amit V., and Jie Tao. "Semantics-based event log aggregation for process mining and analytics." *Information Systems Frontiers* 17, no. 6, 1209-1226, 2015.

[9] Syamsiyah, Alifah. "Ontology-Driven Extraction of Event Logs from Relational Databases.", 2015.

[10] de Murillas, Eduardo González López, Wil MP van der Aalst, and Hajo A. Reijers. "Process Mining on Databases: Unearthing Historical Data from Redo Logs." In *Business Process Management*, pp. 367-385. Springer International Publishing, 2015.

[11] Dunkl, Reinhold. "Data improvement to enable process mining on integrated non-log data sources." In *Computer Aided Systems Theory-EUROCAST 2013*, pp. 491-498. Springer Berlin Heidelberg, 2013.

[12] Rogge-Solti, Andreas, Ronny S. Mans, Wil MP van der Aalst, and Mathias Weske. "Repairing event logs using timed process models." In *On the Move to Meaningful Internet Systems: OTM 2013 Workshops*, pp. 705-708. Springer Berlin Heidelberg, 2013.

[13] de Leoni, Massimiliano, Fabrizio M. Maggi, and Wil MP van der Aalst. "An alignment-based framework to check the conformance of declarative process models and to preprocess event-log data." *Information Systems* 47, 258-277, 2015.

[14] Dumais, Susan, Robin Jeffries, Daniel M. Russell, Diane Tang, and Jaime Teevan. "Understanding user behavior through log data and analysis." In *Ways of Knowing in HCI*, pp. 349-372. Springer New York, 2014.

[15] Alspaugh, S., Archana Ganapathi, Marti A. Hearst, and Randy Katz. "Better logging to improve interactive data analysis tools." In *KDD Workshop on Interactive Data Exploration and Analytics (IDEA)*, 2014.

[16] M. Morisio, I. Stamelos, A. Tsoukias, A new method to evaluate software artifacts against predefined profiles, in: Proceeding of SSEKE'02, ACM, Italy, 2002.

[17] G. Malak, L. Badri, M. Badri, H. Sahraoui, Towards a multidimensional model for web-based applications quality assessment, in: Computer Science Journal, Springer, 2004

[18] Sari, Toufik, Abderrahmane Kefali, and Halima Bahi. "Structural feature-based evaluation method of binarization

۶. نتیجه‌گیری و جمع‌بندی

با توجه به نیاز روزافزون به زمینه فرآیندکاوی در حوزه‌های مختلف، نیاز به پیش در کسب نتایج مناسب و باکیفیت در آن احساس می‌شود.

برای دستیابی به نتایج بهتر در فرآیندکاوی روش‌های مختلفی وجود دارد. یکی از راه‌های کسب نتیجه بهتر، افزایش کیفیت نگاره رویداد است که به عنوان ورودی فرآیندکاوی استفاده می‌گردد. در این پژوهش بعد از بیان طبقه‌بندی شده مسائل و مشکلات تحلیل نگاره رویداد، کارهای مرتبط در زمینه افزایش کیفیت نگاره رویداد و بهبود فرآیندکاوی به صورت دسته‌بندی شده ذکر شده، روش هر مقاله و مسئله‌ای که هر مقاله به آن می‌پردازد بررسی و عنوان گشته و در پایان ارزیابی مبتنی بر ویژگی بر روی کارهای انجام شده، در جدولی بیان شده است.

مراجع

[1] Van Der Aalst, Wil, Arya Adriansyah, Ana Karla Alves de Medeiros, Franco Arcieri, Thomas Baier, Tobias Blickle, Jagadeesh Chandra Bose et al. "Process mining manifesto." In *Business process management workshops*, pp. 169-194. Springer Berlin Heidelberg, 2011.

[2] van der Aalst, Wil MP. "Extracting event data from databases to unleash process mining." In *BPM-Driving innovation in a digital world*, pp. 105-128. Springer International Publishing, 2015.

[3] Bose, R. P., Ronny S. Mans, and Wil MP van der Aalst. "Wanna improve process mining results?." In *Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on*, pp. 127-134. IEEE, 2013.

[4] Mans, Ronny S., Wil MP van der Aalst, and Rob JB Vanwersch. "Data Quality Issues." In *Process Mining in Healthcare*, pp. 79-88. Springer International Publishing, 2015.

[5] Ly, Linh Thao, Conrad Indiono, Jürgen Mangler, and Stefanie Rinderle-Ma. "Data transformation and semantic log purging for process mining." In *Advanced Information Systems Engineering*, pp. 238-253. Springer Berlin Heidelberg, 2012.

techniques for word retrieval in the degraded Arabic document images." *International Journal on Document Analysis and Recognition (IJ DAR)* 19, no. 1, 2016.

[19] Sidi, Fatimah, P. H. Shariat Panahy, Lilly Suriani Affendey, Marzanah A. Jabar, Haidi Ibrahim, and Aouache Mustapha. "Data quality: A survey of data quality dimensions." In *Information Retrieval & Knowledge Management (CAMP), 2012 International Conference on*, pp. 300-304. IEEE, 2012.

[20] Fürber, Christian. "Data Quality." In *Data Quality Management with Semantic Technologies*, pp. 20-55. Springer Fachmedien Wiesbaden, 2016.

[21] Kulkarni, Prerna S., and J. W. Bakal. "Survey on Data Cleaning." *structure* 3, no. 4, 2014.

[22] Devi, Sapna, and Arvind Kalia. "Study of Data Cleaning & Comparison of Data Cleaning Tools." 2015.

[23] Laranjeiro, Nuno, Seyma Nur Soydemir, and Jorge Bernardino. "A Survey on Data Quality: Classifying Poor Data." In *Dependable Computing (PRDC), 2015 IEEE 21st Pacific Rim International Symposium on*, pp. 179-188. IEEE, 2015.

[24] ISO, *ISO/IEC 25012- Software engineering - Software product Quality Requirements and Evaluation (SQuaRE)*, in *Data quality model*. 2008.