

سید محمد



دانشکده کامپیوتر و فناوری اطلاعات

رشته مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیک

پایان نامه کارشناسی ارشد

بخش بندی فعالیت های فرآیندهای کاری بر اساس تحلیل آماری

نگارنده: پیمان دهقان

استاد راهنما:

دکتر مرتضی زاهدی

شهریور ۱۳۹۵



شماره:

تاریخ:

باسمه تعالی

ویرایش:

مدیریت تحصیلات تکمیلی

دانشکده :

گروه :

پایان نامه کارشناسی ارشد آقای / خانم به شماره دانشجویی:

تحت عنوان:

در تاریخ توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد
با درجه مورد پذیرش قرار گرفت.
مورد ارزیابی و

امضاء	اساتید مشاور	امضاء	اساتید راهنما
	نام و نام خانوادگی :		نام و نام خانوادگی :
	نام و نام خانوادگی :		نام و نام خانوادگی :

امضاء	نماینده تحصیلات تکمیلی	امضاء	اساتید داور
	نام و نام خانوادگی :		نام و نام خانوادگی :
			نام و نام خانوادگی :
			نام و نام خانوادگی :
			نام و نام خانوادگی :

سیاس گزاری

خدای را بسی شاکرم که از روی کرم، پدر و مادری فداکار نصیبم ساخته تا در سایه
درخت پر بار وجودشان بیسایم و از ریشه آن ها شاخ و برگ گیرم و از سایه وجودشان
در راه کسب علم و دانش تلاش نمایم. والدینی که بودندشان تاج افتخاری است بر
سرم و نامشان دلیلی است بر بودنم، چرا که این دو وجود، پس از پروردگار، مایه
هستی ام بوده اند و دستم را گرفتند و راه رفتن را در این وادی زندگی پر از فراز و نشیب
آموختند. آموزگارانمی که برایم زندگی، بودند و انسان بودن را معنا کردند.
تقدیم به وجود بارزیشان...

تشکر و قدردانی

از استاد گرامیم جناب آقای دکتر مرتضی زاهدی بسیار سپاسگزارم چرا که بدون راهنمایی‌های ایشان تأمین این پایان نامه بسیار مشکل می نمود.

و

از جناب آقای مهندس یعقوبی به دلیل یاری‌ها و راهنمایی‌های بی‌چشمداشت ایشان که بسیاری از سختی‌ها را برایم آسان تر نمودند.

تعهد نامه

اینجانب **پیمان دهقان** دانشجوی دوره کارشناسی ارشد رشته هوش مصنوعی دانشکده کامپیوتر و فناوری اطلاعات دانشگاه صنعتی شاهرود نویسنده پایان نامه بخش بندی فعالیت های یک فرآیند کاری بر اساس تحلیل آماری تحت راهنمایی آقای دکتر مرتضی زاهدی متعهد می شوم.

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود است و مقالات مستخرج با نام «دانشگاه صنعتی شاهرود» و یا «Shahrood University of Technology» به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است. تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزارها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود است. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

چکیده:

مدیریت فرآیند کسب و کار یک روش ساخت‌یافته و سیستماتیک برای تحلیل، توسعه، کنترل و مدیریت فرآیند با هدف توسعه کیفیت خدمات است. عدم وجود یک سیستم کسب و کار کارآمد باعث ایجاد خدمات کم کیفیت و هزینه‌های غیرضروری و همچنین نارضایتی برای کارمندان و مشتریان هر سازمان می‌شود. مدیریت فرآیند کسب و کار یک راه‌حل جامع برای درک و مدیریت روندهای موجود در هر سازمان است و به سازمان‌ها کمک می‌کند تا بتوانند روابط کاری پیچیده‌ی موجود بین کارکنان، سیستم‌ها و فرآیندها را به نحوی ساده‌سازی کنند که قابل‌فهم شود و از این راه به افزایش بهره‌وری سازمان‌ها و همچنین رضایت مشتریان کمک کند.

در این تحقیق مدیریت فرآیند کسب و کار، کاوش فرآیند و الگوریتم‌های کاوش فرآیند مورد بررسی قرار می‌گیرد و سپس چندین فرآیند به منظور کشف روابط بین آن‌ها کاوش می‌شود تا بتوان فعالیت‌های مشابه از یک فرآیند را که می‌توانند در یک گروه یا کلاس قرار گیرند را شناسایی کرد. در این مرحله ابتدا داده‌های گزارشی ثبت‌شده را بر اساس زمان شروع تراکنش‌ها و همچنین پایان تراکنش‌ها استخراج کرده و در مرحله‌ی بعد به خوشه‌بندی و یافتن روابط نهفته در این داده‌های گزارشی پرداخته می‌شود. نتایج حاصل از پیاده‌سازی روی فرآیندهای واقعی و شبیه‌سازی شده نشان‌دهنده‌ی این است که خوشه‌های تولید شده توسط روش پیشنهادی توانسته میانگین ۸۵,۵٪ شباهت را با خوشه‌های ایجاد شده توسط انسان در فرآیندهای شبیه‌سازی شده‌ی آموزشی و دقت ۷۳,۳٪ را در فرآیندهای واقعی حفظ کند. همچنین با نظر کارشناسان متخصص در این حوزه روش پیشنهادی تقریباً ۱۰٪ دقت بیشتری نسبت به روش فیلتر کردن گره در کاوش فازی دارد ضمن این که زمان پردازش روش پیشنهادی در مقایسه با روش خوشه‌بندی ردپاهای پرونده بسیار بهبود پیدا کرده است.

کلمات کلیدی: مدیریت فرآیند کسب‌وکار، کاوش فرآیند، الگوریتم‌های کاوش فرآیند، داده‌های گزارشی، خوشه‌بندی

فهرست مطالب

فصل اول: مقدمه.....	۱
۱-۱ تاریخچه.....	۲
۲-۱ فرآیند.....	۳
۳-۱ کاوش فرآیند.....	۵
۴-۱ ساختار پایان نامه.....	۶
فصل دوم: مروری بر کارهای انجام شده.....	۷
۱-۲ مرور ادبی.....	۸
۲-۲ الگوریتم‌های معروف در کاوش فرآیند.....	۱۲
۳-۲ رویکردهای ساده‌سازی موجود در کاوش فرآیند.....	۱۴
فصل سوم: سیستم‌های مدیریت فرآیند کسب و کار و کاوش فرآیندها.....	۱۷
۱-۳ سیستم‌های مدیریت فرآیند کسب و کار (BPMS) چیست؟.....	۱۸
۲-۳ چرخه حیات فرآیندهای کسب و کار.....	۲۰
۳-۳ کاربرد مدیریت فرآیندها.....	۲۳
۱-۳-۳ کاربرد BPMS در بیمه: توسعه کسب و کار.....	۲۴
۲-۳-۳ کاربرد برای دولت: پیگیری مکاتبات.....	۲۴
۳-۳-۳ کاربرد برای تولید؛ دسترسی به امکانات.....	۲۵
۴-۳ مزایای مجهز شدن یک سازمان به BPMS.....	۲۶
۵-۳ مدل‌سازی سیستم‌های مدیریت جریان کاری.....	۲۷
۱-۵-۳ زنجیره‌ی فرآیند مبتنی بر رخداد.....	۲۸
۲-۵-۳ UML.....	۲۸
۳-۵-۳ دیاگرام فرآیندهای تجاری.....	۲۸
۴-۵-۳ شبکه‌های پتری.....	۲۹
۶-۳ کاوش فرآیند.....	۳۰
۷-۳ مقایسه فرآیند کاوی با داده کاوی.....	۳۰
۸-۳ انواع روش‌های کاوش فرآیند.....	۳۱
۹-۳ یک مثال ساده از کاوش فرآیند در سیستم آموزش.....	۳۲
۱۰-۳ چالش‌های فرآیند کاوی.....	۳۳

۳۳	۱-۱۰-۳ چالش اول: یافتن، ادغام کردن و تمیز کردن سابقه‌ی رخدادها.....
۳۴	۲-۱۰-۳ چالش دوم: استفاده از سابقه‌ی رخدادهای پیچیده که ویژگی‌های گوناگونی دارند.....
۳۵	۳-۱۰-۳ چالش سوم: ایجاد شاخصه‌های ارزیابی.....
۳۶	۴-۱۰-۳ چالش چهارم: مواجهه با رانش مفهومی.....
۳۷	۵-۱۰-۳ چالش پنجم: ارتقای پیش‌فرض‌های نمایشی که در کشف فرآیند استفاده می‌شوند.....
۳۷	۶-۱۰-۳ چالش ششم: برقراری تعادل بین معیارهای کیفیت نظیر سازگاری، سادگی، دقت و عمومیت.....
۳۸	۷-۱۰-۳ چالش هفتم: کاوش بین سازمانی.....
۳۹	۸-۱۰-۳ چالش هشتم: ارائه پشتیبانی عملیاتی.....
۳۹	۹-۱۰-۳ چالش نهم: ترکیب کاوش فرآیند با سایر روش‌های آنالیز.....
۴۰	۱۱-۱۱-۳ خوشه‌بندی در کاوش فرآیندها.....
۴۰	۱۲-۱۲-۳ معیارهای فاصله.....
۴۱	۱-۱۲-۳ فاصله‌ی اقلیدسی.....
۴۱	۲-۱۲-۳ فاصله‌ی همینگ.....
۴۱	۳-۱۲-۳ فاصله‌ی جاکارد.....
۴۲	۱۳-۱۳-۳ الگوریتم‌های خوشه‌بندی.....
۴۲	۱-۱۳-۳ کا-میانگین.....
۴۳	۲-۱۳-۳ آستانه‌ی کیفیت.....
۴۵	۱۴-۱۴-۳ مقایسه هزینه‌ی زمانی دو الگوریتم آستانه کیفیت و کا-میانگین.....
۴۷	فصل چهارم: شرح مسئله و راه حل پیشنهادی.....
۴۸	۱-۴ مفاهیم و اصطلاحات.....
۴۸	۱-۱-۴ فرآیند.....
۴۹	۲-۱-۴ سابقه‌ی رخدادها.....
۴۹	۳-۱-۴ پرونده.....
۵۰	۴-۱-۴ فعالیت.....
۵۰	۲-۴ نمونه‌ای از سابقه‌ی رخداد یک فرآیند.....
۵۱	۳-۴ معرفی سابقه‌ی رخدادهای مورد استفاده در این تحقیق.....
۵۴	۴-۴ روش انجام کار.....
۵۴	۵-۴ شرح کار انجام شده.....
۵۹	فصل پنجم: پیاده سازی و نتایج.....

۵-۱	روند پیاده‌سازی روی نما.....	۶۰
۵-۲	ارزیابی خوشه‌بندی فرآیندهای شبیه‌سازی شده.....	۶۶
۵-۳	پیاده‌سازی برای فرآیندهای واقعی.....	۶۶
۵-۴	ارزیابی روش ارائه‌شده روی فرآیندهای واقعی.....	۷۶
۵-۵	نتایج به‌دست‌آمده روی الگوریتم فازی.....	۷۷
۵-۶	تفاوت و مقایسه‌ی روش پیشنهادی با روش کاوش فازی.....	۸۴
۵-۷	تفاوت و مقایسه‌ی روش پیشنهادی با روش خوشه‌بندی ردپاها.....	۸۷
	فصل ششم: نتیجه‌گیری و پیشنهادها.....	۸۹
۶-۱	نتیجه‌گیری.....	۹۰
۶-۲	پیشنهاد.....	۹۱
	منابع:.....	۹۲
	پیوست:.....	۹۵
	سیستم نما:.....	۹۵
	پرام:.....	۹۵

فهرست اشکال

- شکل ۱-۱: شکل شماتیک فرآیند..... ۴
- شکل ۱-۳: چرخه‌ی حیات مدیریت فرآیند کسب و کار..... ۲۳
- شکل ۱-۴: فلوچارت روش پیشنهادی..... ۵۷
- شکل ۱-۵: فرآیند فعالیت کلاسی در سیستم نما..... ۶۰
- شکل ۲-۵: نتیجه خوشه‌بندی فرآیند اطلاعات پروپوزال و پایان نامه..... ۶۴
- شکل ۳-۵: خوشه‌بندی فرآیند فعالیت کلاسی..... ۶۵
- شکل ۴-۵: نتیجه خوشه‌بندی فرآیند اخذ پروژه دانشجوی کارشناسی..... ۶۵
- شکل ۵-۵: خروجی الگوریتم آلفا مربوط به فرآیند تعمیر لوازم الکترونیک..... ۶۷
- شکل ۶-۵: خوشه‌بندی برای فرآیند تعمیر لوازم الکترونیک..... ۶۸
- شکل ۷-۵: نتیجه خوشه‌بندی فرآیند جریمه‌های ترافیکی..... ۷۱
- شکل ۸-۵: خروجی الگوریتم آلفا از لاگ Bpi Challenge..... ۷۲
- شکل ۹-۵: اولین زیر فرآیند تشکیل شده توسط داده‌های خوشه‌بندی شده..... ۷۳
- شکل ۱۰-۵: دومین زیر فرآیند تشکیل شده توسط داده‌های خوشه‌بندی شده..... ۷۴
- شکل ۱۱-۵: سومین زیر فرآیند تشکیل شده توسط داده‌های خوشه‌بندی شده..... ۷۴
- شکل ۱۲-۵: چهارمین زیر فرآیند تشکیل شده توسط داده‌های خوشه‌بندی شده..... ۷۵
- شکل ۱۳-۵: زیر فرآیند اول، نتیجه حاصل از خوشه‌بندی مجدد شکل (۵-۹)..... ۷۶
- شکل ۱۴-۵: زیر فرآیند دوم، نتیجه حاصل از خوشه‌بندی مجدد شکل (۵-۹)..... ۷۶
- شکل ۱۵-۵: مدل ایجاد شده توسط کاوش فازی روی فرآیند تعمیر لوازم الکترونیک..... ۷۹
- شکل ۱۶-۵: مدل به دست آمده از کاوش فازی پس از اعمال فیلتر کردن گره روی فرآیند تعمیر لوازم الکترونیک... ۸۰
- شکل ۱۷-۵: نتیجه حاصل از روش پیشنهادی روی مدل فازی برای فرآیند تعمیر لوازم الکترونیک..... ۸۱
- شکل ۱۸-۵: نتیجه فیلتر کردن گره روی فرآیند جریمه‌های ترافیکی..... ۸۲
- شکل ۱۹-۵: نتیجه حاصل از روش پیشنهادی روی فرآیند جریمه‌های ترافیکی..... ۸۳

فهرست جداول

جدول ۴-۱: نمونه‌ای از سابقه‌ی رخداد یک فرآیند.....	۵۰
جدول ۴-۲: لاگ مربوط به فرآیند Bpi Challenge.....	۵۲
جدول ۴-۳: قسمتی از لاگ فرآیند تعمیر لوازم الکترونیکی.....	۵۳
جدول ۵-۱: سابقه‌ی رخداد ایجادشده به وسیله حرکت پرونده در فرآیند.....	۶۲
جدول ۵-۲: امتیازهای کارشناسان به فرآیندهای آموزشی.....	۶۶
جدول ۵-۳: شباهت فرآیندهای خوشه‌بندی شده با خوشه‌های واقعی از نظر کارشناسان.....	۶۶
جدول ۵-۴: خروجی داده‌ای برنامه در فاز اول.....	۶۹
جدول ۵-۵: نتیجه‌ی نهایی خوشه‌بندی.....	۷۰
جدول ۵-۶: ارزیابی خوشه‌بندی روی فرآیندهای واقعی توسط کارشناسان.....	۷۷
جدول ۵-۷: درصد قابل قبول بودن روش پیشنهادی در خوشه‌بندی.....	۷۷
جدول ۵-۸: مقایسه‌ی روش پیشنهادی با فیلتر کردن گره در فرآیند تعمیر لوازم الکترونیکی.....	۸۵
جدول ۵-۹: مقایسه‌ی روش پیشنهادی با فیلتر کردن گره در فرآیند جریمه‌های ترافیکی.....	۸۶
جدول ۵-۱۰: مقایسه‌ی روش پیشنهادی با فیلتر کردن گره در فرآیند درخواست وام بانکی.....	۸۶
جدول ۵-۱۱: مقایسه‌ی نهایی بین روش پیشنهادی و فیلتر کردن گره.....	۸۷
جدول ۵-۱۲: پیش‌پردازش برای خوشه‌بندی به روش خوشه‌بندی ردپاها.....	۸۷
جدول ۵-۱۳: مقایسه‌ی زمانی بین روش پیشنهادی با روش خوشه‌بندی ردپاها.....	۸۸

فصل اول:

مقدمه

۱-۱ تاریخچه

مدیریت فرآیند کسب و کار مجموعه فرآیندهایی است که به سازمان‌ها کمک می‌کند تا کارایی کسب و کارشان را بهبود بخشند. درواقع این مجموعه فرآیندها، کارایی و اثربخشی هر سازمان را با خودکار نمودن فرآیند کسب و کار آن سازمان بهینه می‌سازند. شناسایی فرآیندهای کسب و کار نسبتاً ساده است؛ اما مشخص نمودن بخش‌های مختلف کسب و کار و یافتن صاحبان فرآیندها^۱ دشوار است. مدیریت فرآیند کسب و کار، علاوه بر مدیریت فرآیندها در سازمان‌ها، یکپارچه‌سازی بی‌درنگ فرآیندها را با تامین‌کنندگان، شرکای تجاری و مشتریان بر عهده دارد.

اصولاً با توجه به روند تغییرات محیط‌های کسب و کار، تمرکز اکوسیستم‌های کسب و کار را می‌توان به صورت زیر تقسیم نمود:

- تولید بیشتر (تعداد)، دهه ۱۹۶۰ میلادی
- تولید ارزان (هزینه)، دهه ۱۹۷۰
- تولید بهتر (کیفیت)، دهه ۱۹۸۰
- تولید سریع‌تر (زمان)، دهه ۱۹۹۰
- تنوع تولید بیشتر (سرویس)، قرن ۲۱

از بزرگ‌ترین چالش‌های کسب و کار می‌توان از تغییر بازار و مشتریان، ظهور رقبای جدید و تغییر قوانین کسب و کار سازمان‌ها نام برد که با توجه به این چالش‌ها به مرور زمان نیاز به یک روش و سیستمی برای تعریف، مدیریت، تحلیل و بهینه‌سازی فعالیت‌های کسب و کار احساس می‌شد. در گذشته مرسوم بود که برای هر یک از فرآیندها، سیستم‌های جداگانه‌ای در محدوده‌ی آن فرآیند تهیه می‌شد که به سیستم‌های جزیره‌ای معروف بودند. به مرور زمان مشکلات اساسی این نوع نگرش پدیدار شد که یکی از بزرگ‌ترین آن‌ها این بود که چون هر یک از سیستم‌ها برای یک فرآیند خاص تهیه می‌شد، اتصال

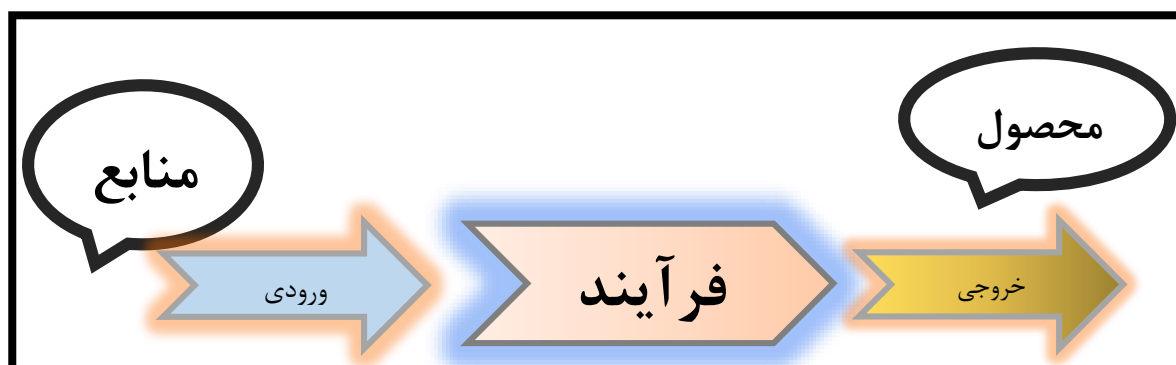
افراد می‌توانند فرآیند برای آن‌ها تعریف شده یا به نوعی در فرآیند نقش مهمی ایفا می‌کنند^۱

سیستم‌ها به یکدیگر مشکل یا حتی غیرممکن بود و از طرفی چون دید طراحی سیستم‌ها در مورد بخشی از سازمان بود و نه کل آن، این سیستم‌ها اهداف کلی سازمان را به نحو مطلوب برآورده نمی‌کردند. بدین صورت ایده‌ی سیستم‌های یکپارچه سازمانی به وجود آمد؛ در این دیدگاه برای سازمان یک سیستم متشکل از تعدادی زیرسیستم مستقل، ولی مرتبط با هم طراحی می‌شد که مشکلات مذکور را مرتفع می‌ساخت. با این حال یک مشکل اساسی که هنوز هم خودنمایی می‌کرد، حجم زیاد کار برای طراحی و پیاده‌سازی این سیستم‌ها بود. به این صورت مدیریت فرآیند کسب و کار مطرح شد؛ مجموعه فرآیندهایی که با داشتن الگوهای متعدد موردنیاز سازمان‌ها، روشی یکپارچه برای تعریف، اجرا و مدیریت فرآیندهای کسب و کار سازمان‌ها را ارائه می‌کرد و با استفاده از متدها و ابزارهای خاص خود حجم کار توسعه راه‌حل‌های اطلاعاتی برای سازمان‌ها را به حداقل می‌رساند. درزمینه‌ی تعریف مدیریت فرآیند کسب و کار و مفهوم آن، بین شرکت‌های مختلف و صاحب‌نظران این رشته اختلافاتی وجود دارد، ولی آنچه عموماً در بین این نظرات مشترک است این است که: مدیریت فرآیند کسب و کار روشی یکپارچه و نظام‌مند برای طراحی، اجرا و پایش فرآیندهای کسب و کار است که ممکن است در هر یک از آن‌ها افراد یا نرم‌افزارهای موجود در سازمان درگیر باشند.

۱-۲ فرآیند

فرآیند مجموعه فعالیت‌های متوالی و مرتبط به هم بوده که محصول خاصی را به وجود می‌آورد و برای ایجاد این محصول به داده‌های خاصی (ورودی) نیاز دارد که زمینه را برای درست عمل نمودن آن فراهم می‌سازند. فرآیند را می‌توان یک سری منطقی از تراکنش‌های مرتبط با یکدیگر که ورودی را به نتایج و خروجی‌ها تبدیل می‌کند تعریف کرد^۲.

² <http://www.netrise.ir/>



شکل ۱-۱: شکل شماتیک فرآیند

فرآیندهای موجود در هر سازمان برای دستیابی به مأموریت سازمان طراحی شده‌اند تا با عملکرد بهتر نیازهای اساسی مشتریان را تامین نمایند. برای پاسخگویی به نیازها و خواسته‌های مشتریان باید فرآیندهای موجود، کارایی و اثربخشی لازم را داشته باشند. یک فرآیند زمانی کارایی لازم را خواهد داشت که به صورت درست انجام گیرد و زمانی از اثربخشی برخوردار خواهد بود که به صورت درست انتخاب و طراحی شده باشد. فرآیند، معرف یک یا دسته‌ای از فعل‌وانفعالات است که به منظور تبدیل داده‌ها به خروجی^۳ انجام می‌گیرد. در یک تعریف معادل می‌توان گفت فرآیند، یک سری منطقی از تراکنش‌های مرتبط با یکدیگر است که ورودی را به نتایج و خروجی‌ها تبدیل می‌کند.

انواع فرآیندها

سه نوع فرآیند کسب‌وکار وجود دارد:

- فرآیندهای مدیریتی:

این فرآیندها، فعالیت‌های نظام‌مند مثل مدیریت راهبردی و نحوه‌ی اداره سازمان را حمایت

می‌کنند.

³ Output

- فرآیندهای عملیاتی (اصلی):

این فرآیندها ارزشی را در راستای کسب و کار اصلی سازمان ایجاد می‌کنند. مثل خرید، تولید، بازاریابی و فروش.

- فرآیندهای پشتیبانی:

این فرآیندها فرآیندهای عملیاتی را پشتیبانی می‌کنند. مثل حسابداری، استخدام و فناوری اطلاعات.

۱-۳ کاوش فرآیند

فرآیندکاوی یا کاوش فرآیند^۴ موضوع نسبتاً جوانی است که بین هوش محاسباتی و داده‌کاوی از یک سو و مدل‌سازی فرآیندهای سازمان و آنالیز از سوی دیگر قرار می‌گیرد. هدف کاوش فرآیند، کشف، نظارت و ارتقای فرآیندهای واقعی از طریق استخراج دانش از داده‌های ذخیره‌شده قابل‌خواندن از سیستم‌های اطلاعاتی امروزه است. کاوش فرآیند بیشتر به آنالیز فرآیندها با استفاده از سابقه‌ی رخدادها^۵ می‌پردازد. فرآیندکاوی یک تکنیک و یک ابزار کمکی برای بازسازی فرآیندهای کسب و کار از رویدادهای ثبت‌شده در یک سیستم اطلاعاتی است. ورودی‌ها برای الگوهای مرتبط با فعالیتی که پس‌از آن به‌عنوان مدل فرآیندهای کسب و کار تکثیرشده‌اند، استخراج می‌شوند^۶. تکنیک‌های فرآیندکاوی از سابقه‌ی رخدادهای ثبت‌شده برای کشف، آنالیز و ارتقای فرآیندهای کسب و کار استفاده می‌کنند. هر رویداد ثبت‌شده به یک فعالیت اشاره دارد و مرتبط با یک مورد (نظیر یک نمونه فرآیند) است.

^۴ Process Mining

^۵ Event Log

^۶ <http://www.zavosh.org/>

در این تحقیق سعی می‌شود با استفاده از اطلاعات ثبت شده در سابقه‌ی رخداد هر فرآیند، فعالیت‌های اجرا شده روی پرونده‌ها را در خوشه‌های مرتبط به هم قرار دهیم که این کار را می‌توان در دسته‌ی کاوش فرآیند جای داد.

۴-۱ ساختار پایان‌نامه

در فصل اول این پایان‌نامه در مورد تاریخچه و مقدماتی از سیستم‌های مدیریت کسب و کار و کاوش فرآیندها توضیح داده شد. در فصل دوم به بیان کارهای انجام‌شده در این زمینه خواهیم پرداخت. در فصل سوم با مفهوم مدیریت فرآیندها و همچنین کاوش فرآیندها بیشتر آشنا خواهیم شد، سپس در فصل چهارم به شرح مسئله و روش پیشنهادی خواهیم پرداخت، در فصل پنجم پیاده‌سازی روش پیشنهادی و نتایج حاصل از آن را بررسی خواهیم کرد و در نهایت در فصل ششم نتایج و پیشنهادهای حاصل از این تحقیق را بیان می‌کنیم.

فصل دوم:

مروری بر کارهای

انجام شده

با توجه به پیشرفت سیستم‌های کامپیوتری و نقش مهم آن در زندگی روزمره‌ی انسان، سیستم‌های مدیریتی نیز با کمک از این پدیده، پیشرفت بسیاری داشته‌اند و توانسته‌اند نقش به‌سزایی در پیشبرد فعالیت‌های انسان و نظارت بر کارها ایفا کنند. سیستم مدیریت جریان کار بخشی از یک نرم‌افزار است که یک زیرساخت برای راه‌اندازی، اجرا و نظارت بر یک گردش جریان کار فراهم می‌کند. درواقع سیستم مدیریت جریان کار یک مدل کلی است که می‌توان تمامی فرآیندهای تجاری را توسط آن مدل کرد. برای ارزیابی سیستم‌ها می‌توان از تکنیک‌های مختلف مدل‌سازی که برای سیستم مدیریت جریان کار فراهم شده استفاده نمود.

یکی از نیازمندی‌های اصلی سیستم مدیریت جریان کار وجود یک طرح جریان کاری است. پس باید بتوان با ارائه‌ی یک مدل، جریان کار در سیستم را به نمایش درآورد. شرط لازم برای ایجاد مدل مدیریت جریان کاری برای هر فرآیند، داشتن اطلاعات کافی در مورد آن فرآیند است. کسب این اطلاعات با استفاده از تکنیک‌هایی از مبحث کاوش فرآیند میسر است که از بررسی و آنالیز فرآیندها بر پایه‌ی سابقه‌ی رخدادهای ثبت‌شده در سیستم به دست می‌آیند.

با استفاده از فرآیندکاوی و آنالیز فرآیند می‌توان به رفتارهای بنیادی و اصلی سیستم پی برد، نقاط ضعف و قوت یک سیستم را شناسایی کرد و برای بهبود سیستم تلاش کرد. طبیعی است که با داشتن رفتارهای اصلی سیستم می‌توان مدل موردنظر را طراحی کرد.

۲-۱ مرور ادبی

کاوش فرآیندها اثبات کرده که یک ابزار ارزشمند برای بررسی عملکرد اجرای فرآیندها بر مبنای داده‌های گزارشی است. متأسفانه درزمینه‌ی فرآیندکاوی با انعطاف زیادی روبه‌رو هستیم که این انعطاف در زمینه کشف و بررسی داده‌های مفید، باعث به‌وجود آمدن مشکلاتی خواهد شد.

ایده‌ی کاوش فرآیندها ایده‌ی جدیدی نیست، آقای کوک و آقای ولف^۷ روی مسائل مشابهی درزمینه‌ی فرآیندهای مهندسی نرم‌افزار مطالعه می‌کردند. در [۱] آن‌ها ۳ روش از کشف فرآیند را توصیف کردند، که یکی از آن‌ها استفاده از شبکه‌های عصبی با نام RNet، دیگری یک روش صرفاً الگوریتمی با نام Ktail و روش مارکوفی بوده است. که به نظر آن‌ها دو روش آخر امیدوارکننده‌تر از شبکه‌های عصبی در این زمینه بود.

روش Ktail یک ماشین حالت متناهی می‌سازد که با بررسی رفتارهای آینده‌ی هر حالت، حالات ممکن برای آن را محاسبه می‌کند. روش مارکوفی نیز از مخلوط روش الگوریتمی و روش‌های آماری برای مقابله با نویز استفاده می‌کند. آقای کوک و آقای ولف کار خود را درزمینه‌ی فرآیندهای موازی نیز گسترش دادند. آن‌ها با پیشنهاد معیارهای خاصی از قبیل (آنتروپی، شمارش رویدادهای یک نوع، نوبت یا دوره رویداد و علیت) و استفاده از آن‌ها به کشف مدل از جریان رویدادها^۸ پرداختند [۲].

آقای کوک و آقای ولف یک معیار اندازه‌گیری برای تعیین اختلاف بین یک مدل فرآیند و رفتارهای واقعی ثبت‌شده بر مبنای داده‌های رویداد محور^۹ ارائه کردند؛ اما آن‌ها نتوانستند موفق به تعریف روشی شوند که یک مدل صریح و دقیق از فرآیند را ارائه کند [۳].

آقای سانگ^{۱۰} روشی ارائه کرده است که در آن داده‌های گزارشی از هم متمایز شده و داخل زیرمجموعه‌هایی متشابه قرار می‌گیرند که می‌توان از هر یک از این زیرمجموعه‌ها یک مدل فرآیند ساخت. در دنیای واقعی ما با فرآیندهایی روبه‌رو هستیم که اکثراً ساخت یافته نیستند، ساخت یافته نبودن بدین معناست که با وجود یک تصور کلی از فرآیند، افراد ممکن است از روند کلی آن خارج یا منحرف شوند. در این تحقیق، مطالعه روی یک شرکت تعمیر لوازم الکترونیک صورت گرفته که ساخت و تعمیر

⁷ Cook and Wolf

⁸ Events Flow

⁹ Event-based Data

¹⁰ Song

محصولات تلفن همراه و جهت‌یابی را بر عهده دارد؛ این شرکت کارهای مختلفی از قبیل دریافت محصول و درخواست برای تعمیر آن، بررسی کردن گارانتی محصول، ارائه فاکتور و... به‌عنوان کارهای صورت گرفته روی یک پرونده^{۱۱} را در داده‌های گزارشی ثبت می‌کند و در انتها تکلیف یک پرونده مشخص می‌شود. با توجه به داده‌های ثبت‌شده ردپایی از رویدادها^{۱۲} ایجاد می‌شود و سپس یک معیار اندازه‌گیری با نام مشخصات فعالیت^{۱۳} که شامل یک شمارش ساده از تمام رویدادهای اتفاق افتاده روی یک پرونده است، تشکیل خواهد شد. همچنین برای نیروی انسانی دخالت‌کننده در هر پرونده نیز یک مشخصه ایجادشده و تعداد دخالت این نیرو در هر پرونده شمارش می‌شود؛ سپس با پیاده‌سازی الگوریتم کلاس‌بندی روی ویژگی‌های تعریف‌شده، پرونده‌های مشابه در یک خوشه قرار می‌گیرند و ردپای مربوط به هر پرونده در این خوشه‌ها می‌تواند به‌عنوان یک زیر فرآیند از فرآیند اصلی محسوب شود [۴].

در [۵] بحث حول محور پویایی فرآیندهای واقعی است و اینکه یک فرآیند ممکن است در طی یک‌زمان طولانی نسخه‌های مختلفی از رفتار را از خود به معرض اجرا بگذارد. مثال ذکر شده در این تحقیق درباره‌ی یک فروشگاه خواروبارفروشی است که نحوه فروش آن در تعطیلات کریسمس و تعطیلات تابستان و حتی تغییر استراتژی فروش برای سازگار شدن با شرایط زمانی را مورد بررسی قرار می‌دهد. هدف از این تحقیق کشف نسخه‌های مختلف یک فرآیند در طول زمان است که با استفاده از تکنیک خوشه‌بندی ردپاها^{۱۴} به‌عنوان پایه‌ی روش ارائه‌شده و افزودن زمان شروع به‌جریان افتادن یک پرونده به مجموعه‌ی ویژگی‌های مورد استفاده در این تکنیک به‌عنوان نوآوری کار، سعی در خوشه‌بندی پرونده‌های شبیه به هم دارد که این خوشه‌ها نمایانگر یک نسخه از فرآیند اصلی هستند؛ به‌بیان دیگر در این تحقیق زمان شروع به‌جریان افتادن یک پرونده را به داده‌های گزارشی سیستم اضافه کرده و به‌عنوان یک ویژگی

¹¹ Case

¹² Trace

¹³ Activity Profile

¹⁴ Trace Clustering

جدید در اختیار روش خوشه‌بندی ردپا می‌گذارد و فرآیند را خوشه‌بندی می‌کند که درواقع از الگوریتم‌های خوشه‌بندی برای رسیدن به این هدف استفاده کرده است.

کاوش فرآیندها امکان استخراج و کشف دانش را بر مبنای سابقه‌ی رخدادها فراهم می‌کند. یک لاگ^{۱۵}، ثبت اجرای فعالیت‌ها در فرآیندهای کاری است. امروزه تکنیک‌ها و ابزارهای فراوانی در کاوش فرآیندها در حال توسعه هستند، هدف اصلی از کاوش فرآیندها یافتن بهترین مدل از فرآیند با استفاده از اطلاعات نهفته در لاگ‌های آن است [۶].

بحث چک کردن انطباق^{۱۶} برای بررسی اینکه آیا رفتار مدل منطبق بر رفتارهای مشاهده‌شده است یا خیر مطرح می‌شود، مفهوم متابعت و چند معیار مربوط به آن در [۷] توضیح داده‌شده است.

خیلی از موارد مطالعه درزمینه‌ی کاوش فرآیندها برای نمایش قابل اجرا بودن آن‌ها انجام‌شده است، در [۸] در مورد روش‌هایی برای مدیریت کردن فرآیندهای پیچیده که از لاگ‌های فرآیندهای واقعی به وجود می‌آیند بحث شده است. آقای گانتز^{۱۷} و آقای آلس^{۱۸} در [۹] یک تکنیک فازی برای کاوش فرآیندها پیشنهاد می‌دهند که به ساده‌سازی مدل یک فرآیند و برجسته‌سازی جریان‌های مهم در فرآیند می‌پردازد.

آقای گرکو^{۱۹} در [۶] از خوشه‌بندی ردپاها برای کشف مدل‌های پرمعنی استفاده کرده‌اند. ایشان این کار را تنها با استفاده از فعالیت‌ها و انتقال‌هایشان^{۲۰} برای ایجاد خوشه‌ها انجام داده‌اند.

کاوش فرآیندها یک تکنیک برای استخراج مدل از لاگ‌های اجراشده است. همان‌طور که ذکر شد فرآیندهای دنیای واقعی از ساختار خاصی پیروی نمی‌کنند که روش‌های سنتی برای کشف این فرآیندها

¹⁵ Log

¹⁶ Conformance Checking

¹⁷ Gunther

¹⁸ Aalst

¹⁹ Greco

²⁰ Transition

ناتوان‌اند. اصطلاحاً به مدل‌هایی که توسط این روش‌ها استخراج می‌شود اسپاگتی مانند^{۲۱} می‌گویند. در تحقیق [۹] روشی برای غلبه بر این مشکل ارائه می‌شود که می‌تواند یک مدل ساده‌شده از یک فرآیند غیر ساخت‌یافته را ارائه کند. کاوش فرآیندها یک مسیر از پژوهش است که تلاش آن بر این است که یک دید انتزاعی یا نمایش بسته‌بندی‌شده را از لاگ‌های یک فرآیند استخراج کند [۱۰].

۲-۲ الگوریتم‌های معروف در کاوش فرآیند

در کاوش فرآیندها با چند الگوریتم معروف روبه‌رو می‌شویم که به‌اختصار در زیر توضیح داده شده‌اند.

- الگوریتم آلفا [۱۱]^{۲۲}: یک الگوریتم با قابلیت بازسازی گردش کار توسط شبکه پتری با استفاده از ترتیب روابط پیداشده در داده‌های گزارشی است. این الگوریتم‌ها نیاز به اطلاعات وسیعی برای رسیدن به کیفیت مناسب دارند به‌عبارت‌دیگر باید یک لاگ قدرتمند که تمامی روابط موجود در فرآیند را شامل می‌شود موجود باشد تا بتوان نتیجه‌ی خوبی از این الگوریتم دریافت کرد.
- روش‌های استنتاج [۱۲]^{۲۳}: یک مجموعه از ۳ الگوریتم مختلف برای استنتاج ماشین حالت محدود^{۲۴} از داده‌های گزارشی وجود دارد که لاگ در این روش به‌صورت دنباله‌ای ساده از علائم^{۲۵} می‌باشد. این الگوریتم‌ها نشان‌دهنده‌ی سطوح مختلف از ترکیب یا همکاری بین دقت و قدرت در مقابله با نویز هستند که الگوریتم مارکوف با الهام از مدل مارکوف امیدوارکننده‌تر از همه به نظر می‌رسد. این الگوریتم کار را با ساخت نموداری از رویدادها با توجه به نتیجه‌ی زنجیره‌ی مارکوف شروع می‌کند و در گام آخر با تبدیل نمودار به ماشین با حالت محدود، فرآیند کشف‌شده را به نمایش می‌گذارد.

²¹ Spaghetti Like

²² α -Algorithm

²³ Inference Methods

²⁴ FSM

²⁵ Symbols

- گراف جهت‌دار غیر مدور [۱۳]^{۲۶}: یک الگوریتم که قادر به ساخت نمودار وابستگی از لاگ یک سیستم گردش کار است. لاگ بایستی شامل تعداد نسبتاً زیادی از اجرای فرآیندهای هم‌روند باشد تا گراف وابستگی بتواند برای فرآیند به‌طور کامل ساخته شود.
- کسب گردش کار القایی [۱۴]^{۲۷}: روشی است که هدف آن یافتن مدل مخفی مارکوف^{۲۸} می‌باشد که این مدل بهترین نمایش‌دهنده‌ی ساختار از یک فرآیند است. ساختار اولیه مدل مخفی مارکوف را نیز می‌توان از لاگ استخراج کرد که دنباله‌ای از علائم را شامل می‌شود، مدل مخفی نهایی را نیز می‌توان با پاک‌سازی بالا به پایین یا بالعکس از مدل مخفی اولیه به دست آورد.
- خوشه‌بندی سلسله مراتبی [۱۵]^{۲۹}: یک الگوریتم که با توجه به یک مجموعه‌ی بزرگ از اجرای دنباله‌ها که به آن داده می‌شود آن‌ها را به خوشه‌هایی تقسیم کرده و نمودار وابستگی را برای هر خوشه به‌طور جداگانه ایجاد می‌کند. خوشه‌های گردش کار در یک درخت سازمان‌دهی می‌شوند و از همین روست که مفهوم سلسله مراتبی روشن می‌شود. سپس بعد از ایجاد مدل گردش کار برای هر خوشه با یک پیمایش پایین به بالای درخت، به درخت واحد برای مدل دست می‌یابیم.
- گراف نمونه‌ها [۱۶]^{۳۰}: یک روش که هدف آن به تصویر کشیدن یک نمای گرافیکی از اجرای فرآیندها است که با پیدا کردن اجرای هر دنباله در لاگ یک نمودار برای نمونه فرآیند^{۳۱} (حرکت یک پرونده از شروع تا پایان فرآیند) ایجاد می‌کند. در این روش برای شناسایی روندهای موازی در هر نمونه گراف از وابستگی‌های یافت شده در کل فرآیند استفاده کرده و خود را محدود به لاگ یک نمونه فرآیند نمی‌کند.

²⁶ Directed Acyclic Graphs

²⁷ Inductive Workflow Acquisition

²⁸ HMM

²⁹ Hierarchical Clustering

³⁰ Instance Graphs

³¹ Process Instance

۲-۳ رویکردهای ساده‌سازی موجود در کاوش فرآیند

تکنیک‌های کاوش فرآیند که برای محیط‌های غیر ساخت‌یافته مناسب‌اند، باید قادر به فراهم آوردن یک دید سطح بالا از فرآیند باشند. به‌عنوان مثال اگر شما یک نقشه را در نظر بگیرید با چند پدیده در این زمینه روبه‌رو خواهید شد که عبارت‌اند از:

تجمع یا تراکم: برای محدود کردن اطلاعات مورد نمایش، به‌عنوان مثال جاده‌های شهری، شهرها و خانه‌ها به‌صورت یک بسته در نقشه‌ی مربوط به بزرگ‌راه‌ها استفاده می‌شود.

انتزاعی کردن: اطلاعات با سطح پایین در زمینه‌ای که مدنظر ما نیست مثل راه‌های مرتبط با دوچرخه‌سواری که هیچ اطلاعات دقیقی را نمی‌تواند در یک نقشه‌ی مربوط به بزرگ‌راه به ما بدهد قابل حذف است.

تأکید: اطلاعات با ارزش بیشتر می‌توانند با جزئیات کامل نمایش داده شوند.

در فرآیندها ۳ رویکرد موجود است که ممکن است با آن‌ها در کاوش فرآیندها روبه‌رو شویم [۱۷].

- رفتارهای با اهمیت بالا که با جزئیات تمام و به‌صورت دست‌نخورده باقی می‌مانند.

- رفتارهای با اهمیت کمتر اما همبستگی زیاد که به‌صورت یک خوشه نمایش داده می‌شوند.

- رفتارهای با اهمیت کمتر و همبستگی کم که در مدل ساده‌شده حذف می‌شوند.

این تحقیق با استفاده از این اطلاعات سعی بر ایجاد یک مدل ساده‌شده از فرآیند با بررسی این پارامترها دارد و روش‌هایی برای ساده‌سازی فرآیندها با استفاده از فیلترینگ لاگ‌ها و حذف رویدادهای کم‌اهمیت از لاگ‌ها را پیشنهاد می‌دهد.

در روش‌های ساده‌سازی از مفاهیم خوشه‌بندی داده‌ها و خوشه‌بندی گراف‌ها استفاده می‌شود؛ در خوشه‌بندی داده‌ها تلاش بر این است که بتوان با یافتن رابطه بین رویدادها با استفاده از خواص آن‌ها و معیارهای فاصله و استنباط روی این معیارها، رویدادهای مربوطه را خوشه‌بندی کرد [۱۸]، در سوی دیگر الگوریتم‌های خوشه‌بندی گراف بر مبنای آنالیز خواص ساختاری گراف می‌باشد که استراتژی‌های پارتیشن‌بندی از آن‌ها مشتق می‌شود [۱۹].

خوشه‌بندی دنباله‌ها^{۳۲} که بیشتر در علم داده‌کاوی مورد استفاده قرار می‌گیرد نیز در زمینه‌ی کاوش فرآیندها به کار گرفته شده است که هدف اصلی آن بخش‌بندی دنباله‌هایی^{۳۳} از واحدهای کاری داخل خوشه‌های با معنی است که در آن هر خوشه شامل یک زنجیره‌ی مارکوف مرتبه‌ی اول است و در آن هر مرحله فقط و فقط به مرحله‌ی قبلی آن وابسته است و هر خوشه شامل یک زنجیره‌ی مارکوف است که می‌تواند دنباله‌های مربوط به کلاس خودش را ایجاد کند [۲۰].

با توجه به جدید بودن مبحث کاوش فرآیندهای کاری می‌توان کارهای مختلفی در این زمینه برای رسیدن به نتایج بهتر و سرعت عمل بهتر در کشف فرآیندها انجام داد. در این تحقیق از الگوریتم‌های معروف خوشه‌بندی استفاده می‌شود و سعی بر آن است با پیشنهاد روشی جدید به نتایج بهتری در زمینه‌ی خوشه‌بندی فرآیند دست پیدا کرد. در این تحقیق تلاش بر این است تا بتوان با ارائه‌ی روشی فعالیت‌های فرآیند را به صورت معنی‌دار خوشه‌بندی کرده و از خوشه‌های ایجادشده در مسیر ساده‌سازی فرآیند استفاده کرد.

³² Sequence Clustering

³³ Sequence

فصل سوم:

سیستم‌های مدیریت فرآیند کسب و کار

و کاوش فرآیندها

امروزه کسب و کارها با تغییرات سریعی روبه‌رو هستند. در فضای رقابتی کنونی، سازمانی می‌تواند به حیات خود ادامه دهد که سازوکارهای لازم را برای رویارویی با این تغییرات دارا باشد. یکی از رویکردهای جدیدی که به حیات سازمان‌ها در فضای پرقاب‌ت امروز کمک می‌کند، رویکرد فرآیند محوری و مدیریت مبتنی بر فرآیند است. سازمان‌ها در گذشته برای تک‌تک فرآیندها، سیستم‌های جداگانه‌ای در محدوده‌ی آن فرآیند تهیه می‌کردند که به سیستم‌های جزیره‌ای معروف بودند ولی به دلیل عدم امکان ارتباط به یکدیگر و نیز طراحی سیستم‌ها در مورد فقط یک بخش جداگانه، اهداف کلی سازمان به‌صورت مطلوب برآورده نمی‌شد؛ بنابراین سیستم‌های سازمانی شکل گرفت. در رویکرد جامع یک سیستم از تعدادی زیرسیستم تشکیل شده که مرتبط با هم طراحی می‌شوند ولی مشکل آن کار زیاد طراحی و پیاده‌سازی آن بود.

درنهایت مدیریت فرآیند کسب و کار مطرح شد که با داشتن الگوهای متعدد در مورد نیاز سازمان‌ها روشی یکپارچه برای تعریف، اجرا، بازبینی و مدیریت فرآیندهای کسب و کار سازمان ارائه می‌کرد و با استفاده از روش‌ها و ابزارهای مربوطه حجم کار را به حداقل می‌رساند و مدیریت امور را سهل و کارآمد می‌نمود.

با توجه به قوت گرفتن نقش فرآیندهای کسب و کار، بحث کاوش فرآیندها نیز برای رسیدن به نتایج بهینه‌تر و هزینه‌های کمتر شکل گرفت که با استفاده از اطلاعات موجود در سیستم سعی در به دست آوردن یک دید کلی نسبت سیستم و یافتن نقاط ضعف و قوت یک سیستم دارد.

۳-۱ سیستم‌های مدیریت فرآیند کسب و کار (BPMS)^{۳۴} چیست؟

سیستم مدیریت فرآیند کسب و کار یا همان BPMS را می‌توان نقطه‌ی اوج همه تجربیات مشترک، تفکرات و پیشرفت‌های حرفه‌ای در مدیریت کسب و کار در طی چند دهه‌ی اخیر دانست.

³⁴ Business Process Management System

BPMS مجموعه‌ای از روش‌ها، ابزارها و فناوری‌هایی است که برای طراحی، اجرا، تجزیه و تحلیل و کنترل فرآیندهای عملیاتی کسب و کار استفاده می‌شود و در آن سیستم‌های اطلاعاتی و افراد به هم پیوند می‌خورند به طوری که مشخص کننده این است که ((چه زمانی، چه فردی، چه کاری را به انجام می‌رساند)).

از آنجاکه فرآیندهای تجاری کلید موفقیت هر سازمان است، داشتن یک سیستم مدیریت فرآیند کسب و کار در سازمان حائز اهمیت است؛ زیرا سازمان را قادر می‌سازد تا فرآیند کسب و کاری جدید را ایجاد کند و فرآیندهای موجود را مطابق با تغییرات محیطی بهبود بخشد.

ابزارهای مدیریت فرآیند کسب و کار برای پیاده‌سازی فرآیندهای کاری از ایجاد هماهنگی فعالیت‌های بین افراد و سیستم‌ها استفاده می‌کنند، عناصر حیاتی این سیستم عبارت‌اند از [۲۱]

موتور فرآیند: چارچوبی برای مدل‌سازی و اجرای نرم‌افزاری فرآیند محور و ابزارهای نظارت^{۳۵} بر روی فرآیندهای موجود است.

سیستم گزارش ساز: توانمندسازی مدیران در جهت شناسایی مسائل، روند و فرصت‌های کسب و کار از طریق گزارش‌هاست.

موتور اجرای قوانین:^{۳۶} مدیریت قوانین بر روی فرآیندها را بر اساس آیین‌نامه‌ها و بخشنامه‌های سازمان فراهم می‌سازد.

ابزارهای تعاملی: از طریق اتاق‌های گفتگو محیط‌های کاری پویا و محل تبادل پیام‌ها مرزهای ارتباطی بین بخش‌های درون و برون سازمانی را حفظ می‌کند.

³⁵ Monitoring

³⁶ Rule Engine

ابزارهای یکپارچه‌سازی: چارچوبی سرویس‌گرا به منظور اتصال، ارتباط و یکپارچه‌سازی کلیه

سیستم‌های موجود در سازمان است.

فرآیند نشان می‌دهد که چگونه کار با شرح وظایف داخلی سازمان انجام می‌شود. مدیریت فرآیند محور روابط تامین‌کنندگان و مشتریان را با فرآیندهای کسب‌وکار نشان می‌دهد.

۲-۳ چرخه حیات فرآیندهای کسب و کار

مدیریت فرآیند تجاری طی یک چرخه اتفاق می‌افتد. این چرخه شامل فازهایی است که به هم مرتبط می‌باشند:

- طراحی و آنالیز

فاز طراحی و آنالیز نقطه شروع این چرخه است. در این مرحله سازمان، فرآیندهای تجاری و محیط فنی و تکنیکی بررسی می‌شوند و بر مبنای این بررسی فرآیندها شناسایی، بازبینی، اعتبارسنجی می‌شوند و در نهایت با مدل‌سازی ارائه می‌شوند. فرآیندهایی که با مدل‌ها و نشان‌گذاری‌های گرافیکی ارائه می‌شوند باعث تسهیل در فهم فرآیندها می‌شوند بنابراین ذینفعان مختلف می‌توانند به صورت مؤثر در تعریف و بهبود فرآیندها شرکت کنند. تکنیک‌های مدل‌سازی، اعتبارسنجی، شبیه‌سازی و تأیید فرآیند در این فاز انجام می‌شوند.

- پی‌کربندی

پس از مدل‌سازی و تأیید فرآیندها، این فرآیندها باید پیاده‌سازی شوند. روش‌های مختلفی برای انجام این کار موجود است. آن‌ها می‌توانند با یک مجموعه از سیاست‌ها و رویه‌هایی که توسط افراد سازمان پذیرفته شده است انجام شوند. در این حالت ممکن است که اجرای فرآیندها بدون هیچ سیستم نرم‌افزاری باشد. در این مرحله سیستم نرم‌افزاری برای اجرای فرآیندها پیاده‌سازی می‌شود. مدل‌سازی

فرآیندها در پیاده‌سازی این سیستم‌ها کمک می‌کند. این سیستم باید مطابق با محیط سازمان و فرآیندهای تأییدشده آن باشد. این پیکربندی شامل تعاملات کارمندان با سیستم و یکپارچه کردن سیستم‌های نرم‌افزاری موجود با سیستم مدیریت فرآیند تجاری است. اغلب در سازمان‌ها فرآیندها با سیستم‌های موجود پشتیبانی می‌شوند در این فاز پیکربندی باید بتوان سیستم‌های موجود را به سیستم مدیریت فرآیند تجاری ضمیمه کرد. سیستم مدیریت اطلاعات فرآیندهای تجاری جنبه تراکنشی دارد.

• اجرای کردن

بعد از پیکربندی سیستم فرآیندها به صورت قانونی درمی‌آیند. این فاز فرآیندهای واقعی و در حال اجرا را شامل می‌شود. فرآیندها برای رسیدن به اهداف کسب و کار راه‌اندازی شده‌اند. این فرآیندها رویدادهای تعریف‌شده را دنبال می‌کنند. سیستم مدیریت فرآیند تجاری به طور مؤثر اجرای فرآیندهای تجاری را کنترل می‌کند. مؤلفه نظارت در سیستم مدیریت فرآیند تجاری وضعیت فرآیندها را مشخص می‌کند. نظارت فرآیندها یک مکانیزم مهم در صحت اطلاعات فرآیندها است. اطلاعات مربوط به جزئیات فرآیندها در سیستم مدیریت فرآیند تجاری قابل دسترسی است. اطلاعات نظارتی بر مبنای وضعیت فرآیندها است. در این فاز داده‌های اجرایی با ارزش به صورت فایل لاگ^{۳۷} جمع‌آوری می‌شوند.

• ارزیابی

در این فاز از اطلاعات موجود برای ارزیابی و بهبود مدل‌های فرآیندی استفاده می‌شود. این ارزیابی با تکنیک‌های مانند کاوش فرآیند و نظارت بر فعالیت‌های تجاری انجام می‌شود.

این تکنیک‌ها به شرح زیر می‌باشند:

• کاوش فرآیند

³⁷ Log File

کاوش فرآیند مجموعه‌ای از روش‌ها و ابزارهای مرتبط با نظارت بر فرآیند است. هدف از کاوش فرآیند، تحلیل داده‌های گزارش‌شده از وقایع ثبت‌شده در سیستم^{۳۸} در طی نظارت بر فرآیند و مقایسه آن‌ها با یک مدل فرآیند منطقی است. این کاوش را از سه منظر می‌توان بررسی کرد:

۱. منظر فرآیندی: در این منظر کنترل گردش کار و علت انجام فرآیندها بررسی می‌شود.

۲. منظر سازمانی: در این منظر ساختار سازمانی و افراد درگیر در فرآیند موردنظر می‌باشند.

۳. منظر داده‌ای: در این منظر داده‌ها مورد بررسی قرار می‌گیرند.

• نظارت بر فعالیت تجاری

نظارت خودکار بر فعالیت فرآیند کسب و کار است و به‌صورت مازولی در سیستم مدیریت فرآیند تجاری پیاده‌سازی می‌شود. به‌منظور استفاده از BPM باید شاخص‌های کلیدی کارایی شناسایی شوند و سیستمی به‌منظور نظارت و پاسخگویی به تغییرات به‌صورت بی‌درنگ ساخته شود.

³⁸ Event Log



شکل ۳-۱: چرخه‌ی حیات مدیریت فرآیند کسب‌وکار

۳-۳ کاربرد مدیریت فرآیندها

چالش‌هایی که سازمان‌ها امروزه با آن روبرو هستند نسبت به گذشته پیچیده‌تر است. سازمان‌ها دائم تحت فشارند تا راه‌حل‌های جدیدی برای مسائل و مشکلات پیش رو پیدا کرده و همیشه سعی دارند خدماتی بهتر ارائه نمایند که از استانداردهای بالایی برخوردار باشند. دستگاه‌های اجرایی ارائه‌کننده خدمات عمومی مجبورند به انتظارات نیازمندان، علاقه‌مندان و مشتریان از جمله شهروندان عادی، مسئولان سیاسی، بخش غیردولتی و غیره پاسخ دهند؛ اما سؤال اینجاست که سازمان‌ها چگونه می‌توانند نقش خود را در چنین سناریویی به‌طور مؤثر ایفا نمایند.

مهندسی مجدد و معماری سازمانی طرح‌هایی جامع برای توصیف فعلی و آتی ساختار و رفتار فرآیندها، سیستم‌های اطلاعاتی، پرسنلی و واحدهای سازمان می‌باشند و درواقع، به دنبال هم‌راستا نمودن آن‌ها با اهداف و استراتژی‌های سازمان هستند. اگرچه عموماً فناوری اطلاعات به‌عنوان جزء اصلی آن‌ها در نظر گرفته می‌شود؛ ولی هدف این‌گونه طرح‌ها بهینه‌سازی سازمان با مشخص ساختن معماری کسب‌وکار، معماری فرآیندها و مدیریت عملکرد مبتنی بر استراتژی است.

سیستم مدیریت فرآیندهای کسب و کار از بروزترین راهکارها به منظور پاسخ به نیازهای جدید و متغیر سازمان‌ها است. در ادامه مثال‌هایی از راهکارهای مبتنی بر سیستم مدیریت فرآیند برای فرآیندهای چند سازمان مختلف را بررسی می‌کنیم.

۳-۳-۱ کاربرد BPMS در بیمه: توسعه کسب و کار

راهکار مبتنی بر BPMS :

زمانی که یک درخواست از مشتری جدید دریافت کنیم، سیستم مدیریت فرآیند کسب و کار اطمینان حاصل می‌نماید که این درخواست شامل اطلاعات معتبر است و تکمیل کارهای مرتبط به این مورد را پیش برده و به صورت خودکار درمی‌آورد، به صورتی که بتوان صورت‌حسابی برای حساب جدید تهیه نمود. زمانی که بیمه‌گر یک درخواست رسمی کارگزار را جهت تعهد دریافت می‌کند، چندین مرحله را تکمیل کرده و فرآیند را به یک متخصص پذیرش درخواست ارائه می‌کند. سیستم سپس الحاقیه، نامه ارسالی به پیمانکار و تفاهم‌نامه را تکمیل کرده و فرم ثبت‌نام را برای کارگزار ارسال می‌کند. زمانی که فرم ثبت‌نام کامل شده دریافت می‌شود، اطلاعات موردنیاز وارد سیستم می‌شوند که در دسترس سیستم‌های مرتبط برای تنظیم حساب قرار می‌گیرند. سپس این درخواست به صورت خودکار به فایل‌های الکترونیکی ثابت فرستاده می‌شوند، بسته‌های تصدیق به متقاضی تحویل داده می‌شود و واحد خدمات مشتری در مورد وجود یک حساب جدید اخطار می‌دهد. تمام کارها در طی این مسیر با هشدارها و آستانه زمانی بررسی و کنترل می‌شوند.

۳-۳-۲ کاربرد برای دولت: پیگیری مکاتبات

راهکار مبتنی بر BPMS:

مکاتبات در یک سیستم مدیریت فرآیند کسب و کار تحت وب وارد و دریافت می‌شوند. فرآیند ایجاد پاسخ به صورت خودکار شروع می‌شود. راه حل، یک منبع مرکزی و قابل جستجو را برای تمام مکاتبات ارائه می‌دهد و پاسخ‌ها به صورتی تعیین شده‌اند که پاسخ‌دهندگان می‌توانند دوباره از مفاد موجود آن استفاده کنند. به کارکنان در مورد موعد زمانی آتی و گذشته اخطار داده می‌شود. پاسخ‌ها به راحتی در میان یک گروه از کارکنان به اشتراک گذاشته می‌شود که این امر خطا را در یک ایمیل فردی از بین می‌برد. زمانی که پاسخ‌ها کامل شوند، پرونده درون سیستم بسته شده و مکاتبات از طریق یک روش منظم ارسال می‌شوند.

۳-۳-۳ کاربرد برای تولید؛ دسترسی به امکانات

راهکار مبتنی بر BPMS :

یک واحد یا مدیر پروژه یک درخواست مجوز امنیتی را در سیستم مطرح می‌نماید، این درخواست شامل نوعی دسترسی مورد نیاز به امکانات است. سپس این درخواست‌ها در صورت نیاز به تصویب یا رد گواهی حمایت و برای حصول اطمینان از امنیت تأیید نهایی توسط افراد مجاز پیگیری می‌شوند. تمام درخواست‌های ثبت شده با هدف تهیه گزارش حفظ می‌شوند. در هر زمان داده شده، کاربران مجاز می‌توانند وضعیت تأییدیه‌ها را به منظور کاهش موانع پیشرفت و تعیین اینکه آیا شرایط ترقی تضمین شده‌اند، مشاهده کنند. اگر یک بررسی درخواست در سیستم از آستانه زمان تعیین شده برای آن بگذرد، یک ایمیل اخطار ارسال می‌شود تا در مورد این تأخیر اخطار دهد. یادآور تمدید نشان به صورت خودکار برای حمایت از کارکنان پس از ۱۱ ماه فرستاده می‌شود. یک پایگاه داده متمرکز که از تمام جهات پذیرش یا امنیت، اطلاعاتی معتبر از یک بازدیدکننده و سطح مناسبی از دسترسی را فراهم می‌کند، می‌تواند مشاهده شود.

چند مثال ذکر شده بخش کوچکی از مثال‌های کاربردی مدیریت فرآیندها در سازمان‌ها است که می‌توان به کاربردهای این پدیده در زمینه‌های مدیریت کسب سود، راهبری پروژه‌ها (مدیریت پروژه)، خدمات مالی (ارائه وام)، مدیریت ریسک، سلامت و بهداشت و... نیز اشاره کرد.

۳-۴ مزایای مجهز شدن یک سازمان به BPMS

مدیریت فرآیند تجاری علاوه بر خودکارسازی گردش کار قابلیت‌های دیگری را فراهم کرده است این قابلیت‌ها به سازمان‌ها این امکان را می‌دهند که بتوانند فرآیندهای بیشتری را کنترل کنند. برخی از مزایای استفاده از BPM به شرح زیر است:

- مستندسازی و تعریف فرآیندها: استانداردهایی مانند BPMN^{۳۹} امکان مستندسازی فرآیندها را فراهم می‌نماید.
- خودکارسازی اجرای فرآیندها: با BPM تمام قوانین و منطق‌های تجاری سازمان خودکار خواهند شد.
- شناسایی فرصت‌ها و بهبود فرآیندها: همچنین BPM معیارهایی را فراهم می‌کند برای اندازه‌گیری هزینه فرآیندها و زمان اجرا که در این صورت بهینه‌سازی بر اساس نتایج واقعی خواهد بود.
- حذف فعالیت‌های غیرضروری: در BPM به کمک مدل‌سازی فرآیندها، سازمان‌ها می‌توانند فرصتی برای حذف کارهای غیرضروری داشته باشند.
- کنترل کارایی فرآیندهای در حال اجرا: BPM با ابزارهای نظارتی، امکان نظارت وضعیت فرآیندها را ممکن می‌سازد در نتیجه این کنترل باعث ثبات و سازگاری فرآیندها برای رسیدن به کیفیت

³⁹ Business Process Modeling Notation

بیشتر و بهینه کردن آن‌ها برای کارایی بیشتر می‌شود همچنین قابلیت اندازه‌گیری آن‌ها باعث می‌شود که دید مدیریتی بهتری داشته باشیم.

- همکاری مشتری‌ها و شرکا در فرآیندهای تجاری: BPM امکان همکاری مشتریان و شرکا را از خارج از سازمان فراهم می‌کند؛ و برای فرآیندهایی که در خارج از محدوده سازمان قرار دارند. بسیار کاربردی خواهد بود.
- کاهش منابع موردنیاز: فرآیندهای کسب و کار به افراد و منابع بسیاری نیاز دارند تا اجرا شوند. BPM می‌تواند تعداد منابع موردنیاز برای یک فرآیند را کاهش می‌دهد.
- افزایش هماهنگی: BPM هماهنگی بین بخش‌های مختلف یک شرکت را از بعد جغرافیایی بهبود می‌بخشد.
- افزایش سرعت اجرای سیکل فرآیندها: BPM با کاهش زمان اجرای فرآیندها و امکان اجرای موازی آن‌ها سرعت کسب‌وکار را بهبود می‌دهد.
- افزایش رضایتمندی مشتری: با کاهش زمان اجرا و اطمینان از درستی آن‌ها باعث می‌شود که مشتریان سریع‌تر و آسان‌تر به نیازمندی‌های خودشان برسند.
- چابکی سازمان: BPM این امکان را برای سازمان‌ها فراهم می‌کند که در صورت تغییر شرایط به راحتی بتوانند تغییرات را در فرآیندها اعمال کنند. به این ترتیب به حفظ موقعیت سازمان در بازار رقابتی کمک می‌نماید.

۳-۵ مدل سازی سیستم‌های مدیریت جریان کاری

برای مدل سازی سیستم‌های مدیریت جریان کار روش‌های مختلفی وجود دارد که با توجه به نیاز سازمان‌ها به چند طریق قابل پیاده‌سازی می‌باشند:

۳-۵-۱ زنجیره‌ی فرآیند مبتنی بر رخداد^{۴۰}

زنجیره‌ی فرآیند مبتنی بر رخداد را می‌توان گرافی مرتب از رویدادها و توابع که دارای ارتباط‌دهنده‌های گوناگونی است که باعث اجرای فرآیندهای موازی هم‌زمان می‌شود دانست [۲۲].

این شیوه از عملگرهای منطقی برای ارتباط استفاده می‌کند که یکی از مهم‌ترین ویژگی‌های آن ساده و قابل فهم بودن آن است و به همین دلیل به‌طور گسترده در فرآیندهای تجاری به کار می‌رود.

۳-۵-۲ UML

UML یک‌زبان مدل‌سازی ویژوال است که برای سندسازی در نرم‌افزارها و پروژه‌ها استفاده می‌شود و باعث ساده‌سازی یک پروژه نرم‌افزاری می‌گردد. این روش برای نمایش از گراف استفاده می‌کند. UML یک‌زبان برنامه‌نویسی نیست بلکه ابزاری است که برای استفاده دیگر زبان‌های برنامه‌نویسی اطلاعاتی را فراهم می‌کند و برای نشان دادن اجزای مختلف سیستم از ۷ دیاگرام استفاده می‌کند که هر یک جنبه‌ای از فعالیت‌ها را زیر نظر دارد [۲۳].

۳-۵-۳ دیاگرام فرآیندهای تجاری^{۴۱}

این دیاگرام به دو منظور طراحی می‌شود:

۱. به منظور ساده کردن و قابل فهم کردن فرآیندهای تجاری کاربرد دارد به گونه‌ای که شما را قادر می‌سازد که یک سیستم را مدل کنید بدون اینکه به اطلاعات خاصی نیازمند باشید.

⁴⁰ Event-driven Process Chain(EPC)

⁴¹ Business Process Diagram

۲. امکان مدل‌سازی فرآیندهای تجاری پیچیده را دارد و به‌سادگی می‌تواند به

زبان‌های اجرایی تجاری نگاشته شود [۲۴].

برای مدل‌سازی یک فرآیند از یک رخداد که به‌عنوان آغازکننده فرآیند است

شروع می‌کنیم سپس فرآیند ادامه پیدا می‌کند تا به رویداد انتهایی برسیم و

فرآیند خاتمه پیدا کند.

۳-۵-۴ شبکه‌های پتری^{۴۲}

این شبکه‌های توسط آدام پتری [۲۵] در اوت ۱۹۳۹ به‌منظور توصیف فرآیندهای فیزیکی مطرح شد.

شبکه‌های پتری یک گراف دوبخشی است که از دو نود تشکیل شده است:

۱. نود گذر^{۴۳} (رویدادی که ممکن است رخ دهد)

۲. نود مکان^{۴۴} (شرایط و حالات را نشان می‌دهد)

یال‌های جهت‌دار: که ارتباط بین نودهای گذر و مکان را برقرار می‌کند. این شبکه‌ها به دو صورت جبر

ریاضی یا گرافیکی نمایش داده می‌شوند که شیوه‌ی گرافیکی آن به درک بیشتر کاربران کمک می‌کند.

⁴² Petri Net

⁴³ Transition

⁴⁴ Place

۳-۶ کاوش فرآیند

در فصل اول به صورت خیلی مختصر کاوش فرآیند تعریف شد، با داشتن یک دید ابتدایی نسبت به این پدیده به مقایسه میان داده کاوی و فرآیند کاوی و انواع چالش‌هایی که در این زمینه مطرح می‌شود خواهیم پرداخت.

۳-۷ مقایسه فرآیند کاوی با داده کاوی

در حقیقت فناوری کاوش فرآیند قدرت داده کاوی و مدل سازی فرآیند را ترکیب می‌کند؛ با تولید خودکار مدل فرآیندها بر مبنای داده‌های وقایع ثبت شده، کاوش فرآیند منجر به ایجاد مدل‌های زنده با قابلیت به روزرسانی بالا می‌شود.

۱. حجم عظیمی از داده‌ها:

کاوش فرآیند مشترکات زیادی با داده کاوی دارد. دقیقاً مشابه داده کاوی فرآیند کاوی هم با چالش پردازش حجم بزرگ داده‌ها که نمی‌توان با دست انجام داد، مواجه است. سیستم‌های فناوری اطلاعات داده‌های زیادی درباره‌ی فرآیندهای تجاری مورد پشتیبانی خود جمع‌آوری می‌کنند. این داده‌ها به خوبی بیانگر آنچه در دنیای واقعی اتفاق افتاده هستند و قابلیت استفاده برای درک و بهبود تجارت را دارند.

۲. دیدگاه فرآیندی:

برخلاف داده کاوی، فرآیند کاوی بر دیدگاه فرآیندی تمرکز می‌کند؛ یعنی شامل جنبه‌های موقتی و نگاه به یک اجرای فرآیند از منظر تعدادی فعالیت اجرا شده است. بیشتر تکنیک‌های داده کاوی الگوهای انتزاعی را در قالبی مانند قوانین و یا درخت تصمیم استخراج می‌کنند؛ اما فرآیند کاوی مدل فرآیندهای کاملی ایجاد می‌کند و سپس از آن‌ها برای شناسایی گلوگاه استفاده می‌کند.

۳. استثناءها نیز مهم هستند:

در داده کاوی عمومی سازی به منظور جلوگیری از سرریز شدن داده ها امری بسیار مهم است. این یعنی می خواهیم تمام داده هایی را که با قانون کلی سازگاری ندارند دور بیندازیم. در فرآیند کاوی نیز عمومی سازی کار کردن با فرآیندهای پیچیده و درک جریان فرآیندهای اصلی لازم است. همچنین در بیشتر موارد درک استثناءها به منظور کشف نقاط ناکارآمد ضروری به نظر می رسد.

۴. تمرکز بر روی اکتشاف:

در داده کاوی معمولاً مدل ها برای پیش بینی نمونه های مشابه در آینده استفاده می شوند. در واقع روش های داده کاوی و یادگیری ماشین کمی وجود دارند که مانند یک جعبه سیاه پیش بینی هایی تولید می کنند بدون اینکه امکان برگشت به عقب و یا بیان علت آن ها را داشته باشند. از آنجاکه فرآیندهای تجاری کنونی خیلی پیچیده هستند پیش بینی های دقیق معمولاً غیر واقعی هستند. دانش به دست آمده و بینش عمیق تر نسبت به الگوها و فرآیندهای کشف شده به رفع پیچیدگی کمک خواهد کرد؛ بنابراین اگرچه داده کاوی و فرآیند کاوی مشترکات زیادی دارند اما تفاوت های پایه ای بین آن ها در کاری که انجام می دهند و جایی که مورد استفاده قرار می گیرند وجود دارد.

۳-۸ انواع روش های کاوش فرآیند

تکنیک های کاوش فرآیند، بر اساس داده های ثبت شده رویداد، به سه دسته کلی تقسیم بندی می شوند:

- تکنیک های کشف فرآیند^{۴۵}
- تکنیک های چک کردن متابعت^{۴۶}

⁴⁵ Process Discovery

⁴⁶ Conformance Checking

• تکنیک‌های کاوش سازمانی^{۴۷}

دسته اول یا همان تکنیک‌های کشف فرآیند، به‌عنوان ورودی، داده‌های ثبت‌شده رویداد را دریافت کرده و یک مدل بدون استفاده از هیچ اطلاعات پیشینی تولید می‌نمایند.

تکنیک‌های چک کردن متابعت بررسی می‌کنند که آیا فرآیند واقعی که در حال اجرا در سازمان بوده و اطلاعات آن در فایل رویداد ثبت‌شده است منطبق با یک مدل هدف اولیه مدنظر است و بالعکس.

تکنیک‌های دسته سوم هم به این موضوع می‌پردازند که آیا می‌شود با استفاده از سابقه‌ی رخدادهای ثبت‌شده، یک فرآیند را ارتقا یا توسعه داد. به‌عنوان مثال با استفاده از برچسب زمانی در داده‌های ثبت‌شده می‌توان مدل را طوری توسعه داد که گلوگاه‌ها، لایه‌های سرویس و زمان توان عملیاتی را نشان دهد.

۳-۹ یک مثال ساده از کاوش فرآیند در سیستم آموزش

به‌منظور آشنایی بیشتر با بحث کاوش فرآیند، سیستم آموزش را در نظر بگیرید. اکثر دانشگاه‌ها دارای سیستم آموزشی می‌باشند؛ اما در بعضی از مواقع ممکن است چارت درسی به‌صورت مشخص و مدونی در دسترس نباشد. داده‌های سیستم آموزش حاوی اطلاعاتی از قبیل انتخاب واحدهای دانشجویان به همراه درسی که انتخاب کرده‌اند و زمان و تعداد انتخاب‌ها در طول ترم‌های مختلف است. با استفاده از یک سیستم کاوش فرآیند می‌توان از روی این داده‌ها، چارت درسی که به‌نوعی فرآیند تحصیلی یک دانشجو است (کشف فرآیند) را استخراج نمود از طرف دیگر می‌توان با استفاده از تکنیک‌های چک کردن متابعت، میزان پیروی دانشجویان جدید از سرفصل‌های از پیش تعیین‌شده را مشخص نمود و مشخص نمود که آیا دانشجو در انتخاب درس‌ها، پیش‌نیازها و هم‌نیازهای آن درس را رعایت کرده است یا خیر، همچنین با استفاده از تکنیک‌های کاوش سازمانی می‌توان گلوگاه‌های سیستم را شناسایی نموده

⁴⁷ Organizational Mining

و فرآیند را ارتقا داد، مثلاً اینکه دانشجویان که درترم خاصی درس مهندسی نرم‌افزار را پاس می‌کنند موفق‌تر هستند و بر مبنای همین اطلاعات به‌دست‌آمده، چارت درسی را تغییر داد. همچنین می‌توان با استفاده از تکنیک‌ها برخی از رویدادها را پیش‌بینی نموده و بر مبنای آن اقداماتی انجام داد. مثلاً پیش‌بینی کرد که آیا این دانشجوی جدیدالورود قادر هست که درس خود را به پایان برساند و یا چند ترمه به پایان خواهد رساند و سپس بر مبنای این اطلاعات، مشاوره‌هایی را به بعضی از دانشجویان داد.

۳-۱۰ چالش‌های فرآیند کاوی

کاوش فرآیند مهم‌ترین ابزار برای سازمان‌های مدرنی است که نیاز به مدیریت مناسب فرآیندهای عملیاتی دارند. از یک‌سو با رشد باورنکردنی حجم داده روبرو هستیم و از دیگر سو فرآیندها و اطلاعات باید به‌طور مناسبی جمع‌آوری شوند تا نیازمندی‌های مربوط به کارایی، انطباق و سرویس‌های مشتری پاسخ داده شود. علی‌رغم کاربردی بودن کاوش فرآیند، هنوز چالش‌های عمده‌ای پیش رو است که باید موردتوجه قرار گیرد. در ذیل به این چالش‌ها اشاره شده است.

۳-۱۰-۱ چالش اول: یافتن، ادغام کردن و تمیز کردن سابقه‌ی رخدادها

در سیستم‌های فعلی نیز انرژی زیادی باید صرف استخراج سابقه‌ی رخدادهای مناسب برای کاوش فرآیند صورت گیرد. به‌طورمعمول، در این زمینه چند مشکل وجود دارد که باید مرتفع گردد. برخی از این مشکلات عبارت‌اند از:

۱. ممکن است داده‌ها بر روی چندین منبع توزیع شده باشد. این اطلاعات باید ادغام گردند. این مشکل زمانی حادتر می‌شود که از چندین شناسه برای منابع مختلف استفاده شود. مثلاً یک سیستم از نام و تاریخ تولد برای شناسایی افراد استفاده کند و سیستم دیگر از شماره امنیتی اجتماعی فرد.

۲. داده‌های سازمانی غالباً شیء^{۴۸} محور می‌باشند و نه فرآیند محور. به‌عنوان مثال محصولات و ظرف‌ها می‌توانند تگ‌های RFID داشته باشند که خودکار منجر به ثبت رکورد گردند. برای رصد کردن سفارش یک مشتری، این اطلاعات شیء محور باید ادغام و پیش‌پردازش شوند.
۳. سابقه‌ی رخدادها ممکن است ناکامل باشند. یکی از رایج‌ترین مشکلات این است که رویدادها به‌صورت صریح به نمونه‌های فرآیند اشاره نمی‌کنند.
۴. سابقه‌ی رخدادها ممکن است حاوی اطلاعات پرت باشد. منظور از داده‌های پرت نمونه‌هایی است که از الگوی عمومی پیروی نکرده و به‌ندرت اتفاق می‌افتند.
۵. لاگ ممکن است حاوی اطلاعاتی با سطوح مختلف دانه‌دانه شدن باشد. داده‌های لاگ بیمارستانی ممکن است به یک تست خون ساده اشاره کند و یا اینکه به یک رویه پیچیده جراحی اشاره نماید. همچنین برچسب زمانی هم می‌تواند از دقت میلی‌ثانیه تا اطلاعات درشت‌تر نظیر روز (۲۸-۹-۲۰۱۱) را شامل شود.

۳-۱۰-۲ چالش دوم: استفاده از سابقه‌ی رخدادهای پیچیده که ویژگی‌های گوناگونی دارند

داده‌های لاگ ممکن است که ویژگی‌های خیلی متنوعی داشته باشند. بعضی از داده‌های لاگ ممکن است که آن‌چنان بزرگ باشند که رسیدگی به آن‌ها دشوار باشد و بعضی از آن‌ها ممکن است آن‌قدر کوچک باشند که نتوان نتایج قابل اطمینانی از آن‌ها استحصال کرد. ابزارهای موجود در مواجهه با داده‌های با ابعاد پتابایت دشواری‌هایی دارند. در کنار تعداد رکوردهای رویدادهای ذخیره‌شده ویژگی‌های دیگری نظیر متوسط تعداد رویدادها در هر حالت، شباهت میان حالت‌ها، تعداد رویدادهای منحصر به فرد و تعداد مسیرهای واحد نیز هستند که باید مورد توجه قرار گیرند. به‌عنوان مثال فایل لاگ داده L1 با مشخصات ذیل را در نظر بگیرید: ۱۰۰۰ حالت، به‌طور متوسط ۱۰ رویداد به ازای هر حالت. فرض کنید فایل لاگ

⁴⁸ Object

L2 حاوی تنها ۱۰۰ حالت باشد اما هر حالت حاوی ۱۰۰ رویداد باشد و همه رویدادها از یک مسیر واحد تبعیت کنند. پرواضح است که آنالیز L2 به مراتب دشوارتر از آنالیز L1 است، علی‌رغم اینکه هر دو فایل سایز برابر و یکسانی دارند.

از آنجایی که داده‌های لاگ تنها حاوی نمونه‌های مثال می‌باشند، بنابراین نباید این‌طور فرض شود که آن‌ها کامل هستند. تکنیک‌های کاوش متن باید با استفاده از «فرض جهان باز» با این عدم کامل بودن کنار بیایند: این واقعیت که اگر پدیده‌ای اتفاق نمی‌افتد به معنای عدم امکان رخداد آن نیست. این موضوع تعامل با داده‌های لاگ با سایز کم و حاوی تغییرات زیاد را دشوار می‌کند.

همان‌طور که پیش‌تر هم اشاره شد، بعضی از فایل‌های لاگ ممکن است حاوی رکوردهایی با سطح انتزاع بسیار پایین باشند. داده‌های با سطح پایین چندان مطلوب ذینفعان نمی‌باشند؛ بنابراین عموماً سعی می‌شود تا داده‌های سطح پایین با همدیگر تجمیع شوند تا داده‌های با سطح بالاتر تولید گردد. به‌عنوان مثال، زمانی که فرآیند تشخیص و درمان گروهی از بیماران آنالیز می‌شود، احتمالاً دیگر علاقه‌مند به دانستن نتایج آزمایش‌های انفرادی افراد نیستیم.

در این‌گونه از موارد، سازمان‌ها لازم است که از روش سعی و خطا استفاده نمایند تا دریابند که آیا داده‌ها مناسب برای کاوش فرآیند می‌باشند یا خیر؟ بنابراین ابزارها باید سرویس آزمایش امکان‌سنجی سریع برای یک پایگاه داده مشخص را فراهم نمایند.

۳-۱۰-۳ چالش سوم: ایجاد شاخصه‌های ارزیابی

کاوش فرآیند تکنولوژی نوظهوری است. همین امر نشان می‌دهد که چرا نیاز به شاخصه‌های ارزیابی است. به‌عنوان مثال تاکنون ده‌ها تکنیک کشف فرآیند ارائه شده است اما گزارش دقیقی از کیفیت این روش‌ها در دسترس نیست. علی‌رغم اینکه تفاوت‌های زیادی در کارایی و عملکرد این تکنیک‌ها وجود دارد، ارزیابی‌شان کار دشوار و پیچیده‌ای است.

بنابراین نیاز به داده‌های استاندارد و همچنین معیارهای کیفیت مناسب به شدت احساس می‌شود. البته در این زمینه کارهای محدودی انجام شده است. از جمله معیارهای ارزیابی ارائه شده به چهار معیار سازگاری، سادگی، دقت و عمومیت می‌توان اشاره نمود. همچنین سابقه‌ی رخدادهای ثبت شده هم در سایت کاوش فرآیند^{۴۹} موجود است. از یک طرف باید شاخص‌ها بر اساس داده‌های واقعی باشد. از طرف دیگر نیاز به تولید پایگاه داده ترکیبی است که ویژگی‌های خاصی داشته باشد.

۳-۱۰-۴ چالش چهارم: مواجهه با رانش مفهومی^{۵۰}

عبارت رانش مفهومی در حوزه کاوش فرآیند به موقعیتی اشاره می‌کند که در آن فرآیند در عین حال که در حال آنالیز شدن است، تغییر نیز می‌کند. به عنوان مثال در ابتدای یک فایل لاگ، ممکن است که دو فعالیت هم‌زمان باشند در حالی که در ادامه در لاگ این دو فعالیت ترتیبی شوند. فرآیندها به دلایل مختلفی ممکن است تغییر کنند. بعضی از این تغییرات بنا به دلایل تغییرهای دوره‌ای است (مثلاً در ماه مهر و اسفند خریدها بیشتر است و یا اینکه بعد از ظهر جمعه کارمندان کمتری در دسترس هستند). بعضی از تغییرات هم به واسطه تغییر شرایط رخ می‌دهند (مثلاً بازار رقابتی تر می‌شود). این تغییرات بر روی فرآیند تأثیر می‌گذارند و ضرورت دارد که به آن‌ها توجه شود.

پدیده رانش مفهومی با استفاده از شکستن فایل لاگ به قطعات کوچک‌تر و آنالیز رد پای^{۵۱} این قطعات قابل شناسایی است. این آنالیز مرتبه دوم^{۵۲} نیازمند سابقه‌ی رخدادهای بیشتر است. با این حال تعداد کمی از فرآیندها در حالت ثابت می‌باشند و فهم رانش مفهومی دارای اولویت بالایی در مدیریت فرآیندها است؛ بنابراین ابزارها و تحقیقات بیشتری برای آنالیز مناسب رانش مفهومی مورد نیاز است.

⁴⁹ Processmining.org

⁵⁰ Concept Drift

⁵¹ Footprint

⁵² Second Order

۳-۱۰-۵ چالش پنجم: ارتقای پیش فرض‌های نمایشی که در کشف فرآیند استفاده می‌شوند

یک تکنیک کشف فرآیند، با استفاده از یک زبان مشخص (BPMN، Petri Net و...) یک مدل فرآیند تولید می‌نماید. به هر حال مهم است که تجسم نتایج، جدای از نمایی باشد که در کشف فرآیند مورد استفاده قرار می‌گیرد. انتخاب یک زبان هدف، غالباً تعدادی فرض ضمنی را هم دربر می‌گیرد. این فرضیات فضای جستجو را محدود کرده و فرآیندهایی که نمی‌توانند با استفاده از زبان مقصد نمایش داده شوند، کشف نخواهند شد. این بدین معناست که پیش فرض‌های نمایشی که در کشف فرآیند استفاده می‌شوند باید با انتخاب آگاهانه همراه گردند و نباید (فقط) بر مبنای اولویت‌های نمایشی گرافیکی انتخاب شوند.

۳-۱۰-۶ چالش ششم: برقراری تعادل بین معیارهای کیفیت نظیر سازگاری، سادگی، دقت و

عمومیت

غالباً داده‌های ثبت شده کامل نیستند. مدل‌های فرآیندی معمولاً محدودیتی برای تعداد نامحدود نمونه فرآیند (در حالت وجود حلقه‌ها) ندارند. از طرفی، بعضی از نمونه‌ها هم نسبت به سایرین، رخداد به مراتب کمتری دارند؛ بنابراین اینکه فکر کنیم هر نمونه فرآیند در فایل وقایع ثبت شده، موجود است، تصور نادرستی است. برای اینکه نشان داده شود که تصور داشتن داده‌های کامل، در عمل امکان‌پذیر نیست، فرآیندی را در نظر بگیرید که شامل ۱۰ فعالیت بوده و این فعالیت‌ها بتوانند به صورت موازی اجرا شوند. همچنین فرض کنید که فایل رویدادهای ثبت شده حاوی ۱۰,۰۰۰ نمونه فرآیند باشد. تعداد حالت‌های کلی (جایگشت‌ها) در یک مدل با ۱۰ فعالیت هم‌زمان، $10! = 3,628,800$ است؛ بنابراین امکان‌پذیر نیست که تمامی این نمونه‌ها در فایل رویدادهای ثبت شده وجود داشته باشد.

وجود داده‌های نویز (داده‌های با رخداد کم) بر پیچیدگی‌ها می‌افزاید. ساخت مدل برای رفتارهایی که به ندرت رخ می‌دهند (داده‌های نویز) کار بسیار دشواری است. در این گونه موارد، برای پردازش این دسته از رفتارها بهتر است که از چک کردن متابعت استفاده شود [۲۶].

نویز و ناکامل بودن، کشف فرآیند را به یکی از چالشی‌ترین مسائل تبدیل کرده است. تعادل برقرار کردن بین معیارهای سادگی، سازگاری، دقت و عمومیت داشتن کار پر چالشی است. به همین دلیل اکثر تکنیک‌های قدرتمند کاوش فرآیند پارامترهای متنوعی را فراهم می‌سازند. الگوریتم‌های جدیدی برای تعادل برقرار کردن بین این معیارها نیاز است.

۳-۱۰-۷ چالش هفتم: کاوش بین سازمانی

به‌طور سنتی، کاوش فرآیند در یک سازمان اجرا می‌گردد؛ اما با گسترش تکنولوژی وب‌سرویس، یکپارچگی زنجیره تامین و محاسبات ابری، سناریوهایی پیش می‌آید که در آن داده‌های چند سازمان برای آنالیز در دسترس است. در حقیقت دو مشخصه برای کاوش فرآیندهای بین سازمانی موجود است. در سناریوی همکاران، سازمان‌های مختلف همگی با هم در جهت رسیدن به اهداف مشخصی همکاری داشته و نمونه فرآیندها بین این سازمان‌ها در جریان است. در این مدل سازمان‌ها همانند قطعات یک جورچین می‌باشند. فرآیند کلی به قطعاتی شکسته شده و بین سازمان‌ها توزیع می‌شود تا هر سازمان وظیفه مربوط به خود را انجام دهد. آنالیز رویدادهای ثبت‌شده در تنها یکی از این سازمان‌ها کافی نیست. به‌منظور کشف فرآیندهای انتها به انتها، رویدادهای ثبت‌شده سازمان‌های مختلف باید با یکدیگر ادغام گردد که کار ساده‌ای نیست.

سناریوی دوم این است که سازمان‌های مختلف درعین حال که از زیرساخت‌های مشترکی استفاده می‌نمایند، فرآیند یکسانی را اجرا نمایند. به‌عنوان مثال Salesforce.com را می‌توانید در نظر بگیرید. این شرکت فرآیند فروش شرکت‌های دیگر را بر عهده دارد و مدیریت می‌کند. از یک‌طرف شرکت‌ها از زیرساخت این سایت استفاده می‌کنند و از طرف دیگر مجبور نیستند که دقیقاً یک فرآیند قطعی را دنبال کنند چراکه سیستم امکان تنظیمات اختصاصی در دنبال کردن فرآیند به آن‌ها می‌دهد.

واضح است که آنالیز این تغییرات بین سازمان‌های مختلف کار جذاب و جالبی است. این سازمان‌ها می‌توانند از همدیگر یاد بگیرند و فراهم‌کنندگان سرویس ممکن است که سرویس‌هایشان را ارتقا بخشند و سرویس‌های ارزش‌افزوده‌ای را بر مبنای نتیجه کاوش‌های بین سازمانی ارائه نمایند.

۳-۱۰-۸ چالش هشتم: ارائه پشتیبانی عملیاتی

در ابتدا، تمرکز کاوش فرآیند بر روی داده‌های قدیمی (که در پایگاه داده سیستم‌های اطلاعاتی موجود است) بود اما امروزه با گسترش تکنولوژی و افزایش پردازش‌های روی خط، کاوش فرآیند نباید محدود به پردازش‌های برون‌خطی باشد. سه نوع پشتیبانی عملیاتی تعریف شده است: شناسایی، پیش‌بینی، توصیه. زمانی که نمونه‌ای از فرآیند مورد انتظار تخطی می‌کند، می‌تواند شناسایی گردد و سیستم می‌تواند یک اخطار دهد. داده‌های قدیمی می‌تواند به‌منظور تولید مدل پیش‌گوی استفاده گردد. مثلاً می‌توان زمان به اتمام رسیدن یک نمونه را پیش‌گویی کرده و بر اساس آن تصمیماتی اخذ کرد.

استفاده از روش‌های کاوش فرآیند در مدل برون‌خطی، چالش‌های جدیدی را برحسب قدرت محاسباتی و کیفیت داده ایجاد می‌کند.

۳-۱۰-۹ چالش نهم: ترکیب کاوش فرآیند با سایر روش‌های آنالیز

یکی از چالش‌ها نحوه ترکیب روش‌های آنالیز نظیر داده‌کاوی و یا تحقیقات عملیاتی با کاوش فرآیند است. به‌عنوان مثال شبیه‌سازی را در نظر بگیرید. کاوش فرآیند برای یادگیری یک مدل شبیه‌سازی بر مبنای داده‌های قدیمی می‌تواند استفاده شود. متعاقباً، مدل شبیه‌سازی برای پردازش‌های روی خط می‌تواند استفاده شود.

همچنین خیلی مطلوب است که روش‌های کاوش فرآیند را با آنالیزهای تصویری ترکیب نماییم. در پردازش داده‌های بزرگ، آنالیزهای بصری می‌تواند از توانایی انسان برای شناسایی الگوها در داده‌های بدون ساختار استفاده نماید.

۱۱-۳ خوشه‌بندی در کاوش فرآیندها

در کاوش فرآیندها نیز از روش‌های خوشه‌بندی برای تعیین دنباله‌های مشابه یا دسته‌بندی کردن لاگ‌ها بر اساس ویژگی‌های مدنظر استفاده می‌شود. در کاوش فرآیندها می‌توان از الگوریتم‌های مشهور خوشه‌بندی استفاده کرد که در این تحقیق دو الگوریتم کا-میانگین^{۵۳} و آستانه‌ی کیفیت^{۵۴} به تفصیل بررسی خواهد شد اما قبل از آن در مورد معیارهای فاصله^{۵۵} که این روش‌ها از آن‌ها استفاده می‌کنند بحث می‌شود.

۱۲-۳ معیارهای فاصله

تکنیک‌های کلاس‌بندی می‌توانند از چندین معیار فاصله برای محاسبه شباهت بین چند مورد استفاده کنند. نتایج کلاس‌بندی نیز تحت تأثیر انتخاب همین معیارها است. در این بخش به توضیح چند معیار اندازه‌گیری که در خوشه‌بندی مورد استفاده قرار می‌گیرند خواهیم پرداخت [۴].

فرض کنیم هر پرونده شامل یک بردار ویژگی شبیه $\langle i_1, i_2, i_3, \dots, i_n \rangle$ است. حال برای محاسبه‌ی فاصله‌ی بین ۲ پرونده می‌توان از معیارهای فاصله استفاده کرد که در این بخش ۳ تا از معیارهای معروف معرفی خواهد شد که عبارت‌اند از: فاصله‌ی اقلیدسی^{۵۶} [۲۷] و فاصله‌ی همینگ^{۵۷} [۲۸] و فاصله‌ی جاکارد^{۵۸} [۲۹].

⁵³ K-means

⁵⁴ Quality Threshold

⁵⁵ Distance Measures

⁵⁶ Euclidean Distance

⁵⁷ Hamming Distance

⁵⁸ Jaccard Distance

۳-۱۲-۱ فاصله ی اقلیدسی

$$distance(c_j, c_k) = \sqrt{\sum_{l=1}^n |i_{jl} - i_{kl}|^2} \quad (۱-۳)$$

۳-۱۲-۲ فاصله ی همینگ

$$distance(c_j, c_k) = \sum_{l=1}^n \delta(i_{jl}, i_{kl}) / n \quad (۲-۳)$$

$$where \delta(x, y) = \begin{cases} 0 & \text{if } (x > 0 \wedge y > 0) \vee (x = y = 0) \\ 1 & \text{otherwise} \end{cases}$$

۳-۱۲-۳ فاصله ی جاکارد

$$distance(c_j, c_k) = 1 - \left(\sum_{l=1}^n i_{jl} i_{kl} \right) / \left(\sum_{l=1}^n i_{jl}^2 + \sum_{l=1}^n i_{kl}^2 - \sum_{l=1}^n i_{jl} i_{kl} \right) \quad (۳-۳)$$

همان طور که گفته شد معیارهای بالا در الگوریتم های خوشه بندی مورد استفاده قرار می گیرند

که معمولاً فاصله ی اقلیدسی برای ویژگی های عددی بیشتر کاربرد دارد و دو معیار دیگر برای

ویژگی های رشته ای معمول تر می باشند.

۳-۱۳ الگوریتم‌های خوشه‌بندی

الگوریتم‌های خوشه‌بندی به‌طور کلی در مد یادگیری بدون نظارت^{۵۹} می‌باشند که با مجموعه‌ای از داده‌ها که با استفاده از معیارهای شباهات (معیارهای فاصله) بایستی گروه‌بندی شوند نمایش داده می‌شوند.

هر الگوریتم تنها به مجموعه‌ای از ویژگی‌ها که هر شیء را توصیف می‌کند دسترسی دارد و اطلاعاتی در مورد اینکه نمونه باید به چه قسمتی تعلق یابد در دسترس نیست. (منظور برچسب‌گذاری در کلاس‌بندی)

۳-۱۳-۱ کا-میانگین

کا-میانگین یک الگوریتم مشهور برای دسته‌بندی است و در دامنه‌های مختلفی به‌کاربرده می‌شود که می‌توان از قطعه‌بندی تصاویر و بازیابی اطلاعات و ... نام برد.

خوشه‌بندی کا-میانگین یک روش رایج برای تقسیم مجموعه داده‌ها در k گروه به‌صورت خودکار است. این روش علی‌رغم سادگی آن، یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر (مانند خوشه‌بندی فازی) محسوب می‌شود. این روش روشی انحصاری و مسطح محسوب می‌شود [۳۰].

برای این الگوریتم شکل‌های مختلفی بیان شده است. ولی همه‌ی آن‌ها دارای روالی تکراری هستند که برای تعدادی ثابت از خوشه‌ها سعی در تخمین موارد زیر دارند:

- به دست آوردن نقاطی به‌عنوان مراکز خوشه‌ها این نقاط درواقع همان میانگین نقاط متعلق به هر خوشه هستند.
- نسبت دادن هر نمونه داده به یک خوشه که آن داده کمترین فاصله تا مرکز آن خوشه را دارا باشد.

⁵⁹ Unsupervised

ابتدا با انتخاب k مراکز خوشه‌ها به صورت تصادفی انتخاب می‌شوند و با روال تکراری گام‌های زیر گام‌های زیر ادامه پیدا می‌کنند.

۱. هر نمونه‌ی d_i به نزدیک‌ترین خوشه اختصاص داده می‌شود.

۲. مراکز هر خوشه با میانگین داده‌های درون خوشه به‌روزرسانی خواهند شد.

این الگوریتم زمانی به مرحله‌ی همگرایی می‌رسد که هیچ تغییری در اختصاص دادن نمونه‌ها به خوشه‌ها وجود نداشته باشد [۳۱].

حال اگر بخواهیم گام‌های این الگوریتم را بر شمیریم:

۱. در ابتدا K نقطه به‌عنوان به نقاط مراکز خوشه‌ها انتخاب می‌شوند.

۲. هر نمونه داده به خوشه‌ای که مرکز آن خوشه کمترین فاصله تا آن داده را دارا است، نسبت داده می‌شود.

۳. پس تعلق تمام داده‌ها به یکی از خوشه‌ها برای هر خوشه یک نقطه جدید به‌عنوان مرکز محاسبه می‌شود. (میانگین نقاط متعلق به هر خوشه)

۴. مراحل ۲ و ۳ تکرار می‌شوند تا زمانی که دیگر هیچ تغییری در مراکز خوشه‌ها حاصل نشود.

۳-۱۳-۲ آستانه‌ی کیفیت

این الگوریتم برای استفاده در زمینه‌ی بیوانفورماتیک توسعه داده شده است یا به صورت انحصاری تر برای خوشه‌بندی ژن‌های با رفتار مشابه مورد استفاده قرار می‌گیرد. اگرچه از نظر محاسباتی از الگوریتم کا- میانگین گران‌تر است اما قابل پیش‌بینی است. (تضمین می‌کند که بعد از چندین اجرا خوشه‌هایی با اعضای یکسان برمی‌گرداند) و نیازی به مشخص کردن تعداد خوشه‌ها از قبل ندارد.

روش خوشه‌بندی آستانه کیفیت تضمین می‌کند که فاصله‌ی بین دو المنت در یک خوشه باید زیر یک حد آستانه‌ی مشخص باشد. این الگوریتم شامل دو پارامتر ورودی است: اولین پارامتر حد فاصله‌ی مجاز برای خوشه‌ها و دومین پارامتر حداقل سائز خوشه‌ها است که فاصله‌ی مجاز را می‌توان با فرمول (۳-۴) به دست آورد:

$$d = \max_{i,j} \left\{ \sqrt{(x_i - x_j)^2} \right\} \quad (۳-۴)$$

و همچنین سائز خوشه‌ها به معنی تعداد المنت‌های داخل یک خوشه است. گام‌های اصلی الگوریتم آستانه کیفیت در زیر آمده است:

۱. برای هر المنت یک خوشه نامزد یا کاندید در نظر می‌گیریم که این خوشه‌ها با استفاده از روش‌های حریصانه ایجاد می‌شوند؛ یعنی با در نظر گرفتن المنت y خوشه موردنظر شامل تمام عناصر نزدیک به y است که از مقدار آستانه‌ی تعیین‌شده بیشتر نباشد. (با در نظر گرفتن هر داده نزدیک‌ترین داده به آن را انتخاب می‌کنیم این روند ادامه یافته تا از حد آستانه خارج نشود و پس از آن به سراغ المنت بعدی خواهیم رفت)
۲. خوشه‌ی کاندید با بیشترین اندازه به‌عنوان یک خوشه صحیح انتخاب می‌شود و عناصر این خوشه از منبع داده‌ها، برای بررسی‌های بعدی حذف خواهند شد و این خوشه به‌عنوان یک خوشه درست در مجموعه خوشه‌ها قرار می‌گیرد و دوباره عضویت خوشه‌ها به‌روزرسانی می‌شود.
۳. اگر بزرگ‌ترین خوشه‌ی باقی‌مانده سائزی بزرگ‌تر از حداقل سائز در نظر گرفته‌شده برای خوشه‌ها باشد به مرحله‌ی ۲ می‌رویم و در غیر این صورت الگوریتم خاتمه می‌یابد.

الگوریتم آستانه کیفیت خوشه‌های بدون تداخل^{۶۰} ایجاد می‌کند و برخی المنت‌ها که خارج از خوشه‌ها قرار می‌گیرند به‌عنوان داده‌های دورافتاده یا پرت در نظر گرفته می‌شوند و نتیجه خوشه‌بندی ایجاد خوشه‌های با قطر محدود است [۳۲].

۳-۱۴ مقایسه هزینه‌ی زمانی دو الگوریتم آستانه کیفیت و کا-میانگین

برای محاسبه‌ی هزینه‌ی اجرا، این الگوریتم شامل ۳ حلقه تودرتو است. حلقه‌ی بیرونی برای انتخاب خوشه‌های واقعی، حلقه‌ی میانی بر روی خوشه‌های کاندید و حلقه‌ی درونی روی المنت‌ها برای ایجاد خوشه‌های کاندید می‌باشند. پس هزینه‌ی اجرا را می‌توان به‌صورت زیر در نظر گرفت:

$$O\left(\frac{N^2}{L}D\right) \quad (۳-۵)$$

که در آن

- N: تعداد المنت‌ها در مجموعه داده‌هاست
- L: میانگین اندازه‌ی یک خوشه
- D: هزینه‌ی محاسبه‌ی فاصله

اما برای پیچیدگی زمانی الگوریتم کا-میانگین هزینه‌ی زمانی به این شرح است:

$$O(LKNT_{dist}) \quad (۳-۶)$$

که در آن

⁶⁰ Non-overlapping

- L : تعداد تکرارها برای همگرا شدن

- K : تعداد خوشه‌ها

- N : تعداد المنت‌ها

- T_{dist} : زمان محاسبه فاصله بین دو المنت

که با توجه به پیچیدگی زمانی این دو الگوریتم مشخص می‌شود که برای مجموعه داده‌های بزرگ الگوریتم کا-میانگین زمان محاسباتی کمتری (خطی) نسبت به الگوریتم آستانه کیفیت (نمایی) دارد.

فصل چهارم:

شرح مسئله

و

راه حل پیشنهادی

کاوش فرآیندها، یک تکنیک مدیریت فرآیند است که به آنالیز و بررسی فرآیندهای کسب و کار با استفاده از سابقه‌ی رخداد می‌پردازد. در کاوش فرآیندها از الگوریتم‌های اختصاصی داده‌کاوی^{۶۱} برای استخراج دانش از سابقه‌ی رخداد ثبت‌شده توسط سیستم اطلاعاتی استفاده می‌شود. هدف کاوش فرآیندها افزایش کارایی و درک فرآیند، فراهم آوردن تکنیک‌ها و ابزارها برای کشف فرآیند، کنترل داده‌ها و ایجاد یک ساختار جامع از سابقه‌ی رخدادها است.^{۶۲}

هدف در این تحقیق کشف فعالیت‌های^{۶۳} (تراکنش‌ها) نزدیک به هم و قرار دادن فعالیت‌های مشابه در یک خوشه است. به بیان دیگر هدف، بخش‌بندی فعالیت‌های یک فرآیند داخل خوشه‌هایی است که فعالیت‌های آن خوشه به هم شبیه می‌باشند. در این فصل ابتدا ما چند اصطلاح پرکاربرد را در کاوش فرآیندها شرح می‌دهیم سپس به ارائه مثالی از سابقه‌ی رخدادها می‌پردازیم و داده‌های مورد استفاده برای این تحقیق را معرفی می‌کنیم و در نهایت به شرح کار انجام‌شده می‌پردازیم.

۴-۱ مفاهیم و اصطلاحات

۴-۱-۱ فرآیند

روش و ترتیب انجام فعالیت‌های یک نوع پرونده‌ی خاص را فرآیند می‌گوییم. اجرای فرآیند با وجود یک پرونده شروع می‌شود و تا زمانی که کار آن به‌طور کامل به اتمام برسد ادامه پیدا می‌کند. زمانی که پرونده در فرآیند در گردش است را دوره عمر^{۶۴} می‌گویند. هر فرآیند می‌تواند درون خود چندین زیر فرآیند را جای دهد.

⁶¹ Data-mining

⁶² <http://www.processmining.org/>

⁶³ Activity

⁶⁴ Lifecycle

معمولاً رویدادها شامل ویژگی‌هایی با عنوان برچسب زمانی^{۶۵} هستند که زمان وقوع رویداد و حتی زمان پایان آن را نمایش می‌دهند. سابقه‌ی رخداد می‌تواند اطلاعاتی از قبیل اسامی افراد درگیر در هر رویداد را که باعث اجرای یک فعالیت نیز می‌شوند به‌عنوان ایفاکننده‌ی^{۶۶} آن رویداد در خود ثبت کند.

۴-۱-۲ سابقه‌ی رخدادها

برای اعمال تکنیک‌های کاوش فرآیند ضروری است که داده‌ها را از منبع اطلاعاتی از قبیل پایگاه داده‌ها^{۶۷}، گزارش تراکنش‌ها^{۶۸} و ... استخراج کنیم. سابقه‌ی رخدادها معمولاً شامل اطلاعاتی در مورد رویدادهایی که به یک فعالیت یا یک پرونده اشاره می‌کنند می‌شود [۸].

۴-۱-۳ پرونده

پرونده که در مواقعی نمونه‌ی فرآیند^{۶۹} نیز نام‌گذاری می‌شود بخشی از فرآیند است که باید مورد بررسی قرار گرفته و روی آن تراکنش‌هایی اعمال شود. به‌عنوان مثال سفارش یک مشتری، یک درخواست کار، خسارت بیمه، یا مجوز ساخت یا مواردی این‌چنین در کاوش فرآیندها به‌عنوان پرونده شناخته می‌شود. هر پرونده در سیستم با یک شناسه‌ی منحصر به فرد مشخص می‌شود و این شناسه امکان ارجاع به هر پرونده را در دوره‌ی محدود عمرش فراهم می‌کند. هر پرونده در نقطه‌ی شروع هر فرآیند ایجاد می‌شود و در مسیر مشخص‌شده‌ای از فرآیند با انجام فعالیت‌های مشخصی که برای آن تعریف شده است، حرکت می‌کند.

⁶⁵ Timestamp

⁶⁶ Performer

⁶⁷ Databases

⁶⁸ Transaction Logs

⁶⁹ Process Instance

۴-۱-۴ فعالیت

فعالیت که نام‌های دیگر آن وظیفه^{۷۰}، عملکرد^{۷۱} یا رفتار^{۷۲} نیز می‌باشد بخشی از عملیاتی که روی پرونده صورت می‌گیرد است.

۴-۲ نمونه‌ای از سابقه‌ی رخداد یک فرآیند

هدف کاوش فرآیندها استخراج اطلاعات از گزارش سابقه‌ی رخداد در مورد تراکنش‌های موجود در فرآیند است. در سابقه‌ی رخداد یک سری ویژگی‌ها معمولاً برای همه‌ی فرآیندها مشترک هستند و بعضی نیز شامل اطلاعات بیشتر برای هر رویداد می‌باشند. جدول (۴-۱) یک نمونه‌ای از یک لاگ است که شامل ۱۹ رویداد، ۵ فعالیت و ۶ ایفاکننده است.

جدول ۴-۱: نمونه‌ای از سابقه‌ی رخداد یک فرآیند

شماره پرونده	شماره فعالیت	مؤسس	برچسب زمانی
پرونده ۱	فعالیت الف	جان	۹-۳-۲۰۰۴:۱۵.۰۱
پرونده ۲	فعالیت الف	جان	۹-۳-۲۰۰۴:۱۵.۱۲
پرونده ۳	فعالیت الف	سو	۹-۳-۲۰۰۴:۱۶.۰۳
پرونده ۳	فعالیت ب	کارول	۹-۳-۲۰۰۴:۱۶.۰۷
پرونده ۱	فعالیت ب	مایک	۹-۳-۲۰۰۴:۱۸.۲۵
پرونده ۱	فعالیت پ	جان	۱۰-۳-۲۰۰۴:۹.۲۳
پرونده ۲	فعالیت پ	مایک	۱۰-۳-۲۰۰۴:۱۰.۳۴
پرونده ۴	فعالیت الف	سو	۱۰-۳-۲۰۰۴:۱۰.۳۵
پرونده ۲	فعالیت ب	جان	۱۰-۳-۲۰۰۴:۱۲.۳۴
پرونده ۲	فعالیت ت	پیتر	۱۰-۳-۲۰۰۴:۱۲.۵۰
پرونده ۵	فعالیت الف	سو	۱۰-۳-۲۰۰۴:۱۳.۰۵
پرونده ۴	فعالیت پ	کارول	۱۱-۳-۲۰۰۴:۱۰.۱۲
پرونده ۱	فعالیت ت	پیتر	۱۱-۳-۲۰۰۴:۱۰.۱۴
پرونده ۳	فعالیت پ	سو	۱۱-۳-۲۰۰۴:۱۰.۴۴
پرونده ۳	فعالیت ت	پیتر	۱۱-۳-۲۰۰۴:۱۱.۰۳
پرونده ۴	فعالیت ب	سو	۱۴-۳-۲۰۰۴:۱۱.۱۸
پرونده ۵	فعالیت ث	کلیر	۱۷-۳-۲۰۰۴:۱۲.۲۲
پرونده ۵	فعالیت ت	کلیر	۱۸-۳-۲۰۰۴:۱۴.۳۴
پرونده ۴	فعالیت ت	پیتر	۱۹-۳-۲۰۰۴:۱۵.۵۶

⁷⁰ Task

⁷¹ Operation

⁷² Action

با توجه به جدول (۱-۴) هر سطر نمایانگر یک رویداد است که به پرونده موردنظر اشاره دارد (شماره پرونده)، هر رویداد به فعالیتی که باید روی پرونده انجام شود اشاره می‌کند (شماره فعالیت)، همچنین هر رویداد به ایفاکننده یا اجراکننده‌ی فعالیت موردنظر نیز اشاره می‌کند (مؤسس) و درنهایت برچسب زمانی که زمان شروع رویداد را نمایش می‌دهد.

جدول (۱-۴) یک نمونه‌ی ساده‌ای از یک لاگ است که از قسمتی از گزارش‌های ثبت‌شده توسط یک فرآیند، ایجادشده است و همان‌طور که شرح داده شد برای اجرای تکنیک‌های کاوش فرآیند ما به سابقه‌ی رخداد آن فرآیند نیازمندیم تا بتوانیم اطلاعات مفید را با استفاده از تکنیک‌های کاوش فرآیند از آن استخراج کنیم.

۳-۴ معرفی سابقه‌ی رخدادهای مورداستفاده در این تحقیق

در این تحقیق ما از مجموعه داده‌ای^{۷۳} با عنوان Bpi Challenge 2012 استفاده کردیم که این مجموعه داده شامل لاگ‌هایی است که از یک موسسه‌ی مالی هلندی گرفته‌شده است. این لاگ‌ها شامل ۲۶۲۲۰۰ رویداد روی ۱۳۰۸۷ پرونده است. به‌غیراز برخی داده‌های ناشناس یا اشتباه، تمام داده‌ها مربوط به پرونده‌های ورودی به این موسسه‌ی مالی هستند. فرآیند ارائه‌شده در این سابقه‌ی رخداد، روند گرفتن وام شخصی یا اضافه برداشت در یک سازمان مالی جهانی است که هر پرونده ویژگی با عنوان میزان درخواست دارد که مقدار وام یا وجه درخواستی در این ویژگی قرار می‌گیرد. از دیگر ویژگی‌های این مجموعه داده شماره پرونده، نام فعالیت یا تراکنش، زمان و تاریخ شروع رویداد و زمان و تاریخ پایان رویداد و همچنین زمان ثبت رویداد است. این سابقه‌ی رخداد ترکیبی از ۳ زیر فرآیند است که منجر به

⁷³ Dataset

ایجاد این فرآیند بزرگ شده است. در جدول (۴-۲) قسمت کوچکی از این سابقه‌ی رخداد نمایش

داده شده است.^{۷۴}

جدول ۴-۲: لاگ مربوط به فرآیند Bpi Challenge

case	event	startTime	completeTime	REG_DATE	AMOUNT_REQ
173688	A_SUBMITTED	10/1/11 2:08	10/1/11 2:08	10/1/11 2:08	20000
173688	A_PARTLYSUBMITTED	10/1/11 2:08	10/1/11 2:08	10/1/11 2:08	20000
173688	A_PREACCEPTED	10/1/11 2:09	10/1/11 2:09	10/1/11 2:08	20000
173688	A_PREACCEPTED	10/1/11 2:09	10/1/11 2:09	10/1/11 2:08	20000
173688	W_Completeren aanvraag	10/1/11 13:06	10/1/11 13:06	10/1/11 2:08	20000
173688	A_ACCEPTED	10/1/11 13:12	10/1/11 13:12	10/1/11 2:08	20000
173688	O_SELECTED	10/1/11 13:15	10/1/11 13:15	10/1/11 2:08	20000
173688	A_FINALIZED	10/1/11 13:15	10/1/11 13:15	10/1/11 2:08	20000
173688	O_CREATED	10/1/11 13:15	10/1/11 13:15	10/1/11 2:08	20000

مجموعه داده‌ی دیگری که از آن استفاده شده است فرآیندی از یک شرکت تعمیرات لوازم الکترونیک است که روند تعمیر یک دستگاه را در بردارد. این سابقه‌ی رخداد شامل شماره پرونده، نام فعالیتی که باید روی پرونده انجام شود، زمان شروع و پایان انجام فعالیت، شرح کار انجام شده، نوع نقص، تعداد دفعات تعمیر، نوع دستگاه و همچنین رفع عیب است. این لاگ نیز شامل ۱۱۸۵۶ رویداد و ۱۱۰۰ پرونده است که در جدول (۴-۳) قسمت کوچکی از آن را می‌توان مشاهده کرد.

⁷⁴ <https://www.win.tue.nl/>

جدول ۴-۳: قسمتی از لاگ فرآیند تعمیر لوازم الکترونیکی

case	event	Start Time	Complete Time	description	Defect Type	Number Repairs	Phone Type	Defect Fixed
1	Register	1/2/70 14:53	1/2/70 14:53	Simulated process instance				
1	Analyze Defect	1/2/70 14:53	1/2/70 14:53	Simulated process instance				
1	Analyze Defect	1/2/70 15:00	1/2/70 15:00	Simulated process instance	6		T2	
1	Repair (Complex)	1/2/70 15:01	1/2/70 15:01	Simulated process instance	6		T2	
1	Repair (Complex)	1/2/70 15:19	1/2/70 15:19	Simulated process instance	6		T2	
1	Test Repair	1/2/70 15:19	1/2/70 15:19	Simulated process instance	6		T2	
1	Test Repair	1/2/70 15:25	1/2/70 15:25	Simulated process instance	6	0	T2	TRUE
1	Inform User	1/2/70 15:40	1/2/70 15:40	Simulated process instance	6	0	T2	TRUE
1	Archive Repair	1/2/70 15:40	1/2/70 15:40	Simulated process instance	6	0	T2	TRUE

همان‌طور که ملاحظه می‌شود سابقه‌ی رخداد می‌تواند با توجه به فرآیندی که توصیف می‌کند ویژگی‌هایی برای پرونده موردنظر داشته باشد که در سابقه‌ی رخداد دیگر وجود ندارد اما معمولاً در تمام این مجموعه داده‌ها ویژگی‌هایی از قبیل شماره پرونده، نام رویداد، زمان شروع و پایان رویداد وجود دارد. از سابقه‌ی رخداد واقعی دیگری با عنوان سیستم اطلاعات مدیریت جریمه‌های جاده‌ای در این تحقیق استفاده شده است که در این مجموعه داده ویژگی‌هایی مانند شماره پرونده، نام رویداد، زمان شروع و زمان پایان فعالیت، نوع وسیله نقلیه، مهلت پرداخت ثبت شده است که شامل ۱۶۰۰۰۰ رویداد و ۷۸۱۰ پرونده می‌باشد که در این مجموعه داده نیز ویژگی‌های مهم از قبیل شماره پرونده، نام فعالیت و زمان اجرا موجود است.

کار انجام شده در سیستم نما^{۷۵} که شامل فرآیندهایی به صورت شبیه سازی شده می باشد اجرا شده است، سیستم نما در بخش پیوست معرفی شده است. در مورد نتایج کار روی این سیستم در فصل پیاده سازی و نتایج به تفصیل بحث شده است.

۴-۴ روش انجام کار

هدف در این تحقیق یافتن شباهت بین فعالیت ها یا رویدادهای اتفاق افتاده روی هر پرونده است به طوری که تعیین می شود رویدادهایی که روی هر پرونده اتفاق می افتد چقدر می توانند در یک گروه قرار گیرند که اعضای آن گروه از لحاظ برخی ویژگی ها می توانند شبیه به هم باشند. راه حل پیشنهادی برای رسیدن به این هدف و کاری که انجام شده در ادامه شرح داده می شود.

۴-۵ شرح کار انجام شده

تمام سابقه ی رخدادها که مورد بررسی قرار گرفت و حتی آن هایی که بررسی نشد، قطعاً ویژگی هایی با عنوان زمان شروع و زمان پایان یک رویداد برای یک پرونده را دارا می باشند. این ویژگی ها از دو بخش تاریخ و زمان اجرا و پایان رویداد تشکیل می شوند. مشخص است که رویدادهایی که از نظر زمانی به یکدیگر نزدیک می باشند، به نوبت پشت سر هم اجرا شده اند یا فاصله ی کمی در زمان اجرا با یکدیگر دارند. از طرفی تمامی فرآیندها از یک سیکل مرتب پیروی می کنند. به عنوان مثال در فرآیند تعمیرات لوازم الکترونیکی ابتدا نوع نقص دستگاه مورد بررسی قرار گرفته و سپس دستگاه تعمیر می شود، یا اینکه ابتدا دستگاه تعمیر و عیب آن رفع شده و بعد از آن وارد قسمت تست می شود. پس می توان رویدادهایی که نزدیکی زیادی از نظر زمانی به یکدیگر دارند را داخل یک گروه یا خوشه قرار داد. به عنوان مثال در

سیستم نما یک نرم افزار مدیریت فرآیند است که می توان با استفاده از آن کار تحلیل و بررسی فرآیندها را نیز انجام داد ⁷⁵

فرآیند تعمیر لوازم الکترونیکی، بلافاصله بعد از مراجعه مشتری به بررسی عیب آن دستگاه می‌پردازند که می‌توان این دو رویداد را در یک خوشه یا کلاس در نظر گرفت یا اینکه بخش تست که بعد از بخش تعمیر دستگاه است را نیز در یک گروه قرار داد.

برای این منظور ابتدا هر پرونده را به صورت مجزا در نظر گرفته می‌شود یعنی یک پرونده از زمان شروع و قرار گرفتن در جریان فرآیند تا زمانی که کار آن به صورت کامل به اتمام رسیده باشد و از فرآیند خارج شود مورد بررسی قرار می‌گیرد، سپس با استفاده از الگوریتم کا-میانگین فعالیت‌هایی که از نظر زمانی (در نظر گرفتن دو ویژگی^{۷۶} برای هر رویداد که اولی زمان شروع رویداد و دومی زمان پایان رویداد است) به یکدیگر شبیه هستند در یک خوشه قرار می‌گیرند. این کار را برای تمامی پرونده‌های موجود در سابقه‌ی رخداد مربوط به فرآیند تکرار می‌شود زیرا ممکن است مسیر یک پرونده با پرونده‌ی قبل و بعد خود متفاوت باشد که در این صورت نحوه‌ی خوشه‌بندی فعالیت‌ها نیز با خوشه‌بندی فعالیت‌های اتفاق افتاده روی پرونده‌های قبلی و بعدی متفاوت است.

حال به عنوان نتیجه‌ی کار انجام شده با یک مجموعه داده روبه‌رو هستیم که شبیه به سابقه‌ی رخدادهای توضیح داده شده است اما ویژگی جدید را شامل می‌شود که این ویژگی نام خوشه‌ای که فعالیت‌ها در آن قرار دارند را دارا می‌باشد. در فاز دوم روش پیشنهادی مجموعه داده‌ی جدید دوباره با استفاده از الگوریتم کا-میانگین خوشه‌بندی می‌شود، به این صورت که تعداد تکرار هر خوشه برای هر فعالیت به عنوان ویژگی آن فعالیت در نظر گرفته می‌شود. به عنوان مثال فعالیتی با نام Register برای ۱۱۰۰ پرونده مورد بررسی قرار گرفت. نتایج الگوریتم خوشه‌بندی در فاز اول روش پیشنهادی، تعداد ۱۰۳۰ تکرار برای خوشه‌ی شماره یک، ۵۰ تکرار برای خوشه‌ی شماره ۲ و ۲۰ تکرار برای خوشه‌ی شماره ۳ را نشان می‌دهد. حال در مرحله‌ی دوم تعداد خوشه‌های شماره‌ی یک، دو و سه به عنوان ویژگی‌های این فعالیت در نظر

⁷⁶ Feature

گرفته خواهند شد. باید توجه داشت که تعداد ویژگی‌های فاز دوم به تعداد خوشه‌های ایجادشده در فاز اول بستگی دارد که در مثالی که ذکر شد تعداد خوشه‌های ایجادشده در مرحله‌ی اول برابر ۳ بود.

در مرحله‌ی دوم خوشه‌بندی فعالیت‌ها توسط ویژگی‌هایی که در مرحله‌ی قبل برای آن‌ها ساخته شده بود انجام می‌شود که مشخصاً داده‌هایی که از نظر ویژگی‌های داده‌شده نزدیک به هم باشند در یک خوشه یا گروه قرار می‌گیرند. خروجی کار فرآیندی است که فعالیت‌های موجود در آن با توجه به شباهت آن‌ها در زمان اجرا خوشه‌بندی شده است.

این کار، هم برای استخراج اطلاعات مفید از سابقه‌ی رخدادها همانند سایر تکنیک‌های کاوش فرآیند بسیار مفید است هم در امر تقسیم بهینه‌ی منابع^{۷۷} بسیار مفید خواهد بود.

شکل (۴-۱) یک طرح‌واره به صورت فلوچارت را از روش پیشنهادی و کار انجام‌شده به طور خلاصه ارائه می‌دهد.

⁷⁷ این منابع هم می‌توانند از نظر مکانی و زمانی و... مورد بررسی قرار گیرند



شکل ۴-۱: فلوچارت روش پیشنهادی

فصل پنجم:

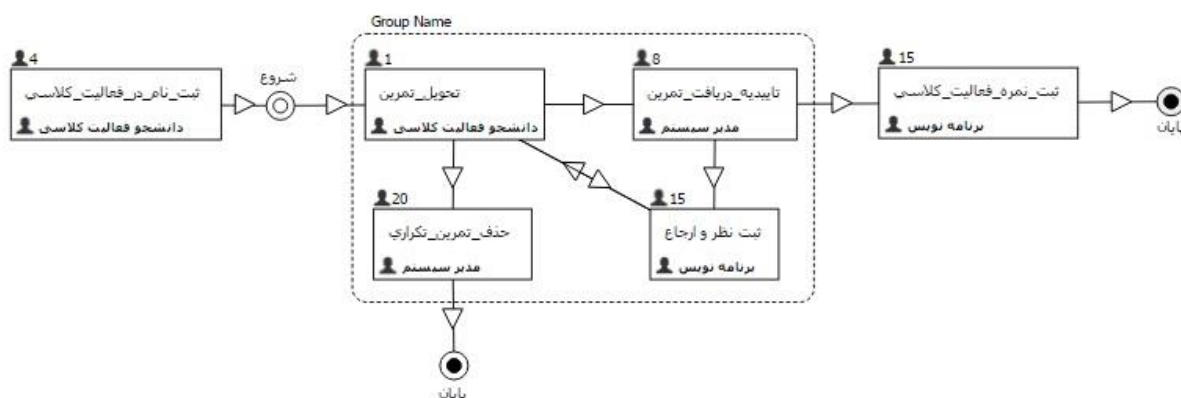
ساده سازی

و نتایج

در این فصل در مورد پیاده‌سازی کاری که در فصل قبل بیان شد بحث می‌شود، به شرح پیاده‌سازی روی قسمت شبیه‌سازی فرآیند در سیستم نما می‌پردازیم، سپس پیاده‌سازی که روی داده‌های واقعی انجام دادیم را بررسی کرده و نتایج را مورد تحلیل قرار می‌دهیم.

۵-۱ روند پیاده‌سازی روی نما

سیستم نما یک سیستم مدیریت فرآیند است که توضیحات کامل‌تر آن در بخش پیوست بیان شده است. در این سیستم قسمتی با نام شبیه‌سازی فرآیند تعبیه شده است، در این قسمت این امکان وجود دارد تا روند فرآیندها را شبیه‌سازی کرد. فرآیندهایی از قبیل فرآیند اطلاعات پروپوزال و پایان‌نامه، اخذ پروژه‌ی کارشناسی، تغییر رشته یا گرایش دانشگاه و سایر فرآیندهای آموزشی در این قسمت موجود است.



شکل ۵-۱: فرآیند فعالیت کلاسی در سیستم نما

شکل (۵-۱) نمایانگر فرآیند فعالیت کلاسی در سیستم نما است که پس از ثبت نام دانشجو در سیستم پرونده در فرآیند جریان پیدا می‌کند و می‌تواند با توجه به شرایط تحت تأثیر فعالیت‌هایی از قبیل تحويل تمرین، تأیید دریافت تمرین حذف تمرین تکراری ثبت نظر و ارجاع یا ثبت نمره قرار گرفته و به مرحله پایان برسد و کار پرونده تمام شود.

در این سیستم امکان دیگری نیز فراهم است که توسط آن قادر هستیم داخل فرآیندها پرونده تزریق کنیم و از مسیری که پرونده‌ها برای رسیدن به نقطه‌ی پایان استفاده می‌کنند مطلع شویم.

برای پیاده‌سازی روش ارائه‌شده در این تحقیق روی سیستم نما ابتدا لازم بود تا یک فرآیند را انتخاب کرده و بعد از آن با استفاده از امکان شبیه‌سازی، چند پرونده داخل فرآیند تزریق شود و بعد از انجام این کار نیاز به ساختن یک سابقه‌ی رخداد داشتیم. به عبارت دیگر کاری که در سیستم نما انجام شد به دلیل نداشتن لاگ کافی دقیقاً برعکس کار انجام شده روی فرآیندهای واقعی با داشتن لاگ و اطلاعات کافی از فرآیند بود. برای برطرف کردن این مورد پس از ایجاد هر پرونده و حرکت آن درون فرآیند، ردپاهای پرونده را در سابقه‌ی رخداد ثبت کردیم. همچنین زمان ورود پرونده داخل یک فعالیت و زمان خروج آن از فعالیت را نیز در سابقه‌ی رخداد قرار دادیم. به بیان دیگر سعی کردیم سابقه‌ی رخداد ساختگی را کاملاً شبیه به یک سابقه‌ی رخداد واقعی طراحی کنیم. تنها تفاوتی که مجموعه داده‌ی ساختگی با مجموعه داده‌ی واقعی داشت، نداشتن تاریخ در برچسب‌های زمانی شروع و پایان بود که دلیل این کار نیز کاملاً مشخص است، چون با یک سیستم شبیه‌ساز طرف هستیم انتظار نمی‌رود که پرونده‌ها مانند یک فرآیند واقعی چندین روز یا ماه در فرآیند جریان داشته باشند و فقط می‌خواهیم که یک دید کلی نسبت به روند حرکتی پرونده‌ها و فعالیت‌های اجراشده روی آن به دست آوریم. اگر تعداد خیلی زیادی پرونده وارد فرآیند در سیستم شبیه‌ساز شود انتظار می‌رود کار تمام پرونده‌ها در یک روز به اتمام برسد.

جدول ۵-۱: سابقه‌ی رخداد ایجادشده به‌وسیله حرکت پرونده در فرآیند

caseid	start	end	time	class	id
721	11:45:03	11:45:03	11:45:03	\N	1
721	11:45:03	11:45:03	11:45:03	\N	73
721	11:45:04	11:45:04	11:45:04	\N	8
927	11:45:05	11:45:05	11:45:04	\N	1
927	11:45:05	11:45:05	11:45:04	\N	143
721	11:45:05	11:45:05	11:45:05	\N	16
304	11:45:06	11:45:06	11:45:05	\N	1
304	11:45:06	11:45:06	11:45:05	\N	73
721	11:45:06	11:45:06	11:45:06	\N	119
721	11:45:06	11:45:07	11:45:06	\N	121
321	11:45:08	11:45:07	11:45:07	\N	1
321	11:45:07	11:45:07	11:45:07	\N	73
304	11:45:07	11:45:07	11:45:07	\N	8
150	11:45:08	11:45:09	11:45:08	\N	1
150	11:45:09	11:45:09	11:45:08	\N	143
927	11:45:10	11:45:10	11:45:09	\N	82

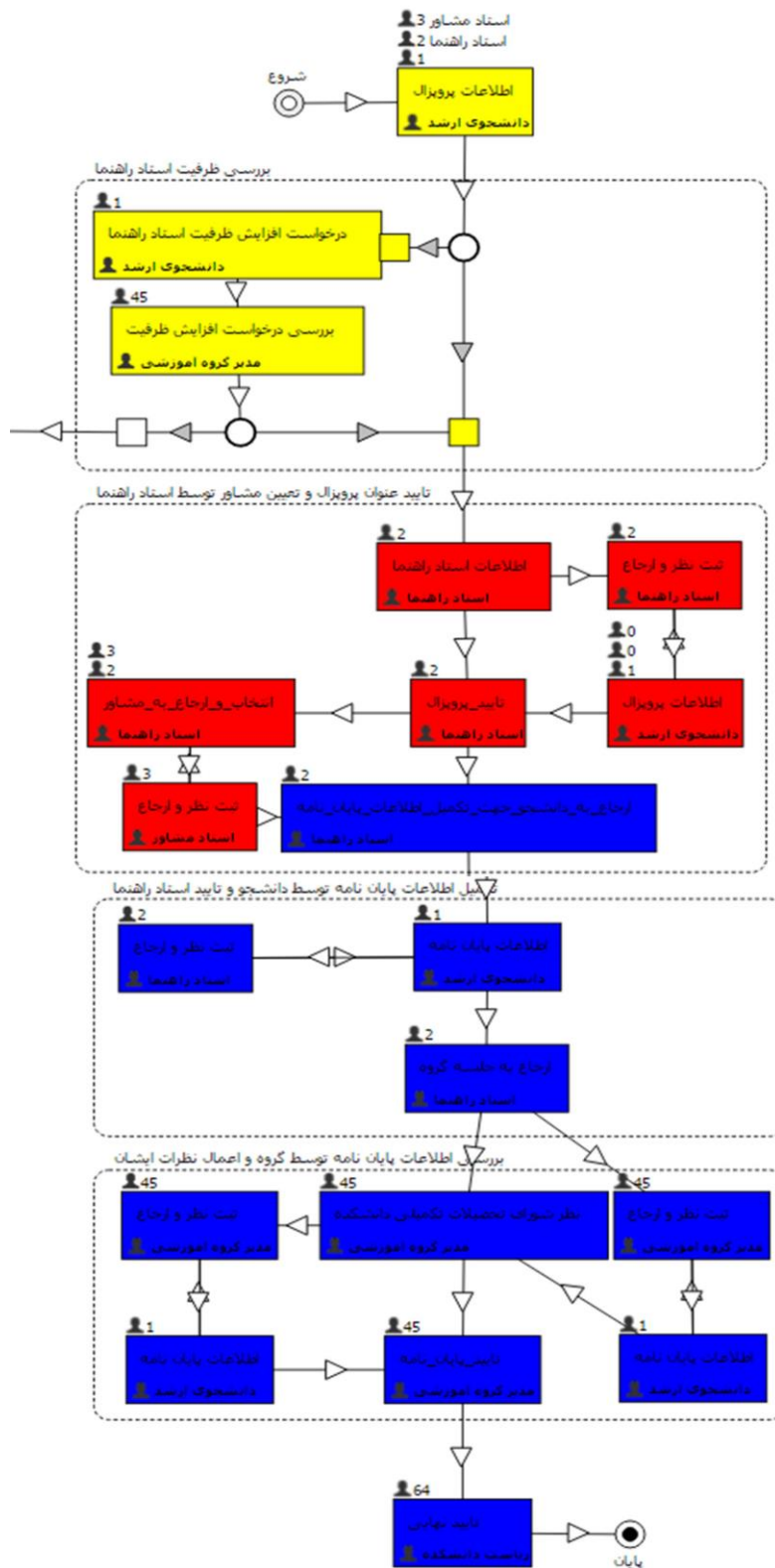
جدول (۵-۱) سابقه‌ی رخداد که با ردیابی پرونده‌ها ساخته شد را نشان می‌دهد که همانند داده‌ی رویداد فرآیندهای واقعی شامل شماره پرونده، نام یا شماره فعالیت (در اینجا از شماره هر فعالیت به‌عنوان نام آن استفاده شده) و همچنین برچسب زمانی است.

کار انجام‌شده که در فصل قبل نیز بیان شد را روی سیستم نما برنامه‌نویسی کرده و بخشی به‌عنوان خوشه‌بندی فرآیند به آن اضافه نمودیم تا بتوانیم عملیات انجام‌شده در هر فرآیند را خوشه‌بندی کنیم. عمل خوشه‌بندی نیز با استفاده از اطلاعات سابقه‌ی رخداد ساخته‌شده در قسمت شبیه‌ساز انجام شد. شکل (۵-۲) فرآیند مربوط به اطلاعات پروپوزال و پایان‌نامه را نشان می‌دهد که با روش پیشنهادی خوشه‌بندی شده است. خط‌چین‌های کشیده شده دور فعالیت‌ها نمایش‌دهنده‌ی خوشه‌هایی است که اپراتور سیستم به‌صورت دستی و با دانش قبلی روی فرآیند اعمال نموده است اما رنگ‌های موجود برای هر فعالیت توسط برنامه‌ی نوشته‌شده و پیاده‌سازی روش ارائه‌شده در این تحقیق ایجادشده است.

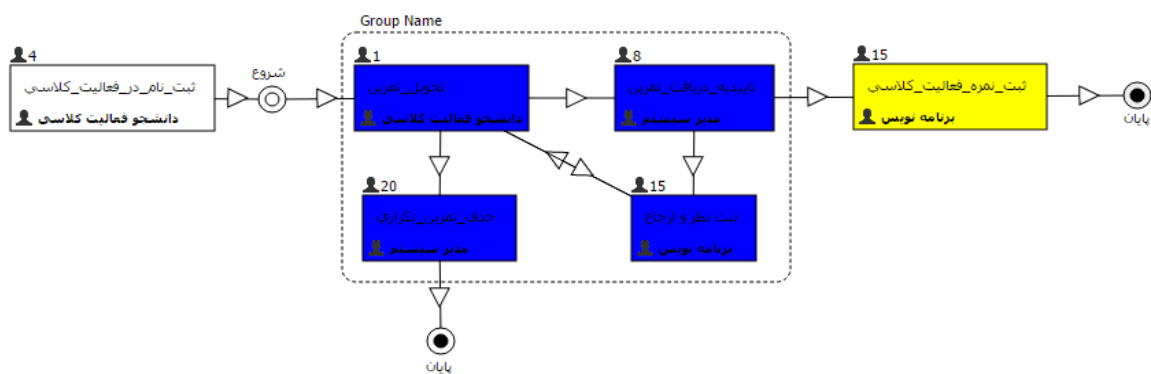
فعالیت‌های با رنگ‌های مشابه به این معنی هستند که این فعالیت‌ها در یک خوشه قرار خواهند گرفت، حال با مقایسه‌ی خوشه‌بندی به صورت دستی و با دانش از فرآیند و همچنین خوشه‌بندی به صورت آماری و توسط روش ارائه‌شده در این تحقیق درمی‌یابیم که این روش تا حد قابل قبولی توانسته است فرآیند را به خوشه‌های منطقی تقسیم کند، با توجه به اختلاف تعداد خوشه‌ها در خوشه‌بندی به صورت دستی و خوشه‌بندی به روش ارائه‌شده دو خوشه‌ی آخر در شکل (۵-۲) که به صورت دستی ایجاد شده‌اند داخل یک خوشه که به صورت خودکار ایجاد شده قرار گرفته است.

شکل (۵-۳) مربوط به فرآیند فعالیت کلاسی است. با توجه به توضیحاتی که در مورد شکل (۵-۲) بیان شد روش ارائه‌شده فعالیت‌های موجود در فرآیند را مطابق خوشه‌بندی دستی و به صورت صحیح خوشه‌بندی کرده است. نکته حائز اهمیت در این شکل قرار نگرفتن فعالیت ثبت‌نام در فعالیت کلاسی در خوشه‌بندی می‌باشد که علت این اتفاق قرار داشتن این فعالیت قبل از نقطه‌ی شروع فرآیند است. دانشجو می‌بایست در فعالیت کلاسی ثبت‌نام کرده باشد تا قادر به ارسال تکالیف خود بشود، پس تمامی فرآیندها پس از نقطه‌ی آغاز خودشان شروع به کار خواهند کرد.

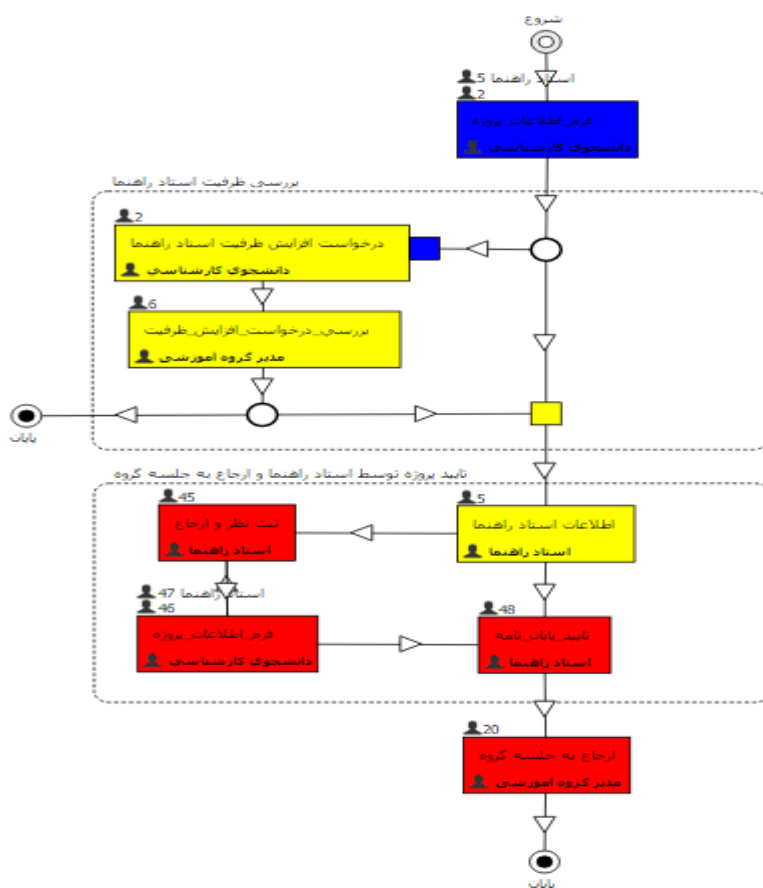
شکل (۵-۴) نتیجه‌ی خوشه‌بندی به روش پیشنهادی روی فرآیند اخذ پروژه‌ی کارشناسی توسط دانشجو را نمایش می‌دهد. روش ارائه‌شده با استفاده از اطلاعات آماری که زمان شروع فعالیت و پایان آن را به عنوان ویژگی‌های مؤثر در نظر می‌گیرد سعی در خوشه‌بندی فرآیندها دارد پس نباید انتظار داشت که خوشه‌بندی به صورت دقیق شبیه به خوشه‌بندی یک عامل انسانی شود، با توجه به شکل (۵-۴) مشاهده می‌شود که روش خوشه‌بندی ارائه‌شده فعالیت‌ها را به صورت منطقی دسته‌بندی کرده و فعالیت خیلی دور یا پرتی را داخل هیچ خوشه‌ای قرار نداده است.



شکل ۵-۲: نتیجه خوشه‌بندی فرآیند اطلاعات پروپوزال و پایان‌نامه



شکل ۳-۵: خوشه بندی فرآیند فعالیت کلاسی



شکل ۴-۵: نتیجه خوشه بندی فرآیند اخذ پروژه دانشجوی کارشناسی

۵-۲ ارزیابی خوشه‌بندی فرآیندهای شبیه‌سازی شده

پس از خوشه‌بندی کردن فرآیندهای شبیه‌سازی شده‌ی آموزشی تصمیم بر آن شد تا برای ارزیابی روش ارائه‌شده، این فرآیندها در اختیار چند کارشناس آموزش قرار گیرد و از آن‌ها خواسته شد با توجه به خوشه‌های ایجادشده به‌صورت دستی و همچنین ارزیابی شخصی، به خوشه‌های تشکیل‌شده توسط روش پیشنهادی امتیازی بین ۱ تا ۱۰ به هر فرآیند بدهند که امتیازهای داده‌شده در جدول (۵-۲) لیست شده است.

جدول ۵-۲: امتیازهای کارشناسان به فرآیندهای آموزشی

نام فرآیند	کارشناس اول	کارشناس دوم	کارشناس سوم	کارشناس چهارم	کارشناس پنجم
پروپوزال و پایان‌نامه	۸	۸	۷	۸	۶
فعالیت کلاسی	۱۰	۹	۹	۱۰	۸
اخذ پروژه‌ی کارشناسی	۹	۸	۹	۸	۷

در جدول (۵-۳) مجموع امتیازهای داده‌شده برای هر فرآیند را محاسبه کرده و درصد شباهت بین خوشه‌ی تشکیل‌شده‌ی دستی و خودکار در این جدول آورده شده است. با توجه به نتایج به دست آمده، این خوشه‌بندی از دید کارشناسان با میانگین ۸۲٫۶٪ از شباهت خوبی با خوشه‌های واقعی موجود در ۳ فرآیند تعریف‌شده در این تحقیق برخوردار است.

جدول ۵-۳: شباهت فرآیندهای خوشه‌بندی‌شده با خوشه‌های واقعی از نظر کارشناسان

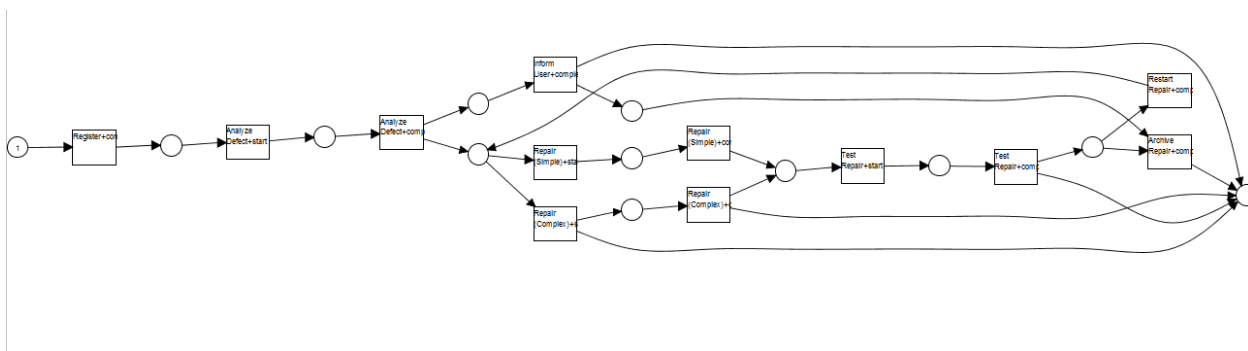
نام فرآیند	پروپوزال و پایان‌نامه	فعالیت کلاسی	اخذ پروژه‌ی کارشناسی
درصد شباهت	۷۴٪	۹۲٪	۸۲٪

۵-۳ پیاده‌سازی برای فرآیندهای واقعی

بعد از پیاده‌سازی روش پیشنهاد شده روی داده‌های شبیه‌سازی شده، نوبت به آن می‌رسد که کار روی داده‌های واقعی و برای فرآیندهای موجود در دنیای واقعی انجام شود. چند داده‌ی رویداد واقعی که

برای پیاده‌سازی از آن‌ها استفاده شده در فصل قبل توضیح داده شد و در مورد اطلاعات موجود در آن‌ها نیز بحث شد. پس‌ازاینکه فرآیند خوشه‌بندی شده و فعالیت‌های مربوط به هر خوشه در آن‌ها مشخص شد برای بیان نتایج به‌صورت طرح‌واره و اینکه درک راحت‌تری نسبت به نتایج به‌دست‌آمده به‌وجود بیاید از الگوریتم کاوش آلفا استفاده شد.

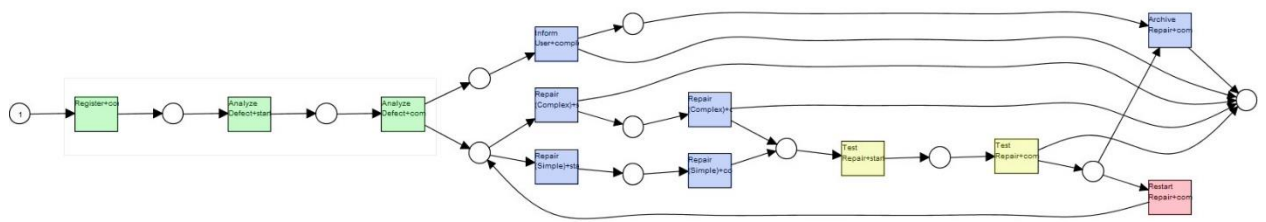
روش کار بدین‌صورت است که با استفاده از نرم‌افزار پرام^{۷۸} (پیوست) و افزونه‌ی کاوش آلفا^{۷۹} که برای کاوش فرآیندها روی آن نوشته‌شده است برای استخراج فرآیند به‌صورت شبکه پتری بهره گرفتیم [۳۳]. سابقه‌ی رخدادهای موجود را به این نرم‌افزار وارد کرده تا خروجی، فرآیند کاوش شده به‌صورت شبکه پتری در اختیار قرار گیرد. شکل (۵-۵) خروجی مربوط به لاگ فرآیند شرکت تعمیرات لوازم الکترونیک به‌صورت شبکه‌ی پتری می‌باشد که توسط الگوریتم آلفا کاوش شده است. بعد از گرفتن خروجی از پرام و با توجه به برجسب‌هایی که برنامه نوشته‌شده برای هر فرآیند زده بود، فعالیت‌های مربوط به یک خوشه را با یک رنگ خاص مشخص نمودیم. شکل (۵-۶) نشان‌دهنده‌ی فرآیند مربوط به داده‌های کمپانی تعمیر لوازم الکترونیکی است که توسط روش پیشنهادی خوشه‌بندی شده است.



شکل ۵-۵: خروجی الگوریتم آلفا مربوط به فرآیند تعمیر لوازم الکترونیک

⁷⁸ Prom

⁷⁹ Alpha Mining



شکل ۵-۶: خوشه‌بندی برای فرآیند تعمیر لوازم الکترونیک

با توجه به نتیجه‌ی به‌دست‌آمده حاصل از خوشه‌بندی، مشاهده می‌شود که گروه‌بندی به‌صورت منطقی صورت گرفته است، به‌این‌ترتیب که فعالیت‌های ثبت درخواست و بررسی عیب دستگاه در یک خوشه (رنگ سبز در شکل ۵-۶)، فعالیت‌های اطلاع به مشتری، انجام تعمیرات ساده، انجام تعمیرات پیچیده و بایگانی کردن دستگاه تعمیر شده در یک خوشه (رنگ آبی در شکل ۵-۶)، فعالیت مربوط به تست در یک خوشه (رنگ زرد در شکل ۵-۶) و درنهایت فعالیت تعمیر دوباره‌ی دستگاه در خوشه‌ی مجزا (رنگ قرمز) قرار گرفتند. نکته‌ی جالب از این خوشه‌بندی قرار دادن فعالیت تعمیر دوباره در یک خوشه‌ی جداگانه بود که پس از بررسی داده‌ی رویداد مربوط به این فرآیند مشخص شد که این فعالیت خیلی کم در لاگ حضور دارد، بدین معنی که دستگاه‌های خیلی کمی پس از گذراندن مرحله‌ی تعمیر و تست احتیاج به دوباره تعمیر شدن داشتند و دلیل این‌که این فعالیت در خوشه‌ی فعالیت تست قرار نگرفت همین نکته بود.

جدول (۴-۵) و جدول (۵-۵) جدول‌های خروجی را برای هر مرحله نشان می‌دهند. جدول (۴-۵) در مرحله‌ی اول روند هر پرونده را دنبال کرده و فعالیت‌ها را بر اساس زمان ورود و خروج در فعالیت خوشه‌بندی می‌کند و به هرکدام یک کلاس اختصاص می‌دهد. به‌عنوان مثال اولین خوشه‌بندی برای پرونده‌ای با شماره‌ی ۱ انجام می‌پذیرد و پس از اتمام خوشه‌بندی و مشخص شدن فعالیت‌های مشابه پرونده‌ی دوم که در شکل با شماره‌ی ۱۰ ظاهر شده است خوشه‌بندی شده و این روند تا زمانی که

خوشه‌بندی فعالیت‌های تمامی پرونده‌ها به پایان برسد ادامه می‌یابد تا فیلد کلاس در مجموعه داده به صورت کامل پر شود.

جدول ۵-۴: خروجی داده‌ای برنامه در فاز اول

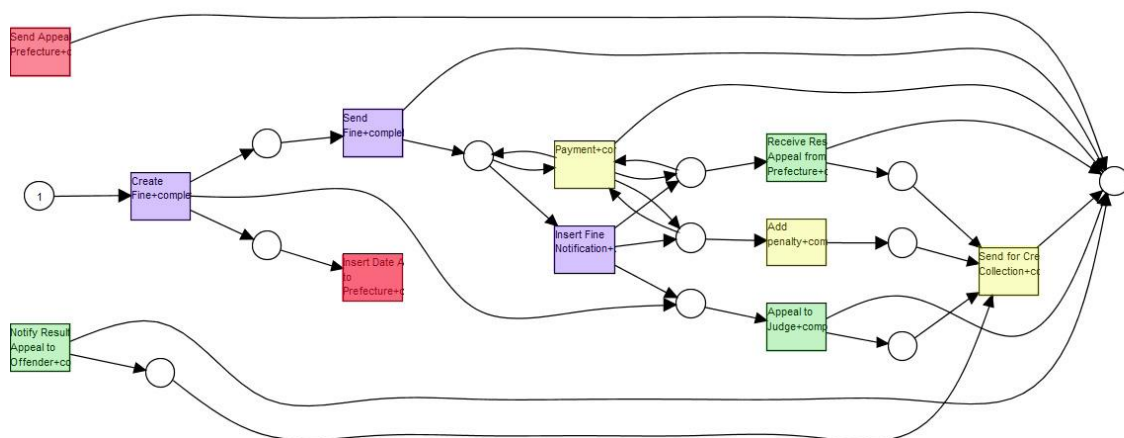
start	end	time	class	id	caseid
1/2/70 14:53	1/2/70 14:53	1/2/70 14:53	\N	Register	1
1/2/70 14:53	1/2/70 14:53	1/2/70 14:53	\N	Analyze Defect	1
1/2/70 15:00	1/2/70 15:00	1/2/70 15:00	\N	Analyze Defect	1
1/2/70 15:01	1/2/70 15:01	1/2/70 15:01	\N	Repair (Complex)	1
1/2/70 15:19	1/2/70 15:19	1/2/70 15:19	\N	Repair (Complex)	1
1/2/70 15:19	1/2/70 15:19	1/2/70 15:19	\N	Test Repair	1
1/2/70 15:25	1/2/70 15:25	1/2/70 15:25	\N	Test Repair	1
1/2/70 15:40	1/2/70 15:40	1/2/70 15:40	\N	Inform User	1
1/2/70 15:40	1/2/70 15:40	1/2/70 15:40	\N	Archive Repair	1
1/1/70 13:39	1/1/70 13:39	1/1/70 13:39	\N	Register	10
1/1/70 13:39	1/1/70 13:39	1/1/70 13:39	\N	Analyze Defect	10
1/1/70 13:45	1/1/70 13:45	1/1/70 13:45	\N	Analyze Defect	10
1/1/70 14:05	1/1/70 14:05	1/1/70 14:05	\N	Repair (Simple)	10

جدول (۵-۵) خوشه‌های ایجادشده در مرحله‌ی قبل را شمارش کرده و این ویژگی در اختیار الگوریتم خوشه‌بندی در مرحله‌ی بعد قرار خواهد گرفت. به عنوان مثال فعالیتی با نام Register دارای ۳ ویژگی می‌باشد که ویژگی اول با نام زرد برابر ۹۷۴، ویژگی دوم با نام آبی برابر ۷۲ و ویژگی سوم با نام قرمز برابر ۵۸ است. این ویژگی‌ها برای تمامی فعالیت‌ها استخراج شده و در اختیار الگوریتم کا-میانگین قرار می‌گیرد و در نهایت با خوشه‌بندی فعالیت‌ها، خروجی، فرآیندی است که برای هر کدام از فعالیت‌های آن دسته‌ای مشخص شده که اعضای آن دسته فعالیت‌هایی هستند که از نظر زمانی به یکدیگر نزدیک می‌باشند.

جدول ۵-۵: نتیجه‌ی نهایی خوشه‌بندی

event	class	counter	label
Register	yellow	974	1
Register	blue	72	1
Register	red	58	1
Analyze Defect	yellow	1424	1
Analyze Defect	red	643	1
Analyze Defect	blue	141	1
Repair (Complex)	red	529	2
Repair (Complex)	blue	861	2
Repair (Complex)	yellow	59	2
Test Repair	blue	2281	3
Test Repair	yellow	75	3
Test Repair	red	660	3
Inform User	blue	734	2
Inform User	red	336	2
Inform User	yellow	32	2
Archive Repair	blue	933	2
Archive Repair	yellow	29	2
Archive Repair	red	38	2

شکل (۷-۵) فرآیند مربوط به سیستم مدیریت جریمه‌های ترافیکی و نتیجه‌ی خوشه‌بندی این فرآیند را نمایش می‌دهد.



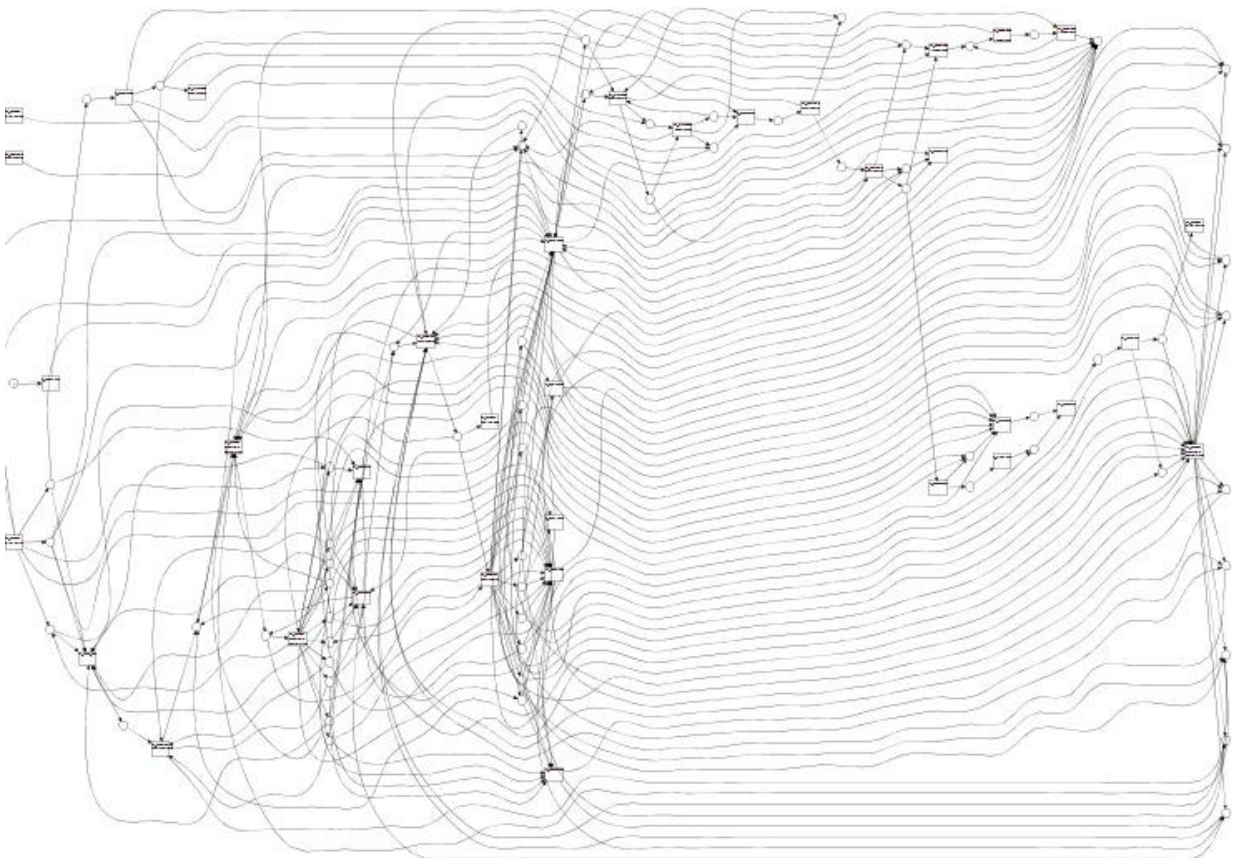
شکل ۷-۵: نتیجه خوشه‌بندی فرآیند جریمه‌های ترافیکی

در این فرآیند، روش ارائه‌شده نتیجه‌ی منطقی از خوشه‌بندی را ارائه می‌کند به‌طوری‌که فعالیت‌های ایجاد جریمه، ارسال جریمه و هشدار دریافت جریمه در یک خوشه (رنگ بنفش در شکل ۷-۵)، فعالیت‌های مربوط به پرداخت، محاسبه‌ی اضافه جریمه و ارسال برای بخش اعتبار سنجی در یک خوشه (رنگ زرد در شکل ۷-۵) همچنین فعالیت‌های مربوط به درخواست تجدیدنظر، دریافت درخواست تجدیدنظر توسط ریاست و اعلام نتیجه‌ی تجدیدنظر به مجرم در یک خوشه (رنگ سبز در شکل ۷-۵) و درنهایت ارسال درخواست به ریاست و درج تاریخ درخواست ارسالی به ریاست نیز در خوشه‌ی جداگانه قرار می‌گیرد (رنگ قرمز در شکل ۷-۵). با توجه به عناوین فعالیت‌های ذکر شده‌ی موجود در هر خوشه مشخص می‌شود که خوشه‌بندی به‌صورت منطقی فعالیت‌های موجود در این فرآیند را تقسیم‌بندی نموده است.

در شکل (۸-۵) فرآیند مربوط به داده‌های Bpi Challenge را مورد بررسی قرارداده‌ایم که در فصل پیش در مورد این داده‌ها به‌طور مفصل بحث شد. همان‌طور که مشاهده می‌شود این فرآیند، یک فرآیند بسیار

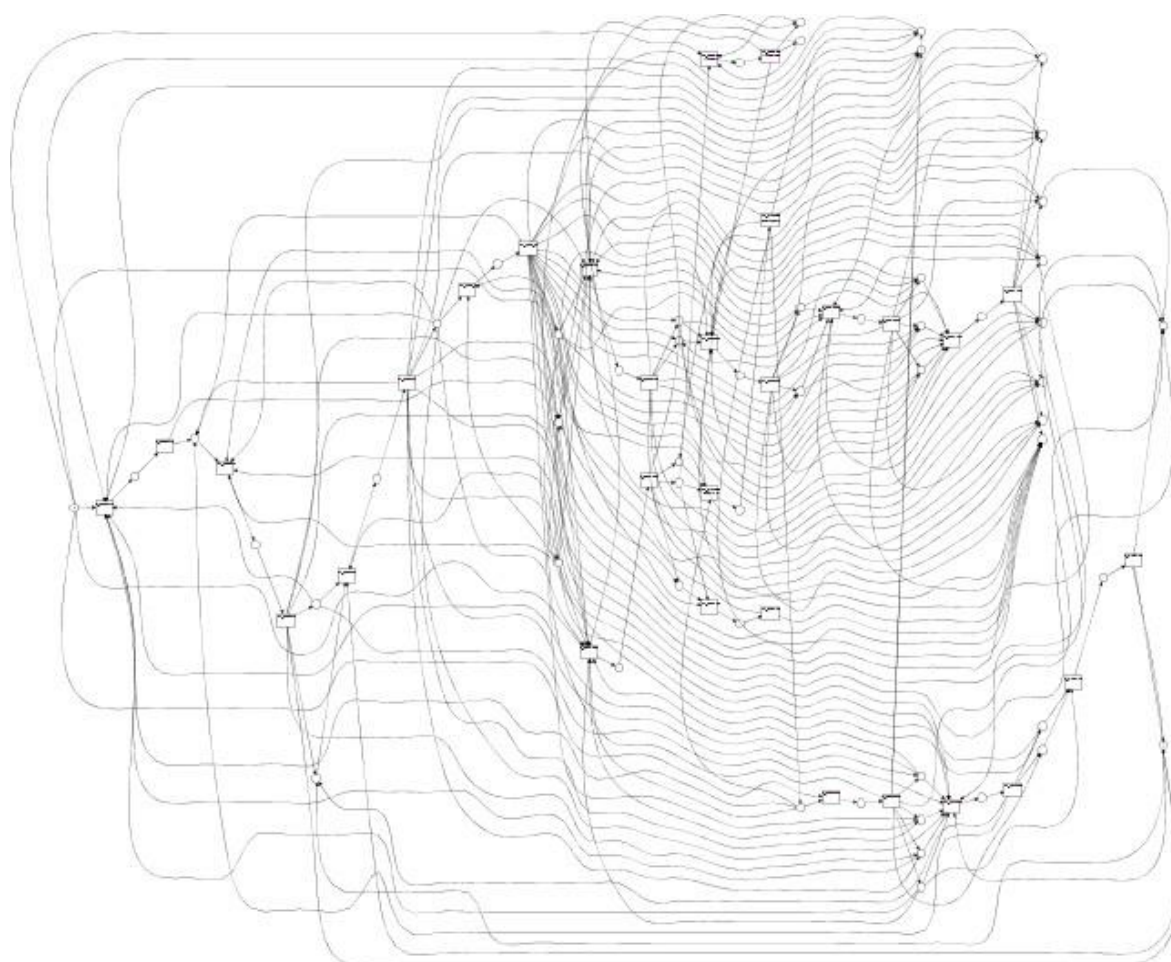
پیچیده است و ارتباط بین فعالیت‌ها به صورت رشته‌های ماکارونی^{۸۰} درهم‌تنیده‌اند که این در هم تنیدگی هم به دلیل تعداد زیاد لاگ‌ها و هم به دلیل ادغام شدن چند فرآیند در یک فرآیند است. پس از این که فعالیت‌های این فرآیند توسط روش پیشنهادی خوشه‌بندی شد، فعالیت‌های موجود حاصل از خوشه‌بندی را که در خوشه‌های یکسان قرار داشتند از لاگ مربوطه جدا کرده و هر کدام به صورت یک لاگ جداگانه به الگوریتم آلفا داده شد تا بتوان بررسی کرد که این خوشه‌بندی در امر ساده‌سازی یک فرآیند نیز می‌تواند کمک‌کننده باشد یا خیر.

شکل‌های (۵-۹، ۵-۱۰، ۵-۱۱، ۵-۱۲) نتایج حاصل از خوشه‌بندی فرآیند شکل (۵-۸) را با استفاده از روش پیشنهادی نشان می‌دهند که این فرآیند بزرگ به چند زیر فرآیند شکسته شده است.



شکل ۵-۸: خروجی الگوریتم آلفا از لاگ Bpi Challenge

⁸⁰ Spaghetti Like



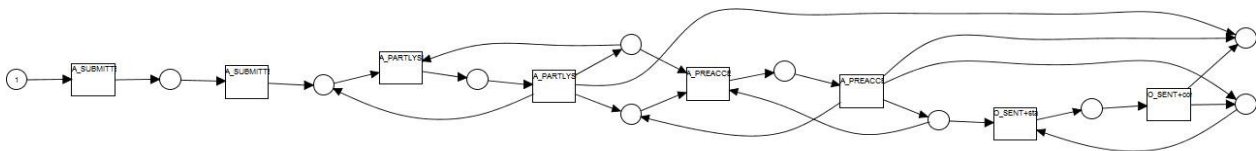
شکل ۵-۹: اولین زیر فرآیند تشکیل شده توسط داده‌های خوشه‌بندی شده

شکل (۵-۹) فرآیند به دست آمده از خوشه‌ی اول است که لاگ‌های مربوط به این خوشه از سابقه‌ی رخداد اصلی جدا و به الگوریتم آلفا داده شده است. همان‌طور که مشخص است این شکل هنوز شامل پیچیدگی فراوانی است که علت آن نیز تعداد فعالیت‌های زیادی است که در این خوشه قرار گرفته است.

شکل (۵-۱۰) فرآیند حاصل از خوشه‌ی دوم است. همان‌طور که مشاهده می‌شود یک فرآیند کوچک از فرآیند اصلی می‌باشد که پس از جدا کردن لاگ‌های مربوط به این فعالیت‌ها از لاگ اصلی و استفاده از

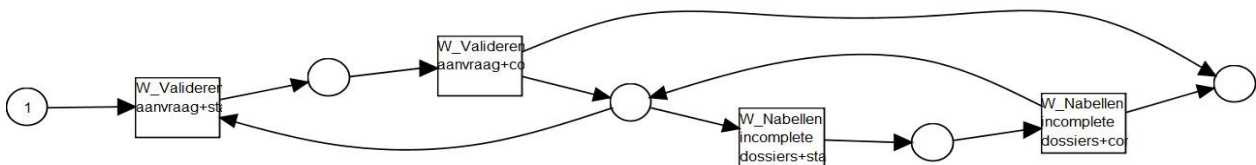
الگوریتم آلفا به این شکل درآمده است. شکل‌های (۵-۱۱، ۵-۱۲) نیز با انجام همین روند به وجود آمده‌اند.

در شکل (۵-۱۰) فعالیت‌های مربوط به ارائه دادن مدارک، بررسی مدارک ناقص، پذیرش اولیه مدارک و ارسال مدارک در یک دسته قرار گرفته‌اند که می‌توان گفت این دسته عملیات مربوط به بررسی مدارک را در خود جای داده است.



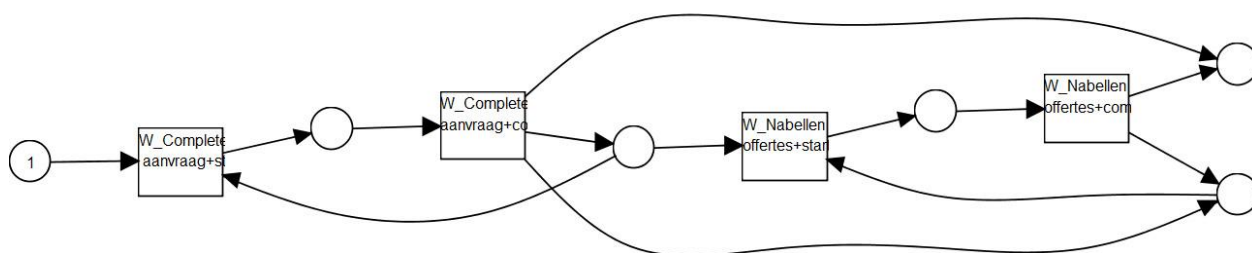
شکل ۵-۱۰: دومین زیر فرآیند تشکیل شده توسط داده‌های خوشه‌بندی شده

در شکل (۵-۱۱) فعالیت‌های اعتبار سنجی پرونده‌ها و اطلاع دادن پرونده‌های ناقص داخل یک دسته قرار گرفته‌اند که به نوعی می‌توان این مجموعه را به فعالیت‌های اعتبار سنجی پرونده در فرآیند درخواست وام مربوط دانست.



شکل ۵-۱۱: سومین زیر فرآیند تشکیل شده توسط داده‌های خوشه‌بندی شده

شکل (۵-۱۲) نیز نشان‌دهنده‌ی این است که فعالیت‌های درخواست تکمیل پرونده‌های ناقص و فرستادن پرونده در مناقصه‌ی دریافت وام در یک دسته قرار گرفته‌اند.

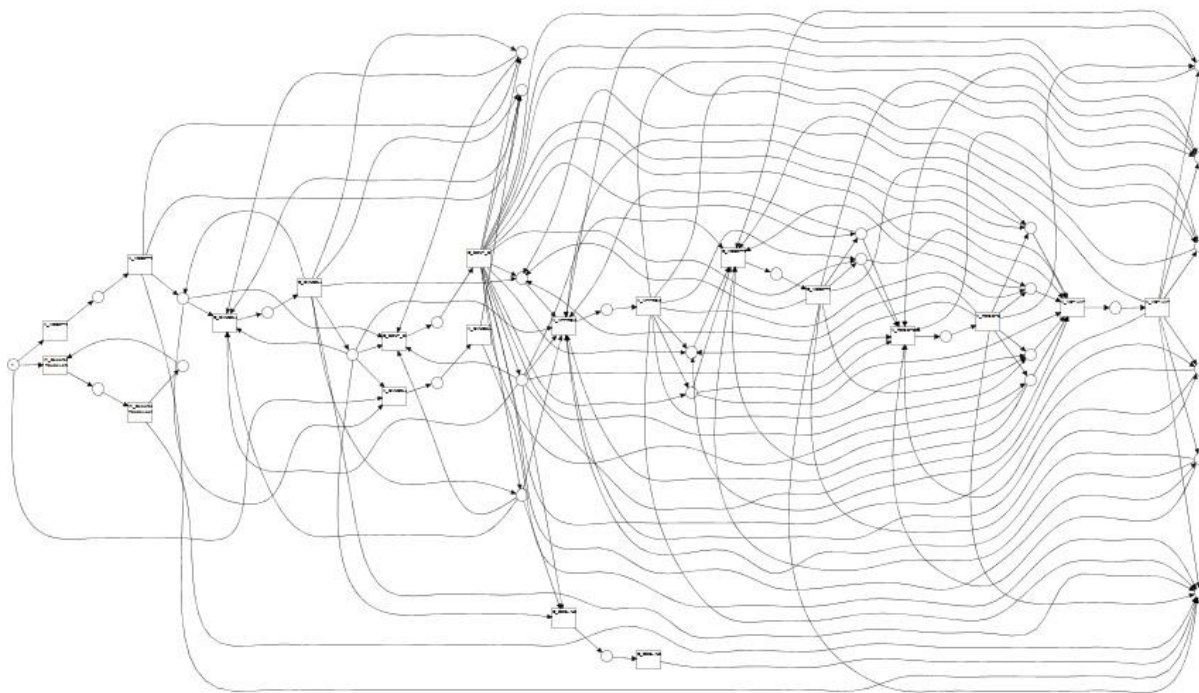


شکل ۵-۱۲: چهارمین زیر فرآیند تشکیل شده توسط داده‌های خوشه‌بندی شده

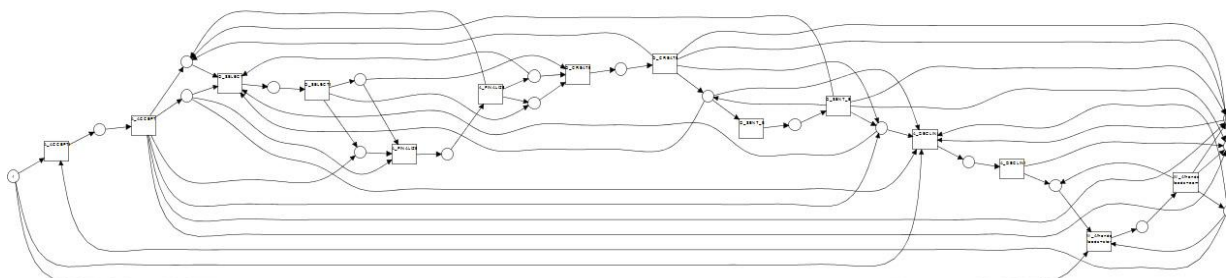
با توجه به این که در خوشه‌بندی انجام‌شده تعداد فعالیت‌های مربوط به خوشه اول زیاد بود، مشاهده شد که فرآیند حاصل از الگوریتم آلفا نیز یک مدل پیچیده از فرآیند را ایجاد کرد به همین دلیل لاگ فرآیند شکل (۵-۹) را دوباره به الگوریتم خوشه‌بندی پیشنهادی دادیم که فعالیت‌ها مجدداً به خوشه‌های مجزا تقسیم شوند که دوباره روند ذکرشده را پیاده کرده و به زیر فرآیندهای شکل (۵-۱۳، ۵-۱۴) رسیدیم.

در شکل (۵-۱۳) فعالیت‌های ثبت‌نام، تأیید، پذیرش، فعال کردن درخواست برای دریافت مدارک و منتفی شدن جریان درخواست وام و بازگرداندن درخواست در یک دسته قرار گرفته‌اند که به‌نوعی قدم‌های اول دریافت وام را شامل می‌شود.

همچنین در شکل (۵-۱۴) فعالیت‌های انتخاب پرونده‌های پذیرش‌شده، نهایی شدن کار برای شروع فرآیند دریافت وام و ایجاد شماره پرونده برای درخواست‌های قبول‌شده در یک دسته قرار گرفته‌اند.



شکل ۵-۱۳: زیر فرآیند اول، نتیجه حاصل از خوشه‌بندی مجدد شکل (۵-۹)



شکل ۵-۱۴: زیر فرآیند دوم، نتیجه حاصل از خوشه‌بندی مجدد شکل (۵-۹)

۵-۴ ارزیابی روش ارائه‌شده روی فرآیندهای واقعی

بعد از خوشه‌بندی به روش پیشنهادی روی فرآیندهای موجود در دنیای واقعی برای ارزیابی عملکرد این روش در میزان صحیح بودن دسته‌های به وجود آمده، نتایج حاصل در اختیار کارشناسان متخصص قرار گرفت و از آن‌ها درخواست شد تا امتیازی بین ۱ تا ۱۰ به خوشه‌های ایجادشده در فرآیندهای مورد بررسی بدهند که نتایج حاصل از این ارزیابی در جدول (۵-۶) آمده است.

جدول ۵-۶: ارزیابی خوشه‌بندی روی فرآیندهای واقعی توسط کارشناسان

نام فرآیند	کارشناس اول	کارشناس دوم	کارشناس سوم	کارشناس چهارم	کارشناس پنجم
تعمیر لوازم الکترونیک	۹	۹	۸	۸	۹
جریمه‌های ترافیکی	۹	۸	۷	۷	۸
درخواست وام	۶	۵	۵	۶	۵

با توجه به جدول (۵-۶) مجموع امتیازهای داده‌شده برای هر فرآیند، محاسبه شده و در جدول (۵-۷) درصد قابل قبول بودن خوشه‌بندی روی این فرآیندها آورده شده است که با محاسبه‌ی میانگین این درصدها به عدد ۷۲,۶٪ خواهیم رسید که امتیاز قابل قبولی می‌باشد.

جدول ۵-۷: درصد قابل قبول بودن روش پیشنهادی در خوشه‌بندی

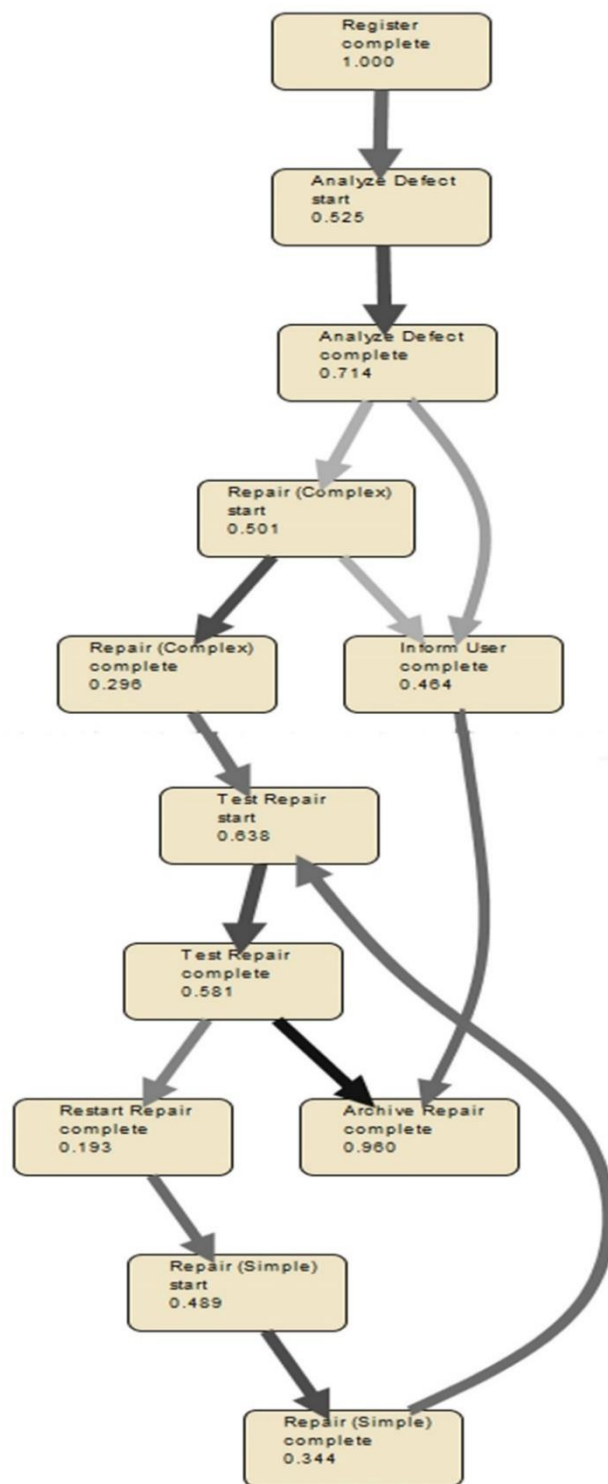
نام فرآیند	تعمیر لوازم الکترونیک	جریمه‌های ترافیکی	درخواست وام
درصد	۸۶٪	۷۸٪	۵۴٪

۵-۵ نتایج به دست آمده روی الگوریتم فازی

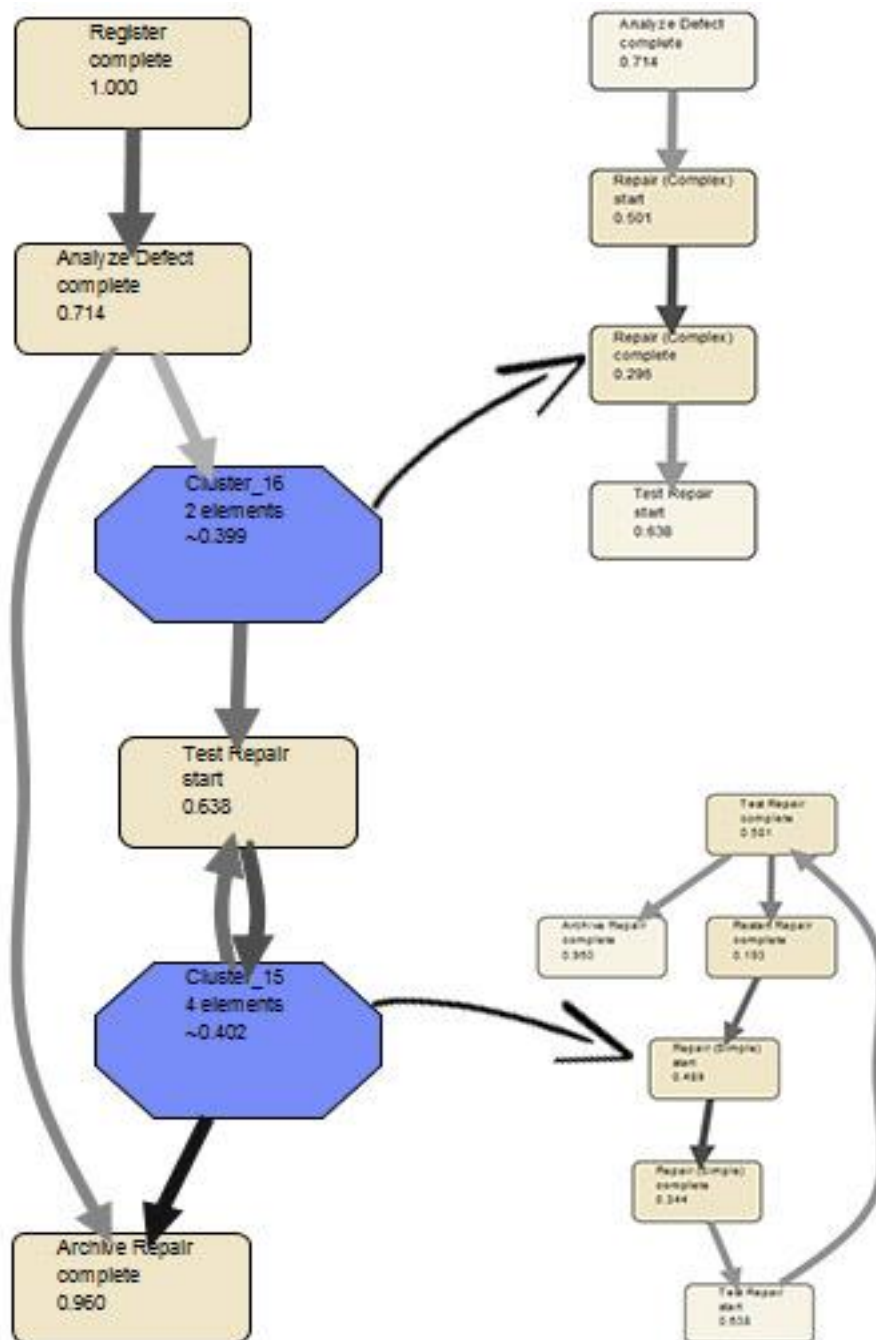
در نرم افزار کاوش فرآیند پرام افزونه‌ی دیگری به نام کاوش فازی وجود دارد که با استفاده از لاگ‌های داده‌شده یک مدل فرآیند ارائه می‌دهد. این مدل فرآیندی با مدلی که در روش آلفا ذکر شد متفاوت است، در روش آلفا مدل خروجی به صورت یک شبکه‌ی پتری در اختیار قرار می‌گیرد اما در روش فازی خروجی به صورت یک گراف همبند و وزن دار است که با وزن این گراف نیز با توجه به داده‌ی رویداد مربوط به فرآیند و پردازش‌هایی که روی این اطلاعات صورت می‌گیرد به وجود می‌آیند. شکل (۵-۱۵) نمونه‌ای از خروجی فرآیند مربوط به کمپانی تعمیر لوازم الکترونیکی با استفاده از کاوش فازی است. داشتن ویژگی فیلتر کردن گره^{۸۱} باعث شد نتایج به دست آمده از خوشه‌بندی روی این مدل نیز نمایش داده شود. اگر به طور خلاصه روش فیلتر گره را توضیح دهیم می‌توان این روش را بدین صورت شرح داد

⁸¹ Node Filtering

که پس از ایجاد مدل (گراف همبند) با استفاده از خواص گراف و وزن یال‌های هر گره می‌توان گره‌هایی که از ارزش کمتری برخوردار هستند را داخل یک خوشه قرار داد و اگر درجه‌ی خوشه‌بندی را بالاتر برده شود حتی گره‌های کم‌اهمیت می‌توانند از این گراف حذف شوند [۹]. روش فیلتر کردن گره در شکل (۵-۱۶) برای فرآیند مذکور نمایش داده شده است. هشت ضلعی‌ها نشان‌دهنده‌ی خوشه‌ها می‌باشند که تعداد آن‌ها با توجه به درجه‌ی خوشه‌بندی می‌تواند کم یا زیاد شود، همان‌طور که در شکل مشاهده می‌شود هر گره از این گراف نمایش‌دهنده‌ی فعالیت‌های موجود در فرآیند و یال‌های گراف نقش ارتباط‌دهنده‌ی بین این گره‌ها را بازی می‌کنند که این اطلاعات نیز از داده‌ی رویداد مربوط به هر فرآیند استخراج می‌شود. در این نوع خوشه‌بندی گره‌های با وزن کمتر داخل خوشه‌هایی قرار می‌گیرند که در هر کدام از خوشه‌ها نیز گره‌های همسایه که وزن آن‌ها از یک حد آستانه کمتر می‌باشد قرار می‌گیرند و بدین ترتیب هر خوشه نیز حاوی یک گراف همبند است. نکته‌ی حائز اهمیت در این روش، حذف گره‌های کم‌اهمیت در زمان خوشه‌بندی می‌باشد، یعنی هنگامی که درجه‌ی خوشه‌بندی از یک حد خاصی بالاتر برود گره‌های کم‌ارزش‌تر حذف خواهند شد.



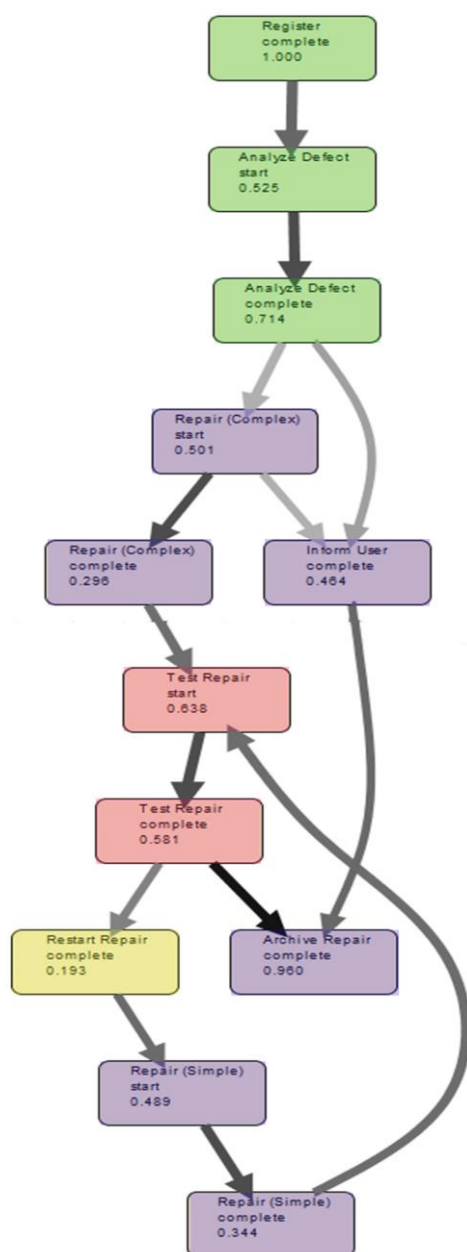
شکل ۵-۱۵: مدل ایجادشده توسط کاوش فازی روی فرآیند تعمیر لوازم الکترونیک



شکل ۵-۱۶: مدل به دست آمده از کاوش فازی پس از اعمال فیلتر کردن گره روی فرآیند تعمیر لوازم الکترونیک

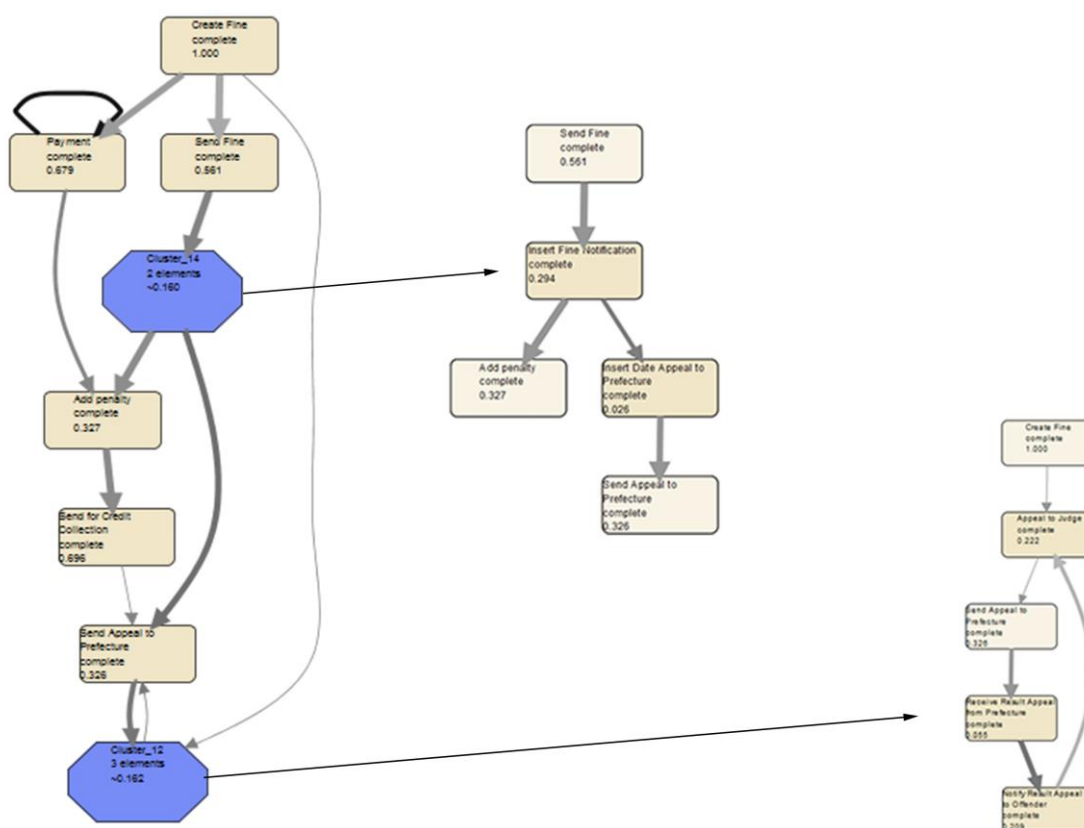
شکل (۵-۱۷) نتیجه خوشه‌بندی به روش پیشنهادی را که روی فرآیند شرکت تعمیر لوازم الکترونیکی می‌باشد، نمایش داده است که در این شکل همان‌طور که مشاهده می‌شود بعد از خوشه‌بندی گره‌ای از فرآیند حذف نخواهد شد و خوشه‌بندی با این روش کاملاً مستقل از وزن گره‌ها و یال‌ها در مدل می‌باشد

که دلیل این کار نیز خوشه‌بندی سابقه‌ی رخداد است نه اطلاعاتی که از داده‌ی رویداد استخراج خواهد شد.

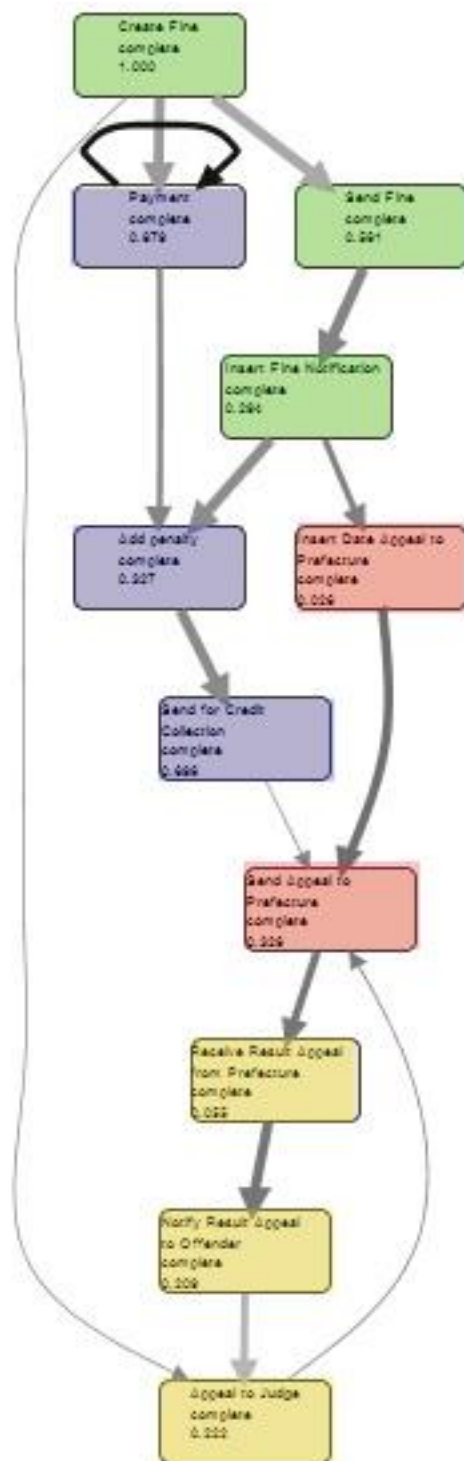


شکل ۵-۱۷: نتیجه حاصل از روش پیشنهادی روی مدل فازی برای فرآیند تعمیر لوازم الکترونیک

شکل های (۵-۱۸، ۵-۱۹) به ترتیب نتیجه‌ی اجرای فیلتر کردن گره و روش پیشنهادی روی فرآیند جریمه‌های رانندگی را نمایش می‌دهند.



شکل ۵-۱۸: نتیجه‌ی فیلتر کردن گره روی فرآیند جریمه‌های ترافیکی



شکل ۵-۱۹: نتیجه‌ی حاصل از روش پیشنهادی روی فرآیند جریمه‌های ترافیکی

۵-۶ تفاوت و مقایسه‌ی روش پیشنهادی با روش کاوش فازی

مبحث فیلترینگ گره در کاوش فازی که نام دیگر آن ادغام و انتزاع گره^{۸۲} می‌باشد، با حذف گره‌های غیرضروری و انتزاعی کردن گره‌هایی که ضرورت کمتری برای مدل دارند سعی در ساده‌سازی مدل می‌کند. ضروری یا غیرضروری بودن گره‌ها با توجه به وزنی که مدل فازی با استفاده از کاوش داده‌ی رویداد برای گره‌ها تعیین می‌کند مشخص می‌شود. پس در اینجا با یک گراف همبند و وزن دار روبه‌رو هستیم. اگر وزن هر گره از یک حد آستانه کمتر باشد و میزان ضرورت آن خیلی پایین باشد می‌توان آن گره و تمام یال‌های وابسته به آن را حذف کرد اما گره‌هایی که بودنشان در مدل ضروری می‌باشد ولی نسبت به سایر گره‌ها از اهمیت کمتری برخوردارند در دسته‌ی گره‌هایی قرار خواهد گرفت که می‌توان آن‌ها را به حالت انتزاعی برد. یا به عبارت دیگر می‌توان آن‌ها را در یک خوشه قرار داد که این کار با بررسی همسایه‌های هر گره که وزنی کم‌تر از حد آستانه‌ی تعیین شده داشته باشند و میزان اهمیت و وزن خوشه‌های همسایه نیز تا حدی نزدیک به گره موردنظر باشد، داخل یک خوشه یا کلاس در مدل قرار می‌گیرند که این خوشه نیز شامل یک گراف همبند می‌باشد [۹].

در روش پیشنهادی در این تحقیق با ضروری یا غیرضروری بودن گره‌ها کاری نداریم. با این استدلال که وقتی در فرآیندی فعالیتی تعبیه شده که پرونده بایستی از آن گذر کند قطعاً فعالیت‌های موجود در هر فرآیند با اهمیت هستند. همچنین در این روش نسبت به مدل ایجادشده با هر روشی (روش کاوش آلفا، کاوش فازی و ...) کاملاً مستقل هستیم زیرا پردازش و خوشه‌بندی بر پایه سابقه‌ی رخداد صورت می‌پذیرد و درواقع با بررسی شباهت زمانی بین فعالیت‌ها می‌توان آن‌ها را در یک خوشه قرارداد، به بیان دیگر نیازی به ایجاد مدل نداریم و درواقع مدل یک راه کمکی برای درک بهتر خوشه‌بندی با روش پیشنهادی می‌باشد. به عنوان مثال در فصل چهارم هنگامی که از این روش برای خوشه‌بندی Bpi Challenge استفاده

⁸² Node Aggregation and Abstraction

شد، داده‌های مربوط به هر خوشه را به کاوش آلفا داده تا بتوان ساده‌سازی فرآیند را با این روش در قالب مدل به نمایش گذاشت.

روش فیلتر کردن گره بسیار متفاوت با روش ارائه شده در این تحقیق است زیرا درواقع در روش فیلتر کردن گره هدف از ایجاد خوشه، بالا بردن سطح انتزاعی مدل می‌باشد اما در روش پیشنهادی هدف قرار گیری فعالیت‌های مشابه در یک خوشه با استفاده از داده‌های آماری است. با این وجود یک مقایسه بین کار انجام شده و روش فیلتر کردن گره ارائه دادیم تا مشخص شود که کدام یک از این روش‌ها به قابل‌فهم‌تر کردن یک فرآیند کمک بیشتری خواهد کرد، به این منظور نتیجه‌های حاصل از روش فیلتر کردن گره و روش پیشنهادی را در اختیار کارشناسان قرار داده و از آن‌ها خواسته شد تا به خوشه‌بندی فعالیت‌ها توسط هرکدام از این روش‌ها امتیازی بین عدد ۱ تا ۱۰ بدهند که نتایج مربوط به فرآیند تعمیر لوازم الکترونیک در جدول (۵-۸)، نتایج مربوط به فرآیند جریمه‌های ترافیکی در جدول (۵-۹) و نتایج مربوط به فرآیند درخواست وام در جدول (۵-۱۰) اعلام شده است.

جدول ۵-۸: مقایسه‌ی روش پیشنهادی با فیلتر کردن گره در فرآیند تعمیر لوازم الکترونیک

روش فیلتر کردن گره	روش پیشنهادی
کارشناس اول	۷
کارشناس دوم	۷
کارشناس سوم	۸
کارشناس چهارم	۶
کارشناس پنجم	۶

جدول ۵-۹: مقایسه‌ی روش پیشنهادی با فیلتر کردن گره در فرآیند جریمه‌های ترافیکی

روش پیشنهادی	روش فیلتر کردن گره	
۹	۶	کارشناس اول
۷	۵	کارشناس دوم
۷	۶	کارشناس سوم
۶	۷	کارشناس چهارم
۸	۶	کارشناس پنجم

جدول ۵-۱۰: مقایسه‌ی روش پیشنهادی با فیلتر کردن گره در فرآیند درخواست وام بانکی

روش پیشنهادی	روش فیلتر کردن گره	
۵	۴	کارشناس اول
۵	۶	کارشناس دوم
۶	۵	کارشناس سوم
۶	۷	کارشناس چهارم
۷	۵	کارشناس پنجم

پس از میانگین گرفتن از امتیازات داده شده به این دو روش به جدول (۵-۱۱) رسیدیم که نتایج این جدول نشان‌دهنده‌ی این است که روش پیشنهادی تقریباً ۱۰٪ از نظر کارشناسان به قابل‌فهم‌تر کردن فرآیندها کمک کرده است که یکی از مهم‌ترین دلایل آن نیز حذف بعضی فعالیت‌های فرآیند برای بالا بردن سطح انتزاعی مدل بود که در روش ارائه‌شده این کار انجام نمی‌شد.

جدول ۵-۱۱: مقایسه‌ی نهایی بین روش پیشنهادی و فیلتر کردن گره

روش فیلتر کردن گره	روش پیشنهادی
۶۰,۶٪	۷۰,۶٪
درصد	

۵-۷ تفاوت و مقایسه‌ی روش پیشنهادی با روش خوشه‌بندی ردپاها

در این روش ابتدا تمام فعالیت‌های موجود در فرآیند در یک جدول قرار می‌گیرند و سپس پرونده‌ها به‌صورت جداگانه مورد بررسی قرار می‌گیرند، به‌گونه‌ای که اگر فعالیت‌ی روی پرونده موردنظر کاری انجام داد در سطر شماره پرونده موردنظر و زیر فعالیت انجام‌شده عدد ۱ قرار می‌گیرد و اگر یک فعالیت در مسیر پرونده کاری روی آن انجام نداد عدد ۰ زیر فعالیت قرار خواهد گرفت این روند ابتدا برای تمام فعالیت‌های انجام‌گرفته یا انجام نگرفته روی یک پرونده انجام‌شده و پس از اتمام بررسی روی یک پرونده تمامی پرونده‌ها به‌صورت جداگانه با این روش مورد بررسی قرار می‌گیرند، ضمن اینکه علاوه بر بررسی روی فعالیت‌های روی یک پرونده، این بررسی روی اجراکننده‌های هر فعالیت نیز صورت می‌پذیرد، با این تفاوت که در بررسی اجراکننده‌ها، تعداد دخالت یک مجری برای یک پرونده مورد بررسی قرار می‌گیرد. جدول (۵-۱۲) نمونه‌ای از انجام این بررسی‌ها برای چند پرونده را نشان می‌دهد.

جدول ۵-۱۲: پیش‌پردازش برای خوشه‌بندی به روش خوشه‌بندی ردپاها

Case id	Activity profile									Originator Profile								
	A	B	C	D	E	F	G	H	I	John	Mike	Sue	Pete	Jane	Fred	Clare	Robert	Mona
1	1	1	0	1	1	1	1	0	1	1	2	2	1	1	0	0	0	0
2	1	1	1	1	1	0	1	0	1	2	0	0	0	0	1	2	1	1
3	1	1	0	1	1	1	1	0	1	1	1	2	2	1	0	0	0	0
4	1	1	1	1	0	0	0	1	1	2	0	0	0	0	1	3	0	0
5	1	1	1	1	1	0	1	1	0	2	0	0	0	0	1	2	2	0
6	1	1	0	1	0	0	0	1	1	1	1	3	0	0	0	0	0	0

پس از انجام این پیش‌پردازش‌ها و آماده شدن دو ویژگی فعالیت و مجری، الگوریتم خوشه‌بندی روی این داده‌ها پیاده‌سازی شده و پرونده‌ها داخل کلاس‌هایی با اندازه‌ی منطقی قرار می‌گیرند که هر کدام از این کلاس‌ها نماینده‌ای از یک زیر فرآیند از فرآیند اصلی می‌باشند [۴].

در روش پیشنهادی هدف خوشه‌بندی فعالیت‌ها بوده و تمرکز روی خوشه‌بندی فعالیت‌ها می‌باشد نه خوشه‌بندی مسیر پرونده‌ها همچنین در داده‌های واقعی در بعضی موارد حجم سابقه‌ی رخداد بسیار زیاد است که پیش‌پردازش ذکرشده در روش خوشه‌بندی ردپاها هزینه‌ی خوشه‌بندی را بسیار بالا می‌برد در صورتی‌که در خوشه‌بندی به روش پیشنهادی نیازی به ساختن ویژگی جدید نیست و از خصوصیات خود سابقه‌ی رخداد استفاده می‌شود، در نتیجه منطقی است که هزینه زمانی در روش پیشنهادی نسبت به هزینه‌ی خوشه‌بندی ردپاها در داده‌های با حجم بزرگ کمتر باشد. زمان انجام هر کدام از این روش‌ها برای فرآیندهای ذکر شده در این تحقیق یادداشت شد که در داده‌های با حجم بالا اختلاف زمان محاسبات در روش پیشنهادی بسیار کمتر از روش خوشه‌بندی ردپاها می‌باشد که نتایج در جدول (۵-۱۳) آورده شده است. البته لازم به ذکر است که این روش‌ها روی یک دستگاه کامپیوتر خانگی و در شرایط پردازش کاملاً مساوی اجرا شده و زمان به‌دست آمده بر همین اساس است. روشن است که هر چه توانایی سیستم پردازشگر بیشتر باشد زمان‌های محاسبه شده کمتر خواهد شد.

جدول ۵-۱۳: مقایسه‌ی زمانی بین روش پیشنهادی با روش خوشه‌بندی ردپاها

روش خوشه‌بندی ردپا	روش پیشنهادی	
۲ ساعت و ۷ دقیقه	۱ ساعت و ۳ دقیقه	فرآیند تعمیر لوازم الکترونیک
۱۶ ساعت و ۵ دقیقه	۱۲ ساعت و ۳۰ دقیقه	فرآیند جریمه‌های ترافیکی
۳۹ ساعت و ۲۰ دقیقه	۳۲ ساعت	فرآیند درخواست وام

فصل ششم:

نتیجہ گیری

ویشہادہ

در فصل‌های قبل کار انجام‌شده را شرح داده و همچنین نتایج حاصل از روش پیشنهادی را به تصویر کشیدیم و به مقایسه‌ی این نتایج با نتایج سایر کارهای مشابه در این زمینه پرداختیم. در این فصل از کار انجام‌شده نتیجه‌گیری می‌کنیم و پیشنهادهایی برای آیندگان که قصد کار در زمینه‌ی کاوش فرآیند را دارند مطرح خواهیم کرد.

۶-۱ نتیجه‌گیری

کاوش فرآیند یک ابزار ارزشمند است که یک بینش واقعی از چگونگی اجراشدن فرآیندها در دنیای واقعی در اختیار می‌گذارد. به‌ویژه اهمیت آن زمانی بیشتر مشخص می‌شود که با محیط‌های انعطاف‌پذیر روبه‌رو باشیم که در آن بازیگران فرآیند و حتی خود صاحبان فرآیند از ساختار دقیق فرآیند بی‌اطلاع هستند. با این حال در این زمینه الگوریتم‌های کاوش فرآیند تمایل به تولید مدل‌های فرآیند پیچیده و غیر ساخت‌یافته دارند که موجب می‌شود از ارزش کمتری برخوردار باشند. یکی از دلایل این محدودیت‌ها و پیچیدگی‌ها وجود پرونده‌ها و فعالیت‌های زیاد می‌باشد که ممکن است روند آن‌ها برای هر پرونده متفاوت از دیگری باشد و به همین دلیل باعث بزرگ شدن حجم سابقه‌ی رخداد خواهد شد.

در این تحقیق یک تکنیک برای دسته‌بندی کردن فعالیت‌های موجود در هر فرآیند ارائه شد که نتایج زیر را به دنبال داشت:

۱. درک فرآیند را آسان‌تر نمود و با شکستن یک فرآیند به فرآیندهای کوچک‌تر موجب کاهش پیچیدگی در فرآیندها شد.
۲. این روش به سازمان‌ها کمک می‌کند تا تقسیم منابع، شامل منابع مکانی و منابع زمانی را به‌صورت بهینه انجام دهند.
۳. تکنیک ارائه شده زمان خوشه بندی برای داده‌های بزرگ را در مقایسه با روش خوشه بندی ردپاها بهبود بخشید.

۴. با توجه به نتایج حاصل شده دقت خوشه‌بندی در روش پیشنهادی نسبت به روش فیلتر کردن گره در روش کاوش فازی حدود ۱۰٪ بهبود پیدا کرد.

۶-۲ پیشنهاد

خوشه‌بندی فرآیندها با هر روشی و برای هر قسمت از فرآیند سعی در ساده‌سازی و قابل‌فهم‌تر کردن فرآیندها دارد و برای رسیدن به این منظور از اطلاعات موجود در سابقه‌ی رخدادهای مربوط به فرآیند استفاده می‌کنند. خوشه‌بندی به روش پیشنهادی نیز با دسته‌بندی کردن فعالیت‌های مشابه قصد بهینه‌سازی در فضا و زمان را برای رسیدگی به پرونده‌ها دارد.

در این بخش پیشنهاداتی برای آیندگان که قصد کار در زمینه‌ی کاوش فرآیند را دارند مطرح می‌شود:

۱. پیشنهاد می‌شود پس از مشخص شدن فعالیت‌های مشابه و با استفاده از تکنیک‌های کاوش متن برای هر خوشه‌ی تشکیل‌شده عنوانی انتخاب کرد تا در قابل‌فهم‌تر کردن یک فرآیند تأثیر دوچندانی حاصل شود. به‌عنوان مثال در فرآیند تعمیر لوازم الکترونیک پس از خوشه‌بندی فعالیت‌هایی از قبیل تعمیر ساده، تعمیر پیچیده و آرشو دستگاه تعمیر شده در داخل یک خوشه، عنوان یا برچسب به نام بخش تعمیرات به خوشه‌ی ایجادشده نسبت داده شود.

۲. روش پیشنهادی در این تحقیق، با استفاده از اطلاعات موجود در لاگ، فرآیندها را دسته‌بندی می‌کرد که با کمک این روش و محاسبه‌ی شباهت خوشه‌ها با فرآیندهای مشابه که فاقد لاگ هستند، می‌توان به زیرفرآیندهای آنها نیز دست پیدا کرد یا به بیان دیگر بدون داشتن لاگ آنها را دسته‌بندی کرد.

1. Cook, J.E. and A.L. Wolf, *Discovering models of software processes from event-based data*. ACM Transactions on Software Engineering and Methodology (TOSEM), 1998. 7(3): p. 215-249.
2. Weijters, A. and W.M. van der Aalst. *Process mining: discovering workflow models from event-based data*. in *Belgium-Netherlands Conf. on Artificial Intelligence*. 2001. Citeseer.
3. Cook, J.E. and A.L. Wolf, *Software process validation: quantitatively measuring the correspondence of a process to a model*. ACM Transactions on Software Engineering and Methodology (TOSEM), 1999. 8(2): p. 147-176.
4. Song, M. C.W. Günther, and W.M.P. van der Aalst, *Trace Clustering in Process Mining*, in *Business Process Management Workshops: BPM 2008 International Workshops, Milano, Italy, September 1-4, 2008. Revised Papers*, D. Ardagna, M. Mecella, and J. Yang, Editors. 2009, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 109-120.
5. Luengo, D. and M. Sepúlveda. *Applying clustering in process mining to find different versions of a business process that changes over time*. in *International Conference on Business Process Management*. 2011. Springer.
6. Greco, G. et al. *Discovering expressive process models by clustering log traces*. IEEE Transactions on Knowledge and Data Engineering, 2006. 18(8): p. 1010-1027.
7. Rozinat, A. and W.M. van der Aalst, *Conformance checking of processes based on monitoring real behavior*. Information Systems, 2008. 33(1): p. 64-95.
8. van der Aalst, W.M. et al. *Business process mining: An industrial application*. Information Systems, 2007. 32(5): p. 713-732.
9. Günther, C.W. and W.M. Van Der Aalst. *Fuzzy mining—adaptive process simplification based on multi-perspective metrics*. in *International Conference on Business Process Management*. 2007. Springer.
10. Van der Aalst, W.M. et al. *Workflow mining: a survey of issues and approaches*. Data & knowledge engineering, 2003. 47(2): p. 237-267.
11. Van der Aalst, W. T. Weijters, and L. Maruster, *Workflow mining: Discovering process models from event logs*. IEEE Transactions on Knowledge and Data Engineering, 2004. 16(9): p. 1128-1142.
12. Cook, J.E. and A.L. Wolf. *Automating process discovery through event-data analysis*. in *Proceedings of the 17th international conference on Software engineering*. 1995. ACM.

13. Agrawal, R. D. Gunopulos, and F. Leymann. *Mining process models from workflow logs*. in *International Conference on Extending Database Technology*. 1998. Springer.
14. Herbst, J. and D. Karagiannis. *Integrating machine learning and workflow management to support acquisition and adaptation of workflow models*. in *Database and Expert Systems Applications, 1998. Proceedings. Ninth International Workshop on*. 1998. IEEE.
15. Greco, G. A. Guzzo, and L. Pontieri. *Mining hierarchies of models: From abstract views to concrete specifications*. in *International conference on Business Process management*. 2005. Springer.
16. Van Dongen, B.F. and W.M. Van der Aalst. *Multi-phase process mining: Building instance graphs*. in *International Conference on Conceptual Modeling*. 2004. Springer.
17. Günther, C.W. and W.M.P. van der Aalst, *Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics*, in *Business Process Management: 5th International Conference, BPM 2007, Brisbane, Australia, September 24-28, 2007. Proceedings*, G. Alonso, P. Dadam, and M. Rosemann, Editors. 2007, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 328-343.
18. Jain, A.K. M.N. Murty, and P.J. Flynn, *Data clustering: a review*. ACM computing surveys (CSUR), 1999. **31**(3): p. 264-323.
19. Van Dongen, S. *Graph clustering by flow simulation [phd thesis]*. [Utrecht (The Netherlands)]: University of Utrecht. 2000.
20. Ferreira, D. et al. *Approaching process mining with sequence clustering: Experiments and findings*. in *International Conference on Business Process Management*. 2007. Springer.
21. Jeston, J. J. Nelis, and T. Davenport, *Business Process Management: Practical Guidelines to Successful Implementations*. 2008, NV. USA: Butterworth-Heinemann (Elsevier Ltd).
22. Tsai, A. et al. *EPC Workflow model to WIFA model conversion*. in *2006 IEEE International Conference on Systems, Man and Cybernetics*. 2006. IEEE.
23. Ferdian, A. and A. Weinburg, *Comparison of event-driven process chains and UML activity diagram for denoting business processes*. Harburg Technische Universität HamburgHarburg, 2001: p. 1-42.
24. Curtis, B. M.I. Kellner, and J. Over, *Process modeling*. Communications of the ACM, 1992. **35**(9): p. 75-90.
25. Petri, C.A. and W. Reisig, *Petri net*. Scholarpedia, 2008. **3**(4): p. 6477.
26. Rozinat, A. and W.M. van der Aalst. *Conformance testing: Measuring the fit and appropriateness of event logs and process models*. in *International Conference on Business Process Management*. 2005. Springer.
27. Duda, R. and P. Hart, and D. Stork, *Pattern classification*. 2000, Wiley-interscience.
28. Hamming, R.W. *Error detecting and error correcting codes*. Bell System technical journal, 1950. **29**(2): p. 147-160.

29. Tan, P.-N. M. Steinbach, and V. Kumar, *Introduction to data mining. Inc.* 2005, Boston, MA, USA: Addison-Wesley Longman Publishing Co.
30. Alpaydin, E. *Introduction to machine learning.* 2004. Cambridge, Massachusettes: The MIT Press Google Scholar.
31. Wagstaff, K. et al. *Constrained k-means clustering with background knowledge.* in *ICML*. 2001.
32. Bednarik, L. and L. Kovacs, *Efficiency analysis of quality threshold clustering algorithms.* Production Systems and Information Engineering, University of Miskolc, 2012. **6**: p. 1785-1270.
33. de Medeiros, A.A. et al. *Process mining: Extending the α -algorithm to mine short loops.* 2004, BETA working paper series, WP 113, Eindhoven University of Technology, Eindhoven.

پوست:

سیستم نما:

این سیستم به عنوان یک سیستم BPMS به شمار می رود و نرم افزاری آماری است که کار نمایش و نظارت روی فرآیندها را آسان می سازد. این سیستم قادر به پشتیبانی نظارت و کنترل روی فرآیندها می باشد و در صورتی که منابع هر سازمان یا سیستم و حالت های فعلی در آن ها مشخص باشد قادر به آنالیز سیستم است که این کار را با استفاده از رخدادهای ثبت شده انجام می دهد. این نرم افزار مربوط به شرکت گنجینه ی اطلاعات و ارتباطات بوده و با اجازه ی این شرکت در این تحقیق برای انجام تحلیل و بررسی فرآیندهای شبیه سازی شده و همچنین آشنایی با یک سیستم BPMS مورد استفاده قرار گرفت.

پرام:

نرم افزار پرام یک قالب کاری قابل تعمیم است که از تکنیک های کاوش فرآیند در طیف گسترده ای در قالب افزونه ها پشتیبانی می کند. پرام یک پلتفرم مستقل است که با زبان جاوا پیاده سازی شده و قابل دانلود به صورت رایگان می باشد. همچنین تمام افزونه های مربوط به این نرم افزار را می توان به صورت پکیج های جداگانه دانلود نمود. تمام اطلاعات مربوط به این نرم افزار در سایت توسعه دهندگان آن

قابل دسترسی است. <http://www.promtools.org>

Abstract:

Business process management is a structural and systematic procedure to analyze, develop, control and manage the process, aimed to improve the quality of services. The absence of an efficient business system causes to create unnecessary costs and low-quality services, and also discontent for staff and clients in each organization. Business process management is a comprehensive solution for understanding and managing the process in any organization and helps them to make complex relationships work between staff, systems and process, simple in a way that it will be understandable and in this way it helps, to increase efficiency of organizations and clients' satisfaction.

In this research business process management, process mining and process mining's algorithms will be examined. Then some processes will be explored with the aim of exploring the relationships between them, so it can identify similar activities of a process that can be placed in a class or a group. In this step extract event logs based on, time of starting transaction and when the transaction is complete, and in the next step clustering and finding hidden relations in this event log will be investigated.

The results of implementation of the real and simulated processes indicate that clusters which produced by the proposed method was able to maintain the similarity of the clusters which produced by humans in educational simulated processes with an average of 85.5% and also the accuracy of 73.3% in the real processes. Also, as far as the experts in this field are concerned the proposed method is almost 10% accurate than node filtering in the fuzzy mining. In addition, the processing time of the proposed method compared with the trace clustering method is much improved.

Key words: Business process management, process mining, event logs, clustering



Shahrood University of Technology

Faculty of computer

MSc Thesis artificial intelligence and robotics

**Business Processes Activities Segmentation based on Statistical
Analysis**

By: Peyman Dehghan

Supervisor:

Dr. M. Zahedi

september 2016