بسم الله الرحمن الرحيم

دانشکده مهندسی کامپیوتر

پایان‌نامه دکتری هوش مصنوعی

یک چارچوب احتمالاتی برای کنترل زمانی رویدادها در فرآیندهای سازمانی با منابع مشترک

نگارنده: ایمان فیروزیان

استاد راهنما

دکتر مرتضی زاهدی

استاد مشاور

دکتر حمید حسن‌پور

دی ۱۳۹۸

دانشکده: مهندسی کامپیوتر

گروه: هوش مصنوعی

رساله دکتری آقای ایمان فیروزیان

تحت عنوان: یک چارچوب احتمالاتی برای کنترل زمانی رویدادها در فرآیندهای سازمانی با منابع مشترک

در تاریخ ..... توسط کمیته تخصصی زیر جهت اخذ مدرک رساله دکتری ارزیابی گردید و با درجه ..... مورد پذیرش قرار گرفت.

اساتید راهنما	امضاء	اساتید مشاور	امضاء
نام و نام خانوادگی: دکتر مرتضی زاهدی		نام و نام خانوادگی: دکتر حمید حسن پور	

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

تقدیم به مأمن و جان‌پناه‌های زندگیم؛

پدر، مادر و همسر عزیزتر از جانم

تقدیر و تشکر

سپاس و ستایش مر خدای را جل و جلاله که آثار قدرت او بر چهره روز روشن، تابان است و انوار حکمت او در دل شب تار، درفشان. آفریدگاری که خویشتن را به ما شناساند و درهای علم را بر ما گشود و عمری و فرصتی عطا فرمود تا بدان، بنده ضعیف خویش را در طریق علم و معرفت بیازماید.

و استادانی را راهنمایم قرار داد تا چشمم را به زیبایی‌های دنیا باز نمایند؛

بی‌نهایت سپاسگزارم و قدردانی می‌کنم از استاد بزرگوارم، جناب آقای دکتر مرتضی زاهدی که پیش از علم به من، درس ادب و اخلاق آموختند. از استاد محترم جناب آقای دکتر حمید حسن‌پور جهت راهنمایی‌ها و مشاوره‌های ارزشمندشان سپاسگذاری کرده،
و همچنین از تمامی عزیزان و همکاران آزمایشگاهی، آقایان دکتر مهدی حسینی، دکتر مهدی یعقوبی، مهندس محمد عبدالهی و خانم دکتر مرضیه رحیمی که با همکاری‌ها، راهنمایی‌ها و همیاری‌هایشان همواره مشوق من بودند، قدردانی می‌نمایم.

تعهد نامه

اینجانب ایمان فیروزیان دانشجوی دوره دکتری رشته مهندسی کامپیوتر گرایش هوش مصنوعی دانشکده مهندسی کامپیوتر دانشگاه صنعتی شاهرود نویسنده پایان‌نامه "یک چارچوب احتمالاتی برای کنترل زمانی رویدادها در فرآیندهای سازمانی با منابع مشترک" تحت راهنمائی دکتر مرتضی زاهدی متعهد می‌شوم.

- تحقیقات در این پایان‌نامه 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 پایان‌نامه 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 می‌باشد 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 پایان‌نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان‌نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان‌نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان‌نامه، 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 افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می‌باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود . استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

چکیده

امروزه سازمان‌ها از سامانه‌های اطلاعاتی همانند سامانه مدیریت منابع سازمانی (ERP) و سامانه مدیریت فرآیندهای سازمانی (BPMS) برای ثبت فعالیت‌ها و رویدادهای خود در سابقه‌رویدادها (Event Log) استفاده می‌کنند. دانشی که به کشف، نظارت و بهبود فرایندهای سازمانی از درون سابقه‌رویدادها می‌پردازد، فرآیندکاوی نام دارد. یکی از حوزه‌های فرآیندکاوی، مدیریت و کنترل هوشمند زمانی فرآیندهای سازمانی بر اساس اطلاعات زمانی مستخرج از رویدادها می‌باشد که محققان بسیاری را به خود جذب کرده است.

کنترل زمانی، شامل عملیات‌هایی اعم از پیش‌بینی زمان باقی‌مانده پرونده‌ها در فرآیند، تنظیم زمان رسیدن یک نمونه‌فرآیند به یک وظیفه خاص در یک فرآیند و همچنین، ایجاد هم‌زمانی، تقدم و تأخر برای دو نمونه‌فرآیند مختلف می‌باشد. مسئله باز عمومی در کنترل زمانی، ارائه رویکرد پیش‌بینی زمانی با لحاظ حداکثر پارامترها و فاکتورها همانند صف، توزیع‌های احتمالاتی، رانش مفهوم، انشعاب توکن‌ها در نمونه‌فرآیندها، داده‌های نمونه‌فرآیندها، چندفرآیندی و غیره به صورت همزمان می‌باشد. مسئله باز تخصصی‌تر، کاوش وابستگی‌های پردازشی پنهان نمونه‌فرآیندها و تنظیم همزمانی روند نمونه‌فرآیندها برپایه رویکرد پیش‌بینی می‌باشد که این رساله در پی ارائه راهکار برای این مسائل باز می‌باشد.

این رساله با بررسی رویکردهای موجود برای کنترل زمانی نمونه‌فرآیندها، ابتدا چهار رویکرد پیشنهادی برپایه تکنیک تجزیه مسیر و فرمول‌بندی تحلیلی مسیر را با استفاده از مدل‌های ماشین بردار پشتیبان بدست‌آمده از داده‌های پیشین ارائه می‌کند و پس از آن یک رویکرد احتمالاتی با احتساب صف وظیفه‌ها را جهت پیش‌بینی چرخه زمانی باقیمانده نمونه فرآیندها ارائه می‌کند. آزمایش‌ها نشان می‌دهد رویکرد ترکیبی ارائه شده به نسبت رویکرد رگرسیون‌های محلی در پیش‌بینی چرخه زمانی، خطای MAE نرمال سازی شده را مجموعاً برای تمام ۸ پایگاه‌داده به اندازه ۱۲,۸۳ درصد کاهش داد. پس از آن، مفروضات مسئله را با لحاظ رانش مفهوم به دنیای واقعی نزدیک‌تر می‌کنیم و عملیات پیش‌بینی چرخه زمانی نمونه فرآیندها را با فرض تغییر در مدل فرآیندی با ارائه یک رویکرد آماری مبتنی بر سیستم گذار حاشیه‌نویسی شده با خروجی ماشین‌های بردار پشتیبان فازی انجام می‌دهیم. رویکرد پویای چندمدله پیشنهادی برای مقابله با رانش مفهوم بهبود ۲۲,۵۶ درصدی در معیار دقت نسبت به رویکرد تک مدله استاتیک مجموعاً بر روی دو پایگاه داده را نشان داد. درنهایت، یک وابستگی پردازشی پنهان میان جفت نمونه فرآیندهای وابسته از درون سابقه رویدادها به صورت رسمی تعریف و استخراج می‌کنیم و برای آن راهکاری با تنظیم و کنترل زمانی رویدادهای نمونه فرآیندها ارائه می‌نماییم. آزمایش‌ها برروی پایگاه داده BPI Challenge 2012 نشان از کاهش 4.3% چرخه زمانی کلی و کاهش 39% چرخه زمانی نمونه فرآیندهای وابسته نسبت به حالت بدون تنظیم همزمانی دارد.

کلمات کلیدی: مدیریت فرآیندهای سازمانی، چرخه زمانی، پیش‌بینی زمانی، رانش مفهوم، تنظیم همزمانی، نمونه فرآیندهای وابسته.

لیست مقالات ژورنالی مستخرج از پایان نامه

1. Iman Firouzian, Morteza Zahedi and Hamid Hassanpour, *Cycle Time Optimization of Processes Using an Entropy-Based Learning for Task Allocation.* International Journal of Engineering, 2019. 32(8): p. 1090-1100.

2. Iman Firouzian, Morteza Zahedi and Hamid Hassanpour, *Real-time Prediction and Synchronization of Business Process Instances using Data and Control Perspective*. International Journal of Nonlinear Analysis and Applications, 2019. 10(1): p. 217-228.

3. Iman Firouzian, Morteza Zahedi and Hamid Hassanpour, *Investigation of the Effect of Concept Drift on Data-Aware Remaining Time Prediction of Business Processes.* International Journal of Nonlinear Analysis and Applications, 2019. 10(2): p. 153-166.

لیست مقالات کنفرانسی مستخرج از پایان نامه

۱. ایمان فیروزیان، مرتضی زاهدی و حمید حسن پور، ۱۳۹۸، پیش‌بینی مدت زمان باقی‌مانده چرخه زمانی فرآیندهای سازمانی براساس جنبه‌های کنترلی و جریان داده، کنفرانس پنجمین کنفرانس پردازش سیگنال و سیستم‌های هوشمند ایران (ICSPIS)، شاهرود، دانشگاه صنعتی شاهرود.

۲. ایمان فیروزیان و مرتضی زاهدی، ۱۳۹۷، پیش‌بینی گلوگاه در فرآیندهای سازمانی با لحاظ رانش مفهوم، پنجمین کنفرانس مهندسی دانش بنیان و نوآوری (KBEI)، تهران، دانشگاه علم و صنعت.

فهرست مطالب

فهرست جداول **ل**

فهرست اشکال **م**

فصل۱ : مقدمه **۱**

۱-۱ مقدمه ۲

۱-۲ اصطلاحات اصلی ۴

۱-۳ مدیریت فرآیند سازمانی ۹

۱-۳-۱ چرخه حیات فرآیند سازمانی ۱۰

۱-۳-۲ فرآیند خدمات مشتری ۱۱

۱-۴ تعاریف رسمی ۱۳

۱-۵ بیان مسأله ۱۸

۱-۶ متدولوژی ۱۹

۱-۷ نتیجه‌گیری ۲۱

فصل۲ : پیشینه تحقیق **۲۳**

۲-۱ پیش‌بینی وضعیت نمونه‌فرآیند ۲۴

۲-۱-۱ پیش بینی کمی ۲۴

۲-۱-۲ پیش بینی های کیفی ۲۷

۲-۱-۳ پیش بینی فعالیت آتی ۲۸

۲-۲ کنترل زمانی ۲۹

۲-۳ نتیجه گیری ۳۶

فصل۳ :پیش بینی زمان باقیمانده نمونه فرآیندها ۳۷

۳-۱ مقدمه ۳۸

۳-۲ تجزیه مسیر نمونه‌فرآیند ۳۹

۳-۳ رویکردهای پیشنهادی برپایه تجزیه مسیر ۴۳

۳-۳-۱ کشف فرآیند ۴۴

۳-۳-۲ انطباق دنباله های جزئی برروی مدل فرآیند ۴۴

۳-۳-۳ فرمول‌بندی تحلیلی مسیر نمونه فرآیند ۴۶

۳-۳-۴ پیش بینی زمان نمونه فرآیند ۴۸

۳-۳-۵ ساخت بردار ویژگی ۵۰

۳-۳-۶ رویکرد پیش بینی‌کننده تک‌مدله ۵۰

۳-۳-۷ رویکرد پیش بینی کننده چندمدله ۵۲

۳-۴ رویکرد پیش بینی با توزیع‌های احتمالاتی با لحاظ صف ۵۷

۳-۵ ارزیابی ۶۲

۳-۵-۱ پایگاه‌های داده ۶۲

۳-۵-۲ ساخت مجموعه آموزشی و مجموعه آزمایشی ۶۴

۳-۵-۳ معیارهای ارزیابی ۶۵

۳-۵-۴ مراجع ارزیابی ۶۷

۳-۶ ارزیابی نتایج ۶۷

۳-۷ جمع‌بندی ۷۳

فصل۴ : پیش بینی زمانی نمونه فرآیندها با لحاظ رانش مفهوم ۷۵

۴-۱ مقدمه ۷۶

۴-۲ رانش مفهوم ۷۷
۴-۳ نمونه سابقه رویداد حاوی داده ۷۸
۴-۴ تکنیک‌های مورد استفاده در رویکرد پیشنهادی ۷۹
۴-۴-۱ سیستم گذار ۷۹
۴-۴-۲ ماشین بردار پشتیبان فازیFSVM ۸۱
۴-۵ پیش بینی مسیر نمونه فرآیند ۸۴
۴-۵-۱ تخمین احتمالات گذار ۸۵
۴-۵-۲ پیش بینی محتمل ترین مسیر ۸۶
٤-٦ تأثیر رانش مفهوم ۸۷
۴-۷ آزمایش‌ها ۹۰
۴-۷-۱ پایگاه داده شبیه سازی ۹۰
۴-۷-۲ پایگاه داده QUT ۹۱
۴-۸ نتیجه گیری ۹۲
فصل۵ : تنظیم همزمانی روند نمونه فرآیندهای وابسته **۹۵**
۵-۱ مقدمه ۹۶
۵-۲ فاز اول: کشف فرآیند ۹۶
۵-۳ فاز دوم: وابستگی پردازشی پنهان ۹۸
۵-۴ فاز سوم: پیش‌بینی زمانی نمونه‌فرآیندها ۱۰۱
۵-۵ فاز چهارم: تنظیم همزمانی نمونه‌فرآیندها ۱۰۳
۵-۶ نتایج آزمایش‌ها ۱۰۴
۵-۷ نتیجه گیری ۱۰۸

فصل۶ : نتیجه‌گیری و کارهای آتی **۱۰۹**

مراجع **۱۱۴**

فهرست جداول

جدول ۱-۱: سابقه رویداد سناریوی بیان‌شده. هر ردیف مربوط به یک رویداد می‌باشد و رویدادها با برچسب زمانی مرتب شده‌اند. ۱۵
جدول ۱-۲ : سابقه رویداد ساده‌سازی شده به‌عنوان مجموعه‌ای از دنباله‌های سابقه رویداد و فرکانس آن‌ها. ۱۷
جدول ۳-۱: بردار ویژگی و مجموعه داده آموزشی برای پیش بینی طول مدت زمان فعالیت F (رویکرد تک پیش بینی کننده) ۵۴
جدول ۳-۲: بردار ویژگی و مجموعه داده آموزشی برای پیش بینی احتمال انشعاب XOR-32 (رویکرد تک پیش بینی کننده) ۵۵
جدول ۳-۳: بردار ویژگی و مجموعه داده آموزشی برای پیش بینی احتمال انشعاب XOR-32 و K=3 (رویکرد چند پیش‌بینی‌کننده) ۵۶
جدول ۳-۴: بردار ویژگی و مجموعه داده آموزشی برای پیش بینی طول مدت زمان فعالیت F و K=3 (رویکرد چند پیش‌بینی‌کننده) ۵۶
جدول ۳-۵: میانگین خطای وزن‌دار MAE برروی اندازه‌های مختلف دنباله‌های جزئی ۶۹
جدول ۴-۱ : یک مثال از بخشی از سابقه رویداد با رویدادهای مرتب شده توسط برچسب زمانی و گروه بندی شده با شناسه پرونده. ۷۸
جدول ۴-۲: نتایج آزمایش‌ها برروی پایگاه داده شبیه سازی شده. امتیازات زیرخط دار متعلق به بهترین امتیازات برای هر مقدار S می باشد. ۹۱
جدول ۴-۳: مقادیر بهینه برای β و τ و S به همراه دقت و معیار F1 برای پایگاه داده QUT ۹۲

فهرست اشکال

شکل ۱-۱: سه محور اصلی فرآیندکاوی شامل کشف فرآیند، تطابق و بهبود می‌باشد [۱]. ... ۳
شکل ۱-۲: ارتباط میان اصطلاحات اصلی حوزه BPM ۵
شکل ۱-۳: فرآیندها و نمونه فرآیندها[۳] ۹
شکل ۱-۴: مدل غیررسمی از فرآیند خدمات مشتری [۱۱] ۱۲
شکل ۲-۱: سیستم گذار حاشیه‌نویسی شده ارائه شده توسط Aalst و همکاران [۱۳]..... ۳۱
شکل ۲-۲: مدل ارائه‌شده توسط Andreas Rogge-soltie برای حل احتمالاتی فرآیند [۳] . ۳۲
شکل ۳-۱: الگوهای مرسوم مدل فرآیندی الف) ترتیبی، ب) بلوک XOR، ج) بلوک AND، د) حلقه تکرارکاری. ۴۱
شکل ۳-۲: چارچوب کلی رویکرد پیشنهادی ۴۳
شکل ۳-۳: کاربرد الگوریتم عقب گرد در تطبیق دنباله با درخت فرآیندی..... ۴۵
شکل ۳-۴: یک نمونه مدل فرآیندی که در آن وضعیت جاری یک نمونه فرآیند، هاشورزده نشان داده شده است. ۴۷
شکل ۳-۵: تجزیه دفعات تکرار حلقه تکرار کاری ۴۸
شکل ۳-۶: توزیع‌های احتمالاتی مربوط به سه وظیفه در فرآیند شکل الف، (ب) به صورت مجزا، (ج) ترکیب احتمالاتی طبق فرآیند، (د) احتمال مستقل "به اتمام نرسیدن" تا لحظه t، (ه) احتمال "به اتمام نرسیدن" تا لحظه t با احتساب وابستگی طبق فرآیند. ۵۸
شکل ۳-۷: محاسبه احتمالات در مواجهه با کنترل انتخابی XOR ۵۹
شکل ۳-۸: محاسبه احتمالات در مواجهه با کنترل AND ۶۰
شکل ۳-۹: فرآیند غنی‌شده با توزیع‌های احتمالاتی..... ۶۰
شکل ۳-۱۰: وضعیت یک فرآیند با احتساب صف های آن در لحظه مشخص t ۶۲
شکل ۳-۱۱: دقت پیش‌بینی رویکردها در سه پایگاه داده BPI Challenge 2012 A, B, C با معیار MAE برای اندازه‌های مختلف دنباله‌های جزئی ۷۰
شکل ۳-۱۲: دقت پیش‌بینی رویکردها در سه پایگاه داده CR، HelpDesk و Hospital با معیار MAE برای اندازه‌های مختلف دنباله‌های جزئی ۷۱
شکل ۳-۱۳: دقت پیش‌بینی رویکردها در سه پایگاه داده RFTMP وINVOICE با معیار MAE برای اندازه‌های مختلف دنباله‌های جزئی ۷۲
شکل ۳-۱۴: خطای میانگین MAE نرمال سازی شده..... ۷۳
شکل ۴-۱: انواع حالت ها در رانش مفهوم[۶۹] ۷۸

شکل ۴-۲: سیستم گذار حاشیه نویسی شده با توابع بازنمایی حالت و بازنمایی رویداد.... ۷۹
شکل ۴-۳: یک نمونه سیستم گذار مستخرج از سابقه رویداد ۸۵
شکل ۴-۴: رویکرد پیشنهادی برای سیستم گذار حاشیه نویسی شده ۸۶
شکل ۴-۵: تطبیق با رانش مفهوم با رویکرد تقسیم مجموعه داده..... ۸۸
شکل ۴-۶: نتایج پایگاه داده شبیه سازی شده در قالب معیار F1 می باشد. هر سه روش برای مقادیر مختلف S ارزیابی شده اند. ۹۱
شکل ۴-۷: نتایج پایگاه داده QUT در قالب معیار F1 می باشد. هر سه روش برای مقادیر مختلف S ارزیابی شده اند. ۹۲
شکل ۵-۱: مدل فرآیندی مستخرج غنی سازی شده با توزیع های احتمالاتی ۹۸
شکل ۵-۲: وابستگی پردازشی پنهان با خط چین نشان داده شده است. ۱۰۰
شکل ۵-۳: وابستگی پردازشی پنهان میان دو مدل فرآیندی مختلف ۱۰۰
شکل ۵-۴: وابستگی پردازشی پنهان در مسیرهای موازی ۱۰۱
شکل ۵-۵: نمونه ای از مسیر پیش بینی شده و غنی سازی شده با اطلاعات زمانی ۱۰۲
شکل ۵-۶: نمودار چرخه زمانی نمونه فعالیت های به اتمام رسیده ۱۰۵
شکل ۵-۷: نسبت زمان انتظار نمونه فرآیندهای وابسته به چرخه زمانی کلی به ازای بازه های مختلف ورود ۱۰۶
شکل ۵-۸: میزان خطای MAE به ازای تعداد وظایف مختلف در مسیر همزمان‌سازی.... ۱۰۷
شکل ۵-۹: میزان کاهش چرخه زمانی به ازای بازه های مختلف ورود ۱۰۷

فصل۱ : مقدمه

۱-۱ مقدمه

فرآیندها[۱] همه‌جا هستند. یک فرآیند را می‌توان به صورت مجموعه‌ای از فعالیت‌های مرتبط به هم که در طول زمان اتفاق می‌افتد و هدف مشترکی را دنبال می‌کند، تعریف نمود. ما فرآیندها را در زندگی روزانه خود مشاهده می‌کنیم. برای نمونه، هنگامی که می‌خواهیم در جلسه‌ای در یک شهر ناآشنا حضور یابیم، ابتدا آدرس را بر روی نقشه مشخص می‌کنیم، سپس اگر قصد سفر با خودرو را داریم، از مسیریاب خودرو کمک می‌گیریم، در غیراینصورت از آژانس‌های مسافرتی مشاوره می‌گیریم تا بهترین مسیر را بیابیم. پس از مکان‌یابی، یک هتل در مکان مقصد انتخاب می‌کنیم و عملیات رزرو را انجام می‌دهیم. پس از آن، تجهیزات لازم مسافرتی فراهم می‌کنیم و به سمت مقصد حرکت می‌کنیم و پس از رسیدن به مقصد به مراحل اسکان در هتل می‌پردازیم.

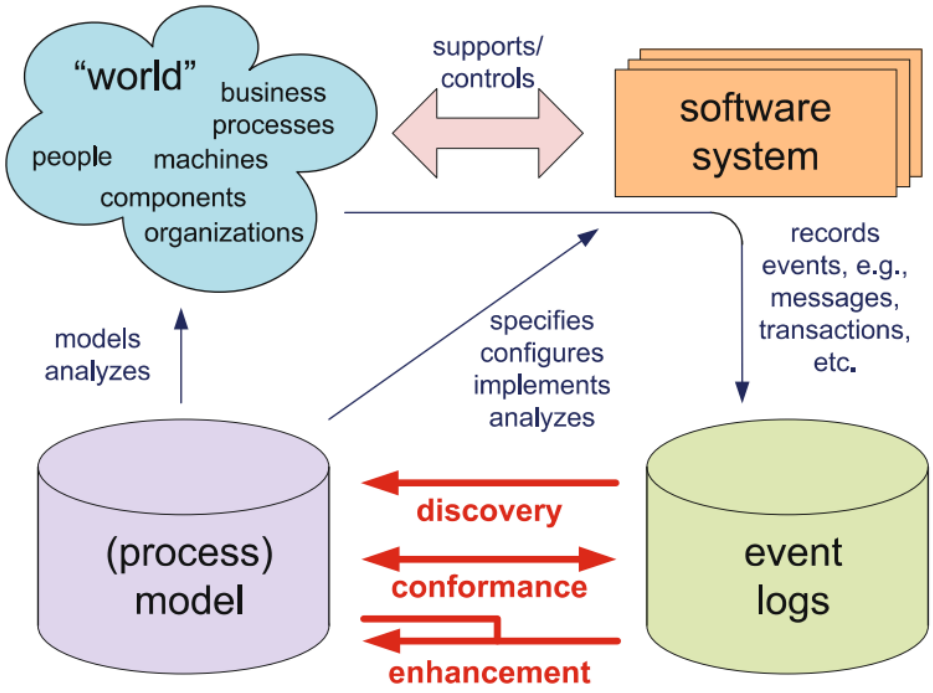
چنین فرآیندهایی را می‌توان در کارخانه‌ها، بیمارستان‌ها، مؤسسات مالی، دانشگاه‌ها و دیگر سازمان‌ها یافت. در حال حاضر، بسیاری از سازمان‌ها، از سیستم‌های اطلاعاتی هوشمند برای پوشش فرآیندهای سازمانی خود استفاده می‌کنند. این سیستم‌ها، سطوح مختلفی از پوشش را فراهم می‌کنند. برای مثال، روال اخذ مجوز ساخت در شهرداری بسیار قانون‌مند می‌باشد و چنین فرآیندهایی توسط سیستم‌های فرآیند-آگاه بسیار دقیق مدیریت می‌شوند تا کاربران ملزم به پیروی دقیق از دنباله خاصی از فعالیت‌ها شوند. در مقابل، جریان‌های کاری دیگری وجود دارد که بسیار متنوع و منعطف می‌باشند، برای مثال، در بیمارستان‌ها، پرونده بیمارها مسیرهای متفاوتی را طی می‌کند. بنابراین سیستم‌های اطلاعاتی در بیمارستان‌ها، اصولاً از یک فرآیند دقیق پیروی نمی‌کنند اما تنها فعالیت‌های پزشکی را برای اهداف مالی ثبت می‌کند. با این وجود، تمام سیستم‌های اطلاعاتی از فعالیت‌های خود "رد پا" به جا می‌گذارند و آنچه را که اتفاق افتاده را ثبت می‌کنند. این ردپاها سابقه رویداد نامیده می‌شود و معمولاً در پایگاه داده یا در فایل سابقه رویداد[۲] ذخیره می‌شود.

---

۱ Process

۲ Event log

رویدادها نقطه شروع تکنیک‌های فرآیندکاوی[۱] می‌باشد. از آنجاییکه سابقه‌رویدادهای سیستم‌های اطلاعاتی، داده‌های عملی و واقعی را در مورد فرآیندهای مربوطه فراهم می‌کند، منبع اطلاعاتی بسیار باارزشی محسوب می‌شود. بسیاری از صاحبان فرآیند، دانش اندکی در مورد فرآیندهای اجرایی خود دارند. بسیاری از سازمان‌ها، فرآیندهای خود را در قالب‌های مشخص، مستندسازی می‌کنند تا برای مثال، قوانین را با اهداف تعیین‌شده بررسی کنند. با استفاده از تکنیک‌های فرآیندکاوی می‌توان (۱) مدل‌های جریان واقعی فرآیند را به صورت اتوماتیک از سابقه رویدادها استخراج نمود (کشف[۲])، (۲) انحرافات را در روال‌های مستندسازی شده تشخیص داد (تطابق[۳])، و (۳) مدل‌های موجود را با تمرکز بر روی تنگناها، دخالت دادن جنبه‌های دیگر و غیره غنی‌سازی نمود. (توسعه[۴]). شکل ۱-۱ محورهای اصلی فرآیندکاوی را به تصویر می‌کشد.



شکل ۱-۱: سه محور اصلی فرآیندکاوی شامل کشف فرآیند، تطابق و بهبود می‌باشد [۱].

[۱] Process Mining
[۲] Discovery
[۳] Conformance
[۴] Enhancement

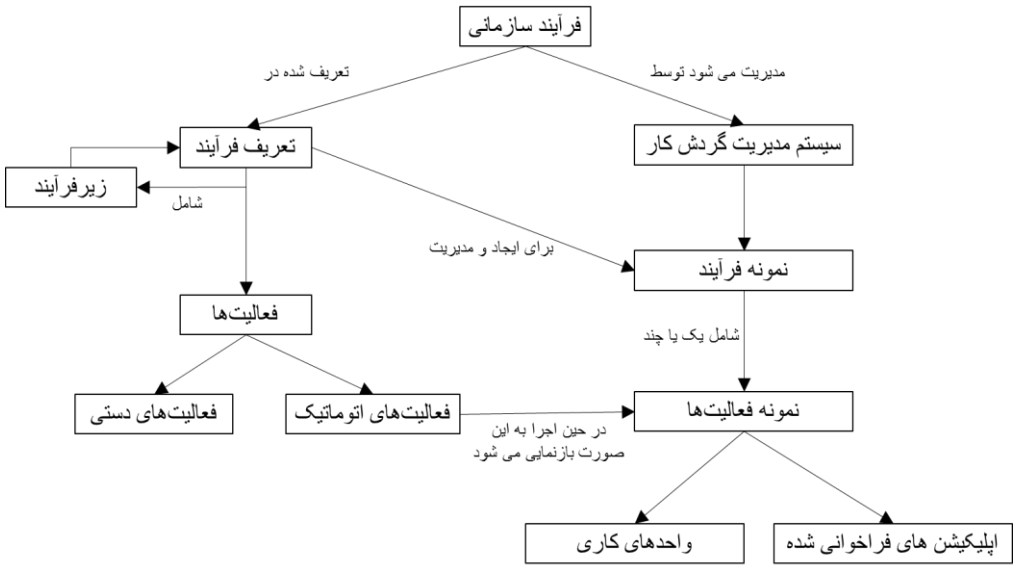
فرآیندکاوی با استفاده از سابقه‌رویدادها، تلاش می‌کند تا یک تصویر واقعی از فرآیندی که در حال رخ دادن است، ایجاد کند و در نتیجه تغییراتی را برای بهبود عملکرد یا تطبیق فرآیند ارائه نماید. ارائه شفاف‌سازی در مورد آنچه حقیقتاً در حال رخ دادن است، بسیار باارزش محسوب می‌شود. علاوه براین، دانش پیرامون وضعیت فعلی یک پیش‌نیاز برای هر نوع بهبود محسوب می‌شود، که جمله معروف "تنها چیزی که قابل اندازه‌گیری است را می‌توان بهبود داد" نیز مهر تأییدی بر آن است.

هر سه بعد فرآیندکاوی، شامل کشف، انطباق و بهبود را باید لحاظ کرد تا بتوان تصویر جامعی از هر فرآیند مفروض بدست آورد. برای مثال، بسیاری شرکت‌ها فرآیندهای خود را مستندسازی می‌کنند و بیشتر علاقه‌مند به تشخیص انحرافات با توجه به روال‌های مستندسازی شده هستند و کمتر علاقه‌مند به کشف مدل‌های جدید هستند، مخصوصاً هنگامی که آن‌ها مجبور به پیروی از یک فرآیند قانونی خاص هستند. علاوه بر این، باید یک مدل فرآیندی کشف‌شده بررسی شود که تا چه میزان نمایانگر واقعیت می‌باشد. توجه داشته باشید که یک مدل کشف‌شده به دلیل وجود نویز و/یا محدودهای موجود در فن‌های کشف، معمولاً با سابقه رویدادها سازگار نیست. نهایتاً برای تشخیص تنگناها و اعمال بهبود، نیاز به اطلاعات اضافی همچون زمان، هزینه، یا منابع می‌باشد. گرچه، حوزه تحقیقاتی این رساله عمدتاً برروی بحث‌های مدیریت زمانی در فرآیندها می‌باشد.

۱-۲ اصطلاحات اصلی

ائتلاف مدیریت جریان کاری[۱]، اصطلاحات حوزه جریان کاری و BPM را برای یکسان‌سازی ادبیات بیان جملات در حوزه، استاندارد کرده است [۲]. ارتباطات میان اصطلاحات اصلی این حوزه به صورت شماتیک در شکل ۱-۲ نشان داده شده است. در ادامه، تعریفی برای هر کدام از اصطلاحات ارائه می‌شود.

[۱] Workflow Management Coalition



شکل ۱-۲: ارتباط میان اصطلاحات اصلی حوزه BPM

تعریف ۱ (جریان کاری): به اتوماسیون کل یک فرآیند سازمانی یا بخشی از یک فرآیند سازمانی جریان کاری گفته می‌شود. این اتوماسیون، بستری را با درنظر گرفتن مجموعه‌ای از روال‌های قانونمند، برای تبادل اسناد، اطلاعات یا وظیفه‌ها از یک شرکت‌کننده به شرکت‌کننده دیگر، برای انجام عملیات خاص، فراهم می‌کند.

تعریف ۲ (سیستم مدیریت جریان کاری[۱]): سیستمی که نحوه اجرای جریان‌های کاری را تعریف، ایجاد و مدیریت می‌کند. این سیستم یک محیط نرم‌افزاری برپایه یک یا چند موتور جریان کاری می‌باشد و دارای قابلیت تفسیر فرآیند، تعامل با شرکت‌کننده‌های جربان کاری و درصورت نیاز، به‌کارگیری ابزارهای هوشمند IT می‌باشد.

تعریف ۳ (فرآیند سازمانی[۲]): مجموعه‌ای از یک یا چند فعالیت مرتبط می باشدکه مجموعاً از هدف فرآیند سازمانی یا سیاست سازمانی آگاه می‌باشند و عموماً در قالب ساختار سازمانی با تعریف نقش‌ها و روابط میان آن‌ها می‌باشد.

---

۱ Workflow Management System

۲ Business Process

تعریف ۴ (فرآیند[۱]): بازنمایی یک فرآیند سازمانی به‌نحوی‌که اعمال تغییرات اتوماتیک را پوشش دهد؛ این تغییرات اتوماتیک همانند مدل‌سازی و عملیاتی سازی توسط سیستم مدیریت جریان کاری انجام می‌شود. فرآیند شامل شبکه‌ای از فعالیت‌ها و روابط میان آن‌هاست. هر فرآیند شروع و پایان مشخص دارد، و در مورد هر فعالیت منفرد اطلاعاتی همانند شرکت‌کننده‌ها، برنامه‌های IT مرتبط و داده‌های دیگر را مشخص می‌کند.

تعریف ۵ (فعالیت[۲]): توصیفی از بخشی از کار که یک گام منطقی را درون یک فرآیند را نشان می‌دهد. یک فعالیت ممکن است فعالیت دستی باشد که اتوماسیون کامپیوتری آن را پوشش نمی‌دهد و یا یک فعالیت اتوماتیک جریان کاری باشد. یک فعالیت جریان کاری نیازمند منابع انسانی و/یا ماشینی برای پوشش اجرای فرآیند می‌باشد. هر منبع انسانی به فعالیت منتسب می‌شود.

تعریف ۶ (فعالیت اتوماتیک): فعالیتی که با استفاده از سیستم مدیریت جریان کاری قادر به اتوماسیون کامپیوتری می‌باشد. این فعالیت قادر به مدیریت طول اجرای فعالیت فرآیند سازمانی که بخشی از آن را تشکیل می‌دهد را می‌باشد.

تعریف ۷ (فعالیت دستی): یک فعالیت درون فرآیند سازمانی که قادر به اتوماسیون نیست و خارج از حوزه سیستم مدیریت جریان کاری می‌باشد. این فعالیت‌ها ممکن است درون یک فرآیند وجود داشته باشند، برای مثال بایگانی دستی پرونده‌ها، یک فعالیت دستی محسوب می‌شود.

تعریف ۸ (نمونه فرآیند یا نمونه فعالیت): بازنمایی نمونه عملیاتی یک فرآیند یا نمونه عملیاتی فعالیت درون یک فرآیند را به همراه داده‌های مرتبط آن گویند. هر نمونه، مشخص‌کننده نخ اجرای منحصربه فردی را از فرآیند یا فعالیت می‌باشد. این نخ‌های اجرایی ممکن است به صورت مستقل کنترل شود و وضعیت داخلی و هویت آشکار خارجی خاص خودش را دارد. در ضمن، این نخ‌های اجرایی را می‌توان به عنوان handle استفاده کرد، برای مثال، برای ثبت یا دریافت داده‌های حسابرسی مرتبط با

---

[۱] Process

[۲] Activity

هر نمونه عملیاتی منفرد.

تعریف ۹ (نمونه فرآیند): هر نمونه فرآیند، بازنمایی هر اجرای منفرد یک فرآیند می‌باشد.

تعریف ۱۰ (نمونه فعالیت): هر نمونه فعالیت، بازنمایی اجرای یک فعالیت در هر اجرای منفرد یک فعالیت می‌باشد.

تعریف ۱۱ (مشارکت‌کننده جریان کاری[۱]): مشارکت‌کننده جریان کاری به منبعی گفته می‌شود که کار مشخص‌شده توسط نمونه فعالیت جریان کاری را انجام می‌دهد. این کار معمولاً در قالب یک یا چند آیتم کاری به مشارکت‌کننده جریان کاری منتسب می‌شود و به لیست کاری آن اضافه می‌شود.

تعریف ۱۲ (آیتم کاری[۲]): بازنمایی کاری است که قرار است (توسط یک شرکت‌کننده جریان کاری) در قالب یک فعالیت درون نمونه فرآیند انجام شود.

تعریف ۱۳ (لیست کاری): لیستی از آیتم‌های کاری منتسب به یک مشارکت‌کننده جریان کاری می‌باشد. در برخی مواقع، به لیستی از آیتم‌های کاری منتسب به گروهی از مشارکت‌کننده‌های جربان کاری که لیست کاری مشترکی را با هم به اشتراک می‌گذارند، گفته می‌شود. لیست کاری بخشی از واسطه میان موتور جریان کاری و handler لیست کاری را تشکیل می‌دهد.

تعریف ۱۴ (هندلر لیست کاری): یک مؤلفه نرم‌افزاری که تعامل میان کاربر (یا گروهی از کاربران) با لیست کاری موجود در موتور جریان کاری را مدیریت می‌کند. به کمک این مؤلفه نرم‌افزاری می‌توان آیتم‌های کاری را از سیستم مدیریت جریان کاری به کاربران ارسال نمود و همچنین پیام‌های اطلاع‌رسانی پیرامون تکمیل یا دیگر شرایط وضعیت کاری را میان کاربر و سیستم مدیریت جریان کاری تبادل نمود.

تعریف ۱۵ (فرآیند): یک دید رسمی به فرآیند سازمانی، که به صورت مجموعه‌ای هماهنگ (موازی و/یا سریال) از فعالیت‌ها می‌باشد که برای رسیدن به یک هدف مشترک به یکدیگر متصل شده‌اند.

---

۱ Workflow Participant

۲ Work Item

تعریف ۱۶ (زیرفرآیند): فرآیندی توسط یک فرآیند (زیرفرآیند) دیگر فراخوانی یا اجرا شده است، و بخشی از آن فرآیند فراخوانی کننده محسوب می‌شود. ممکن است چندین سطح از زیرفرآیند وجود داشته باشد.

تعریف ۱۷ (بلوک فعالیت): مجموعه‌ای از فعالیت‌های درون فرآیند می‌باشد که یک یا چند مشخصه مشترک را به اشتراک می‌گذارند که باعث می‌شود سیستم مدیریت جریان کاری عملیات‌های خاصی را با توجه به بلوک به صورت کلی انجام دهد. برای مثال، در گروه‌بندی فعالیت‌ها در صورت وجود اشتراک در سیاست تخصیص منبع ممکن است به عنوان یک بلوک دسته‌بندی شود.

تعریف ۱۸ (مهلت): یک قید زمانی است برای آنکه یک فعالیت خاص (یا آیتم کاری) در زمان مشخص تکمیل شود.

تعریف ۱۹ (مسیریابی موازی): بخشی از یک نمونه فرآیند درحال اجرا توسط سیستم مدیریت جریان کاری می‌باشد، که در آن دو یا چند نمونه فعالیت به صورت موازی در یک جریان کاری به کمک نخ‌های کنترلی اجرا می‌شوند.

تعریف ۲۰ (مسیریابی ترتیبی): بخشی از یک نمونه فرآیند درحال اجرا توسط سیستم مدیریت جریان کاری می‌باشد، که در آن چند فعالیت به ترتیب تحت یک نخ اجرای منفرد، اجرا می‌شود. در مسیریابی ترتیبی هیچ‌کدام از شرطهای AND-split و AND-join وجود ندارد.

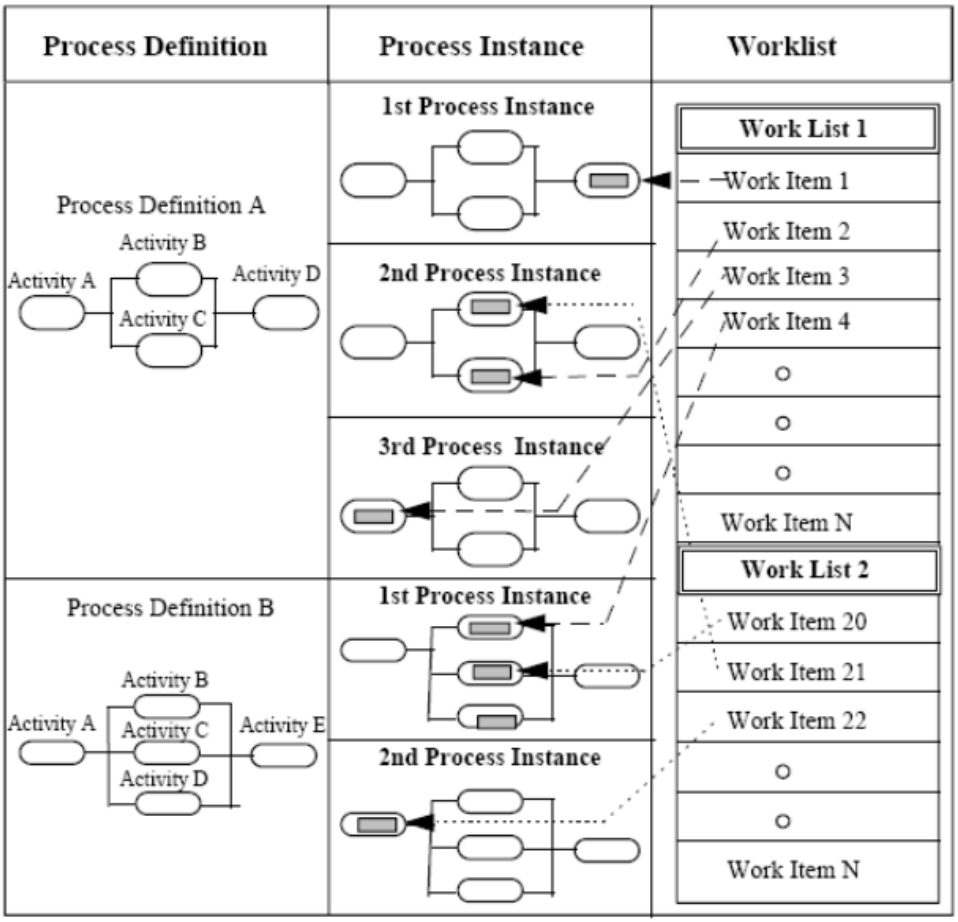
تعریف ۲۱ (تکرار): یک چرخه فعالیت جریان کاری شامل اجرای تکراری یک (یا بیشتر) فعالیت (های) جریان کاری می‌باشد تا شرط پایان برآورده شود.

تعریف ۲۲ (پیش‌شرط): پیش‌شرط یک عبارت منطقی می‌باشد که ممکن است توسط یک موتور جریان کاری ارزیابی شود تا برای شروع نمونه فرآیند یا نمونه فعالیت تصمیم‌گیری کند.

تعریف ۲۳ (پس شرط): پس شرط، یک عبارت منطقی می‌باشد که توسط موتور جریان کاری ارزیابی می‌شود تا برای پایان یک نمونه فرآیند یا نمونه فعالیت را تصمیم‌گیری کند.

تعریف ۲۴ (گذر): نقطه‌ای در طول اجرای یک نمونه فرآیند می‌باشد که یک فعالیت تکمیل می‌شود و نخ کنترلی به فعالیت دیگری که آغاز می‌شود، داده می‌شود.

تعریف ۲۵ (شرط گذر): شرط گذر، یک عبارت منطقی است که توسط موتور جریان کاری ارزیابی می‌شود تا برای دنباله اجرای فعالیت‌ها را درون یک فرآیند تصمیم‌گیری کند.



شکل ۱-۳: فرآیندها و نمونه فرآیندها[۳]

۱-۳ مدیریت فرآیند سازمانی

در دنیای رقابتی امروزی، سازمان‌ها تلاش بسیاری را در جهت بهبود کارایی امور سازمان و اثربخشی آن‌ها انجام می‌دهند [۴, ۵]. ابزارها و تکنیک‌هایی برای پوشش مدیریت فرآیندهای سازمانی به وجود آمده است. نظام ساختاری BPM توسط چرخه حیات فرآیند سازمانی به بهترین نحو بیان می‌شود.

۱-۳-۱ چرخه حیات فرآیند سازمانی

چرخه حیات فرآیند سازمانی [۶] یک نگاه سطح بالا بر روی فازهای فرآیند سازمانی دارد. چرخه حیات شامل این فازها می‌باشد: (۱) ارزیابی، (۲) طراحی و آنالیز، (۳) پیکربندی و (۴) عملیاتی سازی فرآیندهای سازمانی [۵]، که در شکل ۱-۲ نمایش داده می‌شود.

در فاز ارزیابی، فرآیندهای سازمانی as-is یک سازمان بررسی می‌شوند. سپس، از تکنیک‌هایی همانند فرآیندکاوی [۱] برای درک یک بینش مناسب از فرآیندهای در حال اجرا استفاده می‌شود. در فاز طراحی و آنالیز مدیریت فرآیندهای سازمانی، فرآیندها شناسایی می‌شوند و کارشناسان خبره، فرآیندهای to-be را در قالب مدل‌های فرآیند سازمانی تحلیل می‌کنند. اگر مدل‌های فرآیند سازمانی در این فاز وجود داشته باشند، طراحی مجدد [۷] صورت می‌پذیرد. مدل‌های فرآیندی که مجدداً طراحی شده را می‌توان برای تصحیح [۸] و ابزارهای شبیه‌سازی [۹] برای بررسی تغییرات ایجادشده فرآیندهای as-is استفاده می‌شود.

پس یک مدل فرآیند سازمانی برای اجرا انتخاب شد، فاز پیکربندی آغاز می‌شود. در این فاز، فرآیند در محیط عملیاتی توسعه می‌یابد. محیط‌ها را می‌توان به دو دسته اتوماتیک و دستی دسته‌بندی نمود. در محیط‌های دستی، پیکربندی منجر به قاعده‌مند کردن امور کارکنان می‌شود. در محیط‌های اتوماتیک، پیکربندی منجر به کنترل اتوماسیون فرآیندها توسط موتور اجرایی فرآیند می‌شود. برای مثال، در محیط‌های بیمارستانی اتوماسیون به دلیل دستی انجام شدن کارها، استفاده از موتورهای جریان کاری بسیار چالش‌برانگیز است [۱۰]. این عدم امکان به خاطر نیازمندی‌های با انعطاف بالا، پیچیدگی یا رخداد ضعیف آن است. هزینه پیاده‌سازی یک موتور جریان کاری با میزان کارایی بدست آمده از اتوماسیون تناسب دارد.

در فاز عملیاتی، که چرخه حیات مدیریت فرآیند سازمانی را با آن پایان می‌پذیرد، بکارگیری فرآیند سازمانی در محیط عملیاتی صورت می‌پذیرد.

برای توضیح ایده اصلی فرآیندکاوی، از مثال فرآیند خدمات مشتری استفاده می‌کنیم. در این بخش، ابتدا سناریو را با جزییات توضیح می‌دهیم و پس از آن مفاهیم حوزه فرآیند را به کمک آن ارائه می‌کنیم.

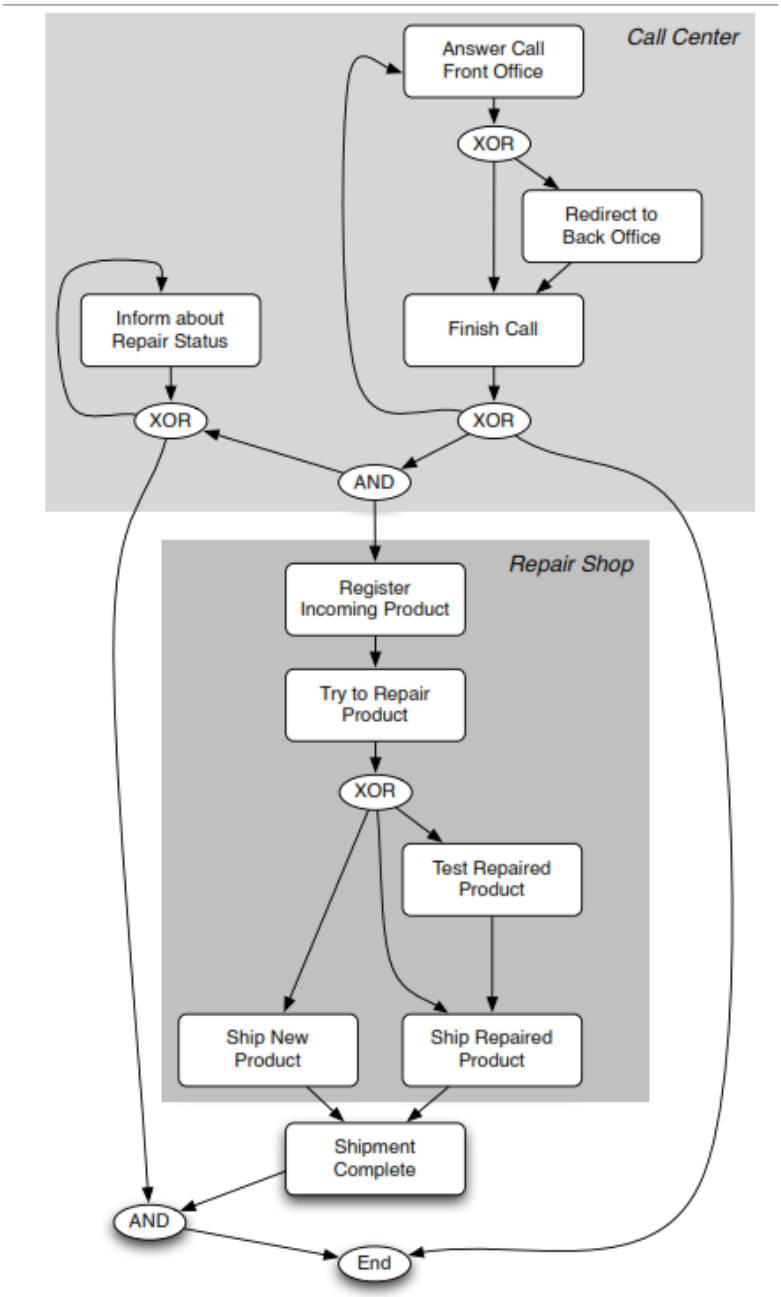
۱-۳-۲ فرآیند خدمات مشتری

شکل ۱-۴ فرآیند خدمات مشتری را در یک شرکت فرضی A نشان می‌دهد. مشتری‌هایی که مشکلی با محصول شرکت A دارند، با مرکز خدمات تلفنی تماس می‌گیرند. مرکز خدمات تلفنی دارای یک دفتر عمومی با کارمندانی با دانش عمومی می‌باشد که می‌توانند مسائل ساده عمومی را حل کنند. اگر مشکل با هیچ‌کدام از کارمندان دفتر عمومی حل نشود، مشتری به کارمند دفتر تخصصی ارجاع داده می‌شود. هر تماس‌گیرنده، یک شماره درخواست سرویس (SR) منحصربه‌فرد دریافت می‌کند که با استفاده از آن، کارمند مرکز خدمات تلفنی قادر خواهد بود تا به تاریخچه کامل خدمات دسترسی پیدا کند. اگر محصول نیاز به تعمیر داشته باشد و هنوز دارای گارانتی باشد، مشتری شماره تعمیر ویژه (R) را دریافت می‌کند و محصول به کارگاه تعمیر ارسال می‌شود. در آنجا، محصولات ورودی در صورت امکان تعمیر می‌شود. اگر تعمیر موفقیت‌آمیز باشد، باید پیش از آنکه به مشتری بازگردانده شود، تست شود. طبق راهنماهای کیفیت، این تست نباید توسط همان شخصی که تعمیر کرده، انجام شود. اگر امکان تعمیر آن وجود نداشته باشد، مشتری محصول جدید را به جای آن دریافت می‌کند.

در شرکت A، مهندس خدمات مشتری مسئول هزینه‌ها و کیفیت فرآیند سرویس می‌باشد. رضایت مشتری مهم‌ترین هدف در راستای کیفیت می‌باشد. تعدادی سؤال برای مهندس خدمات مشتری پیش می‌آید، همانند "مشکل چند مشتری پس از اولین تماس مشکلشان حل شد؟ (و دیگر تماس نگرفتند)"، درصد این تماس‌ها نسبت به کل تماس‌ها، با نرخ دقت اولین-تماس شناخته می‌شود و ثابت شده است که با میزان رضایت مشتری وابستگی دارد. مشتری‌هایی که چندین بار تماس گرفته‌اند از سرویس‌هایی که توسط شرکت A ارائه شده بود، کمتر راضی بودند. سؤال‌های آماری دیگر همانند "عملکرد زمانی فرآیند چگونه است؟ چقدر طول می‌کشد تا فعالیت‌های خاص انجام شود؟"، "چه زمانی پرونده جاری

به نقطه پایان فرآیند می‌رسد؟ چه زمانی پرونده جاری به یک فعالیت خاص می‌رسد؟" نیز می‌توان در نظر گرفت.

در شکل ۱-۴، فرآیند خدمات مشتری به صورت غیررسمی بازنمایی شده است تا مسیر حرکت پرونده‌ها از نقطه آغاز تا پایان به صورت شماتیک نشان داده شود.



شکل ۱-۴: مدل غیررسمی از فرآیند خدمات مشتری [۱۱]

۱-۴ تعاریف رسمی

برای تعریف رسمی مفاهیم سابقه رویدادها و مدل‌های فرآیندی، نیاز به معرفی نمادگذاری‌های زیر داریم.

- $f \in A \to B$ تابعی با دامنه A (dom(f)) و برد B (rng(f)) می‌باشد.
- $f \in A \nrightarrow B$ تابع جزئی می‌باشد که ، دامنه f ممکن است زیرمجموعه‌ای از A باشد $(dom(f) \subseteq A)$.
- یک چند-مجموعه مشابه مجموعه است با این تفاوت که از هر عنصر ممکن است چند تا وجود داشته باشد. برای مثال، $[a, b^2, c^3, d, d, e]$ یک چند-مجموعه است با نه عنصر شامل یک a، دو b، سه c، دو d و یک e می‌باشد. چند-مجموعه کیف هم نامیده می‌شود.
- $B(A) = A \to N$ یک مجموعه‌ای از چند-مجموعه‌ها (کیف‌ها) در دامنه متناهی A می‌باشد. $X \in B(A)$ یک چند-مجموعه است که به ازای هر $a \in A$، X(a) تعداد دفعاتی که a در چند-مجموعه وجود دارد را مشخص می‌نماید.
- $|X| = \sum_{a \in A}^{n} X(a)$ کاردینالیتی برخی چند-مجموعه‌ها بر روی A می‌باشد. این تابع را می‌توان به یک مجموعه نیز اعمال نمود. فرض بر این است که یک مجموعه، یک چند-مجموعه می‌باشد که در آن هر عنصر تنها یک بار وجود دارد.
- P(A) مجموعه توانی A می‌باشد، که $P(A) = \{ X \mid X \subseteq A \}$.
- به ازای هر مجموعه مفروض A، A* مجموعه تمام توالی‌های متناهی بر روی مجموعه A می‌باشد.
- یک دنباله متناهی بر روی A با طول n همان نگاشت $\sigma \in \{ 1, \ldots, n \} \to A$ می‌باشد. هر دنباله با یک رشته نمایش داده می‌شود، به طوریکه $\sigma = \langle a_1, a_2, \ldots, a_n \rangle$ که به ازای $1 \le i \le n$، $\sigma(i) = a_i$ می‌باشد و $|\sigma| = n$ برابر با طول دنباله σ می‌باشد.

- الحاق دو دنباله متناهی $\sigma=<a_1, a_2, \ldots, a_n>$ و $\sigma'=<b_1, b_2, \ldots, b_m>$ توسط $b_1, b_2, \ldots, b_m>$ $\sigma. \sigma'=< a_1, a_2, \ldots, a_n,$ مشخص می‌شود.
- $set(\sigma)$ یک دنباله σ را به یک مجموعه تبدیل می‌کند، که $set(\sigma)=\{ \sigma(i)|1\leq i\leq|\sigma|\}$ می‌باشد.
- $f(\sigma)$ اعمال تابع f را بر روی هر عنصر دنباله σ نشان می‌دهد، که به صورت $f()=<f(\sigma(1)), f(\sigma(2)), \ldots, f(\sigma(|\sigma|))>$ نمایش داده می‌شود.
- فرض کنید $R \subseteq X\times X$ یک رابطه بر روی X باشد. علاوه براین، برای هر $k\epsilon N$، $R^0=\{(x,x) \mid x \epsilon X\}$ و $R^{k+1} = \{(x,z)\epsilon X\times X|(x,y) \epsilon R^k \wedge (y,z) \epsilon R^k\}$ می‌باشد.

در جدول ۱-۱، هر رویداد مربوط به یک ردیف می‌باشد و هر رویداد را می‌توان با چند مشخصه تعریف نمود که این مشخصات در ستون‌های مختلف در جدول ۱-۱ نشان داده شده است. درحالی‌که سابقه رویدادهای مستخرج از سیستم‌های واقعی ممکن است فرمت‌های مختلف خروجی داشته باشد و رویدادهای درون این سابقه رویدادهای ممکن است حاوی انواع مختلف اطلاعاتی باشد، مفهوم یک رویداد با تعدادی مشخصه را به صورت عمومی در ذیل تعریف می‌کنیم:

تعریف ۱(رویداد، مشخصه). فرض می‌کنیم که L مجموعه تمام شناسه پرونده‌های ممکن باشد، A مجموعه‌ای از نام فعالیت‌های ممکن باشد، T حوزه زمان باشد، R مجموعه تمام نام منابع ممکن باشد، و D_x بازه مقدار صفت داده X باشد. همچنین فرض کنید ε فضای رویداد باشد که شامل مجموعه تمام رویدادهای ممکن می‌باشد. رویداد $e \epsilon \varepsilon$ می‌تواند شامل مشخصات مختلف باشد. در حوزه این پایان‌نامه، مشخصات زیر را تعریف می‌کنیم:

- $prop_{Case} \epsilon \varepsilon \rightarrow L$ مشخص‌کننده شناسه پرونده رویداد می‌باشد.
- $Prop_{Act} \epsilon \varepsilon \rightarrow A$ مشخص‌کننده نام فعالیت مربوطه می‌باشد.
- $Prop_{Time} \epsilon \varepsilon \nrightarrow T$ مشخص‌کننده برچسب زمانی یک رویداد می‌باشد.
- $Prop_{Res} \epsilon \varepsilon \nrightarrow R$ مشخص‌کننده نام منبع مجری رویداد می‌باشد.

- $Prop_x \in \varepsilon \nrightarrow D_x$ مشخص‌کننده مقدار برخی صفات داده X مربوط به رویداد مفروض می‌باشد.

جدول ۱-۱: سابقه رویداد سناریوی بیان‌شده. هر ردیف مربوط به یک رویداد می‌باشد و رویدادها با برچسب زمانی مرتب شده‌اند.

شماره پرونده	فعالیت	برچسب زمانی	مجری	داده
پرونده ۱	پاسخ به تماس‌های دفتر عمومی (A)	2007-03-07 11:08:24	کریس	
پرونده ۲	پاسخ به تماس‌های دفتر عمومی (A)	2007-03-07 11:09:05	پت کریگ	
پرونده ۱	پایان تماس (C)	2007-03-07 11:11:44	کری ولش	مشکل = خرابی سخت‌افزاری
پرونده ۲	پایان تماس (C)	2007-03-07 11:14:56	پت کریگ	مشکل = خرابی سخت‌افزاری
پرونده ۲	ثبت محصول ورودی (E)	2007-03-10 08:08:01		
پرونده ۲	تلاش برای تعمیر محصول (F)	2007-03-12 08:08:01	ایوو دی بوئر	
پرونده ۲	آگاهی نسبت به وضعیت تعمیر (D)	2007-03-17 17:10:05	ری اولی	
پرونده ۲	تست محصول تعمیرشده (G)	2007-03-23 16:09:15	کنی	
پرونده ۲	ارسال محصول تعمیرشده (H)	2007-03-24 07:04:05		
پرونده ۲	تکمیل ارسال (J)	2007-03-25 16:34:00	UPS-0987	SF Level = 2
…	…	…	…	…
پرونده ۳	ثبت محصول ورودی (E)	2007-03-10 08:08:01		
…	…	…	…	…

برای مثال، این مشخصات برای رویداد e که در اولین ردیف جدول ۱-۱ نمایش داده شده است، تعریف می‌شود: پرونده ۱ = $prop_{Case}(e)$، پاسخ به تماس‌های دفتر عمومی = $prop_{Act}(e)$، 2007-03-07 11:08:24 = $prop_{Time}(e)$ و کریس ولش =$prop_{Res}(e)$. توجه داشته باشید که $prop_{Problem}$ برای رویداد ردیف اول جدول ۱-۱ تعریف نشده است زیرا که این رویداد خاص در حوزه این تابع نیست.

در تعریف ۱، تمام مشخصات به‌جز $prop_{Case}$ و $prop_{Act}$، به‌عنوان توابع جزئی تعریف شده‌اند. توجه داشته باشید، ممکن است رویدادهایی وجود داشته باشد که مربوط به هیچ نمونه فرآیند یا فعالیتی نباشد اما برای مثال، ورود شخص خاصی را به سیستم ثبت می‌کند. چنین رویدادهایی را می‌توان برای محاسبه قابلیت استفاده منابع بکار برد. بنابراین، می‌توان از این رویدادها برای بدست آوردن تصویر جامعی از عملکرد فرآیند نیز استفاده کرد. گرچه، در این رساله، فرض بر اینست که هر رویداد مربوط به یک نمونه فرآیند و یک فعالیت می‌باشد.

حال، مفهوم سابقه رویداد را بیان می‌کنیم، که در آن رویدادها به یک دنباله خاص پیوند داده می‌شود. به طور خلاصه، یک رویداد سابقه مجموعه‌ای از دنباله‌ها می‌باشد و رویدادهای درون هر دنباله با یک ترتیب خاص قرارگرفته‌اند. علاوه بر این، هر رویداد در سابقه رویداد، منحصربه‌فرد می‌باشد و تنها به یک دنباله خاص پیوند زده می‌شود.

تعریف ۲ (دنباله، رویداد سابقه) یک دنباله، یک توالی رویدادها $\sigma = \langle e_1, e_2, \dots, e_n \rangle \in \varepsilon^*$ می‌باشد به طوریکه هر رویداد تنها یکبار ظاهر می‌شود و تمام رویدادها در دنباله دارای شناسه پرونده یکسان می‌باشد. برای مثال، به ازای هر $1 \leq i \leq j \leq n$، داریم $prop(\sigma(i)) = prop_{Case}(\sigma(j)) \wedge \sigma(i) \neq \sigma(j)$. C مجموعه‌ای از تمام دنباله‌های ممکن (شامل دنباله‌های جزئی) می‌باشد. سابقه رویدادها مجموعه‌ای از دنباله‌های $E \subseteq C$ به طوریکه هر رویداد دقیقاً یکبار در کل سابقه رویداد ظاهر شده باشد، برای مثال به ازای هر $\sigma_1, \sigma_2 \in E$ داریم: $set(\sigma 1) \cap set(\sigma 2) = \emptyset$ یا $\sigma 1 = \sigma 1$.

علاوه براین، نمادگذاری‌های زیر را برای راحتی تعریف می‌نماییم:

- $Event(E) \epsilon P(\varepsilon)$ مجموعه‌ای از رویدادهای موجود در سابقه رویداد مفروض می‌باشد، بطوریکه $events(E) = \{ e\epsilon\sigma \mid \sigma\epsilon E \}$.
- $\alpha(E) \epsilon B(A^*)$ یک رویداد سابقه ساده‌سازی شده می‌باشد، از آنجاییکه هر رویداد در E با نام فعالیت آن جایگزین می‌شود، برای مثال، $\alpha(E) = [prop_{Act}(\sigma) \mid \sigma \epsilon E]$.

گرچه تمام مشخصات رویدادهای موجود در سابقه رویداد در آنالیز فرآیندکاوی بکار می‌رود، اما الگوریتم‌های جریان کنترلی معمولاً برچسب‌های زمانی و داده‌های زمانی را در نظر نمی‌گیرد. این داده‌های زمانی معمولاً در سابقه رویدادهای واقعی موجود می‌باشد و روی ترتیب زمانی فعالیت‌های حقیقی انجام‌شده متمرکز می‌شوند. بنابراین، یک نمونه فرآیند را می‌توان به عنوان دنباله‌ای از فعالیت‌ها در نظر گرفت و یک رویداد سابقه را می‌توان به صورت مجموعه‌ای از دنباله رویدادهای مختلف و فرکانس آن‌ها ساده‌سازی کرد. جدول ۱-۲ چنین سابقه رویداد ساده‌سازی شده را نمایش می‌دهد.

جدول ۱-۲ : سابقه رویداد ساده‌سازی شده به‌عنوان مجموعه‌ای از دنباله‌های سابقه رویداد و فرکانس آن‌ها.

دنباله	تعداد نمونه‌ها
۵۰	ABC
۳۰۰	AC
۱۰	ACAC
۵	ACABC
۲۰	ACEFGHJ
۳۳	ACEFDGHJ
۱۷	ACEFDGHJ
۱۰۱	ACEFHJ

برای مثال، اگر $S=\alpha(E)$ سابقه رویداد ساده‌سازی شده برای سناریوی خدمات مشتری باشد، پرونده ۲ از جدول ۱-۱ با $\sigma=<A, C, E, F, D, G, H, J>\epsilon S$ در جدول ۱-۲ نشان داده می‌شود. علاوه بر این، فرکانس دنباله $\sigma = <A, C, E, F, D, G, H, J>\epsilon S$ برابر با ۳۳ می‌باشد که $S(\sigma)=33$ به معنی اینکه ۳۲ نمونه فرآیند دیگر دقیقاً همان ترتیب فعالیت‌های پرونده ۲ را داشته‌اند.

بررسی دقیق‌تر جدول ۱-۲ با اندکی دانش در حوزه فرآیندکاوی نشان می‌دهد که فرآیند خدمات مشتری امکان تولید دنباله‌های دیگر بسیاری را نیز دارد، که مشابه آن در سابقه رویداد وجود ندارد. برای مثال، مشتری می‌تواند برای دریافت اطلاعات در مورد وضعیت تعمیر پس از ارسال محصول تماس بگیرد (بین H و J). مشتری‌ها ممکن است بیش از دو بار با مرکز خدمات تلفنی تماس بگیرند و سپس دستور تعمیر و غیره را صادر کند. بنابراین، دنباله‌های مشاهده‌شده درون هر سابقه رویداد را باید نمونه‌برداری شده درنظر گرفت اما لزوماً رفتار کاملی ندارد: گرچه این سناریوها (مشتری دو بار تماس می‌گیرد و سپس به تعمیر ارسال می‌کند) در مدت زمانی مشاهده‌شده اتفاق نیفتاده است، دلیل بر هرگز اتفاق نیفتادن آن نیست.

مسئله را با تعدادی از دنباله‌های ممکن قابل تولید توسط فرآیند مفروض، بیشتر بسط می‌دهیم. برای مثال، تعداد 120 = !5 ترکیب احتمالی برای اجرای پنج وظیفه موازی و 3628800 = !10 ترکیب احتمالی برای ده وظیفه موازی وجود دارد. بنابراین، حتی اگر فرآیند در مدت زمان طولانی مورد نظارت قرار بگیرد، احتمال اینکه تمام ترکیبات ممکن مشاهده شود، بسیار بعید است. علاوه براین، اگر فرآیندی حاوی حلقه باشد، تعداد نامحدودی از دنباله‌ها ممکن می‌باشد. بنابراین، نباید انتظار داشت یک سابقه رویداد تمام دنباله‌های ممکن از فرآیند مفروض را به نمایش گذارد. درنتیجه، تکنیک‌های فرآیندکاوی برای تضعیف مفهوم کمال تلاش می‌کند، که حجم اطلاعات یک سابقه رویداد برای کشف مجدد فرآیند کافی باشد.

۱-۵ بیان مسأله

در فرآیندکاوی، به کشف، نظارت و بهبود فرآیندهای سازمانی از درون سابقه‌رویدادهای موجود در سامانه‌های اطلاعاتی می‌پردازد. فرآیند سازمانی، مجموعه‌ای از فعالیتها با هدف ارائه یک سرویس یا محصول می‌باشد که طبق یک روال مشخص به یکدیگر متصل هستند. هر رویداد در سابقه‌رویداد با

مجموعه‌ای از مشخصه‌ها اعم از اطلاعات زمانی مشخص می‌گردد. مدیریت و کنترل هوشمند فرآیندهای سازمانی بر اساس این اطلاعات زمانی، یکی از چالش‌های حوزه فرآیندکاوی می‌باشد.

کنترل زمانی، شامل عملیات‌هایی اعم از پیش‌بینی زمان باقی‌مانده پرونده‌ها در فرآیند، تنظیم زمان رسیدن یک نمونه‌فرآیند به یک وظیفه خاص در یک فرآیند و همچنین، ایجاد هم‌زمانی، تقدم و تأخر برای دو نمونه‌فرآیند مختلف می‌باشد.

از آنجاییکه در هرکدام از کارهای پیشین، پیش بینی چرخه زمانی با پارامترها و فاکتورهای محدودی مورد آزمایش قرار گرفته است. برای نمونه، Aalst و همکاران [۱۲, ۱۳] به جنبه صف نمونه فرآیندها توجه نکرده اند؛ Rogge-soltie و همکاران [۳] به جنبه داده نمونه فرآیندها توجه نکرده اند؛ Polato و همکاران [۱۴] به جنبه صف نمونه فرآیندها توجه نکرده اند. از این رو، صورت مسئله باز عمومی در کنترل زمانی، ارائه رویکرد پیش‌بینی زمانی با لحاظ حداکثر پارامترها و فاکتورها همانند صف، توزیع‌های احتمالاتی، رانش مفهوم، انشعاب توکن‌ها در نمونه‌فرآیندها، داده‌های نمونه‌فرآیندها، چندفرآیندی و غیره به صورت همزمان می‌باشد.

صورت مسئله باز تخصصی‌تر، کاوش وابستگی‌های پردازشی پنهان نمونه‌فرآیندها و تنظیم همزمانی روند نمونه‌فرآیندها برپایه رویکرد پیش‌بینی می‌باشد که این رساله در پی ارائه راهکار برای این مسائل باز می‌باشد. از آنجاییکه وابستگی‌های پردازشی در نمونه فرآیندها موجب افزایش چرخه زمانی کلی در فرآیند می‌شوند، ارائه راهکار جهت شناسایی، استخراج و تنظیم تأخیر و تعجیل نمونه‌فرآیندهای وابسته جهت کاهش چرخه زمانی و همچنین افزایش کیفیت سرویس‌دهی، به عنوان صورت مسئله باز تعریف می‌گردد.

۱-۶ متدولوژی

تعداد شرکت‌هایی که از سیستم‌های اطلاعاتی فرآیند-آگاه برای مدیریت فرآیندهای سازمانی خود بهره می‌برند، در حال گسترش می‌باشد. تمام این سیستم‌ها وقایع مربوط به هر فعالیت را در قالب

سابقه‌رویداد ذخیره می‌کنند. این سابقه‌رویدادها تحت جوانب مختلفی تحلیل و آنالیز می‌شوند. این رساله به مبحث پیش‌بینی، برنامه‌ریزی، تنظیم و کنترل زمانی اجرای نمونه‌فرآیندها در شرایطی که عوامل متعددی برروی روند نمونه فرآیندها تأثیرگذار می‌باشد، می‌پردازد. این عوامل در قالب مفروضات این رساله به صورت موردی در ذیل آمده است:

۱- سابقه رویداد حاوی مشخصات زمانی رویدادها می‌باشد.

۲- سابقه رویداد می‌تواند حاوی چند فرآیند باشد.

۳- فرآیندها در حین عملیات مدیریت پرونده‌ها، ممکن است تحت تأثیر رانش مفهوم قرار گیرند.

۴- فرآیندها دارای تعداد محدودی از منابع می‌باشد و منابع از نوع مصرفی نیستند.

۵- برای هر وظیفه‌ای در فرآیند، یک صف از پرونده‌ها در نظر گرفته‌شده است.

۶- برای جلوگیری از پیچیدگی محاسباتی، تعداد فرآیندهای با منابع مشترک، برابر با دو در نظر گرفته‌شده است.

۷- تعداد وظیفه‌ها در هر فرآیند، حداکثر ۳۰ در نظر گرفته‌شده است.

۸- نرخ ورود پرونده‌ها به فرآیندها، ثابت نیست و از توزیع پوآسن پیروی می‌کند.

۹- مدت‌زمان انجام وظیفه توسط هر منبع، ثابت نمی‌باشد و از یک تابع توزیع احتمال گاوسی پیروی می‌کند.

کنترل زمانی نمونه فرآیندها با استفاده از اطلاعات زمانی موجود در سابقه رویداد، با در نظر گرفتن منابع محدود یکی از چالش‌های این حوزه می‌باشد.کنترل زمانی، شامل عملیات‌هایی اعم از پیش‌بینی زمان باقی‌مانده پرونده‌ها در فرآیند، تنظیم زمان رسیدن یک نمونه‌فرآیند به یک وظیفه خاص در یک فرآیند و همچنین، ایجاد همزمانی، تقدم و تأخر برای دو نمونه‌فرآیند مختلف می‌باشد.

این رساله، با بررسی رویکردهای موجود برای کنترل زمانی نمونه‌فرآیندها، چهار رویکرد پیشنهادی برپایه تکنیک تجزیه مسیر و فرمول‌بندی تحلیلی مسیر با استفاده از مدل‌های ماشین بردار پشتیبان بدست‌آمده از داده‌های پیشین را به همراه یک رویکرد احتمالاتی با احتساب صف وظیفه‌ها جهت

پیش‌بینی چرخه زمانی باقیمانده نمونه فرآیندها ارائه می‌کند. پس از آن، مفروضات مسئله را با لحاظ رانش مفهوم به دنیای واقعی نزدیک‌تر می‌کنیم و عملیات پیش‌بینی چرخه زمانی نمونه فرآیندها را با فرض تغییر در مدل فرآیندی با ارائه یک رویکرد آماری مبتنی بر سیستم گذار حاشیه‌نویسی شده با خروجی ماشین‌های بردار پشتیبان فازی انجام می‌دهیم. درنهایت، یک وابستگی پردازشی پنهان میان جفت نمونه فرآیندهای وابسته از درون سابقه رویدادها به صورت رسمی تعریف و استخراج می‌کنیم و برای آن راهکاری با تنظیم و کنترل زمانی رویدادهای نمونه فرآیندها ارائه می‌نماییم. این رساله، در قالب سه فصل، رویکردهای پیشنهادی و آزمایش‌ها و نتیجه‌گیری خود را ارائه می کند.

۱-۷ نتیجه‌گیری

در این فصل، ابتدا مقدمه‌ای از مفاهیم فرآیند و فرآیندکاوی ارائه گردید و سپس هدف رساله تشریح داده شد. نقطه آغازین در مباحث فرآیندکاوی سابقه‌رویداد می‌باشد. سابقه‌رویداد می‌تواند حاوی مشخصات زمانی باشد. این مشخصات زمانی در غنی‌سازی مدل‌های فرآیندی بکار می رود. به کمک فرآیندهای غنی شده، وضعیت فرآیند را در آینده می توان پیش بینی نمود و برای رسیدن به یک وضعیت مطلوب برنامه‌ریزی نمود. در ادامه رساله در فصل دوم، کارهای پیشین مرتبط با بحث های زمانی فرآیند را شرح داده شده است. رویکردهای پیشنهادی رساله به همراه معرفی پایگاه داده ها و آزمایش‌ها نیز در سه فصل سوم و چهارم و پنجم ارائه شده است.

فصل۲ : پیشینه تحقیق

۲-۱ پیش‌بینی وضعیت نمونه‌فرآیند

مرور ادبیات تحقیق نشان می‌دهد که تعداد مقالات چاپ شده در حوزه پیش‌بینی و نظارت فرآیند در سالهای اخیر به صورت چشم‌گیری افزایش یافته است. با تحلیل آماری، مشخص گردید که ۱۴ مقاله از ۵۵ مقاله مرتبط در بین سالهای ۲۰۰۶ تا ۲۰۱۳ منتشر شده اند و ۴۱ مقاله از آنها در چهار سال اخیر منتشر شده است (۱۷ مقاله ژورنالی و ۳۸ مقاله کنفرانسی).

مهم ترین معیار در دسته بندی تکنیک های پیش بینی و نظارت، نوع پیش بینی می باشد [۱۵]. در ادامه، برخی از پژوهش های انجام شده را در سه دسته کمی، کیفی و پیش بینی فعالیت بعدی نشان می‌دهیم.

۲-۱-۱ پیش بینی کمی

می توان پژوهش های مربوط به پیش بینی های کمی را براساس خروجی پیش بینی به دو دسته تقسیم نمود:

- پیش بینی زمانی؛
- پیش بینی هزینه ای.

پیش بینی زمانی. این دسته از پژوهش ها برروی جنبه زمانی نمونه فرآیندها تمرکز می‌کنند که گروه بسیار بزرگی هم هستند. در پژوهش Aalst و همکاران [۱۶]، محققان یک مجموعه از رویکردهایی را ارائه می‌دهند که در آن سیستم‌های گذار برپایه یک انتزاع مفروض از رویدادها در سابقه‌رویداد، با اطلاعات زمانی مستخرج از سابقه‌رویداد حاشیه‌نویسی می‌شوند. به صورت خاص، اطلاعات در مورد زمان سپری شده، زمان انتظار و زمان باقیمانده در هر حالت از سیستم گذار گزارش می‌شود. پس آن، از اطلاعات جهت پیش بینی زمان تکمیل دنباله های جاری بهره برده می‌شود. توسعه های دیگر این رویکرد در پژوهش های [۱۴, ۱۷] ارائه شده است که محققان از تکنیک‌های یادگیری ماشین جهت

حاشیه‌نویسی سیستم‌های گذار بهره می‌برند. به صورت جزئی تر، در پژوهش Polato و همکاران [۱۷]، سیستم های گذار با مدل های یادگیری ماشین همانند بیز ساده و مدل های رگرسیون بردار پشتیبان حاشیه نویسی شده اند. در پژوهش دیگر Polato و همکاران [۱۴]، دو رویکرد جدید برپایه ماشین های بردار پشتیبان به همراه یک رویکرد پیشنهادی دیگر برپایه جنبه داده‌ای توکن‌ها را با توسعه بر پژوهش [۱۷] انجام می‌دهد. محققان همچنین سه رویکرد پیشنهادی را در دوحالت ایستا و غیرایستا از نظر شرایط رانش مفهوم مورد آزمایش قرار دادند. بهبودهای دیگری بر رویکرد Aalst و همکاران [۱۳]، که هدف آنها نیز پیش بینی زمان باقیمانده یک دنباله جاری می باشد، پژوهش های [۱۸, ۱۹] می باشد. در این پژوهش ها، سیستم گذار حاشیه‌نویسی‌شده با رویکرد خوشه‌بندی پیش‌بینی‌کننده متن-محور ترکیب می‌شود. ایده اصلی خوشه‌بندی پیش‌بینی‌کننده آنست که سناریوهای متفاوتی با پیش‌بینی‌کننده‌های مختلف در آن می توان تعریف نمود. علاوه بر آن، می‌توان از اطلاعات متنی به همراه اطلاعات جریان کنترلی [۱۹] و یا اطلاعات کنترل و منبعی [۱۸] جهت پیش بینی بهره برد.

رویکرد دیگری برپایه استخراج مدل درخت‌های دنباله در پژوهش Ceci و همکاران [۲۰] جهت پیش بینی زمان اتمام و تعیین فعالیت بعدی نمونه فرآیند جاری ارائه شده است. مدل درخت دنباله مشابه با رویکرد خوشه‌بندی پیش‌بینی‌کننده، خوشه‌بندی دنباله‌ها را به کمک دنباله‌های مشابه از فعالیت‌ها انجام می‌دهد و یک مدل پیش‌بینی‌کننده را برای هر نود از درخت دنباله با بهره‌گیری از اطلاعات داده نمونه فرآیند می‌سازد. در پژوهش Rogge-solti و همکاران [۲۱]، از شبکه های پتری تصادفی با گذارهای با توزیع نرمال (GDT-SPN) برای پیش بینی زمان باقیمانده یک پرونده بهره می برند. به صورت خاص، رویکرد، مدل فرآیند تصادفی را که پیشتر استخراج شد را به همراه دنباله جاری و اطلاعات دیگری از وضعیت فرآیند در لحظه جاری به عنوان ورودی می‌گیرد تا پیش بینی زمان باقیمانده را به عنوان خروجی برگرداند. در پژوهش [۳]، Rogge-solti و همکاران، زمان سپری شده از رویداد آخر را استخراج می‌کنند تا پیش‌بینی‌های دقیق‌تری برای زمان باقیمانده و همچنین تخمین احتمال تجاوز از یک مهلت زمانی ارائه کند.

در رویکردی متفاوت با رویکردهای فوق‌الذکر Van Dongen و همکاران در پژوهش [۲۲]، پیش‌بینی‌ها را تنها براساس سابقه رویداد انجام می دهند. به صورت جزئی تر، آنها رویکردی جهت پیش‌بینی چرخه زمانی باقیمانده یک پرونده با استفاده از رگرسیون غیرپارامتریک و استفاده از طول مدت زمان فعالیت‌ها و رویدادها به همراه داده‌های هر پرونده توسعه داده اند. در پژوهش [۲۳, ۲۴]، بهبودی بر رویکرد برپایه خوشه‌بندی متنی [۱۸, ۱۹] ارائه شده است و برروی یکی از محدودیت‌های رویکردهای برپایه سیستم گذار تمرکز می‌کند که نیازمندی به تحلیلگر برای انتخاب تابع انتزاع را نشان می‌دهد و این مشکل را با جایگزینی آن با الگوریتم های رگرسیون استاندارد حل می نماید. علاوه براین، در پژوهش [۲۳]، مؤلفه خوشه‌بندی در رویکرد بهبود داده شد تا راهکاری جهت حل مسائل دقت و مقیاس پذیری آن ارائه کند. در پژوهش Pandey و همکاران [۲۵]، مدل های مخفی مارکوف برای انجام پیش بینی مدت زمان باقیمانده بکار گرفته می‌شود. یک ارزیابی مقایسه‌ای نشان می‌دهد که مدل‌های مخفی مارکوف نتایج دقیق تری را نسبت به سیستم‌های گذار حاشیه‌نویسی‌شده و مدل‌های رگرسیون ارائه می‌کند. در پژوهش Senderovich و همکاران [۲۶]، پیش بینی زمان اتمام دنباله‌های جاری را با پیش‌بینی برپایه ویژگی میان‌پرونده‌ای ارائه می‌کند. رویکردهای پیشنهادی نه تنها از اطلاعات مربوط به پرونده جاری بهره می‌برد، بلکه از وضعیت پرونده‌های جاری دیگر نیز به جهت پیش بینی استفاده می‌نماید. کدگذاری پیشنهادی موجب بهبود نتایج اعمال شده به دو سابقه‌رویداد را نشان می‌دهد. نوعی از پیش‌بینی که بسیار به پیش‌بینی چرخه زمانی نزدیک می‌باشد، پیش‌بینی تأخیر پرونده جاری می‌باشد. در پژوهش Senderovich و همکاران [۲۷]، تئوری صف برای پیش‌بینی تأخیرهای محتمل در اجرای فرآیندهای سازمانی بکارگرفته می‌شود. مؤلفان آن پژوهش رویکردهایی را برای ارتقاء رویکردهای سنتی برپایه سیستم‌های گذار همانند رویکرد Aalst و همکاران [۱۳] ارائه می‌کنند که اثرات صف را دخیل می‌کند.

پیش بینی هزینه. دومین دسته از پژوهش ها برروی پیش بینی هزینه متمرکز می شود. در این دسته نیز از پژوهش های برپایه مدل فرآیندی می توان نام برد، همانند پژوهش Tu و همکاران [۲۸]

که در آن پیش‌بینی‌های هزینه از طریق مدل فرآیندی مستخرج و گراف دنباله حاشیه‌نویسی‌شده با هزینه‌ها و اطلاعات دیگری همانند تولید، حجم و زمان انجام می شود.

۲-۱-۲ پیش بینی های کیفی

دسته دوم از رویکردهای پیش‌بینی مقادیر کیفی را پیش‌بینی می‌کند. دو نوع اصلی پیش بینی در این دسته شامل:

- پیش بینی ریسک؛
- پیش بینی خروجی کیفی.

پیش‌بینی های ریسک. برای نمونه Conforti و همکاران [۲۹]، تکنیکی را ارائه می‌کنند که با دریافت مدل فرآیندی به عنوان ورودی ریسک‌های فرآیند را کاهش می‌دهد. ایده اصلی، پشتیبانی از مشارکت‌کننده‌های فرآیند در اتخاذ تصمیم‌های ریسک-آگاه به واسطه پیش‌بینی‌های مربوط به اجرای فرآیند می‌باشد. درخت‌های تصمیم از سابقه اجراهای فرآیندهای پیشین بواسطه اطلاعات مربوط به داده‌ها، منابع، دفعات اجرا ساخته می‌شوند. درخت های تصمیم پس از آن پیمایش می شوند و پیش‌بینی‌های مربوط به ریسک به کاربر بازگردانده می‌شود. Conforti و همکاران در دو پژوهش دیگر [۳۰, ۳۱] توسعه هایی از پژوهش خود [۲۹] را ارائه کردند. در پژوهش [۳۱]، چارچوب تصمیم‌های ریسک-آگاه به سناریوهایی که در آن چندین پرونده همزمان اجرا می‌شوند، توسعه داده می شود. به صورت خاص، در راهکار حل پرونده های مختلف از یک فرآیند، از تکنیکی با رویکرد برنامه نویسی خطی با اعداد صحیح جهت محاسبه الگوریتم انتصاب بهینه منابع به وظایف آماده اجرا استفاده می‌شود. Conforti و همکاران در پژوهش [۳۰] پژوهش خود در [۲۹] را به نحوی توسعه دادند که دیگر اجراهای فرآیند در آن به صورت مجزا در نظر گرفته نمی شود و اطلاعات پیرامون ریسک‌ها به صورت اتوماتیک به پرونده های جاری مشابه از همان فرآیند به صورت زمان-واقعی منتشر می‌شود تا پیش‌بینی‌های

زودهنگام انجام شود.

Metzger و همکاران [۳۲] سه رویکرد متفاوت در پیش بینی تخطی از قیود در پرونده ها را بررسی می‌کنند: یادگیری ماشین، ارضاء قیود و تجمیع کیفیت سرویس‌دهی. محققان علاوه بر اعلام نتایج قابل قبول برای سه رویکرد مذکور، تفاوت ها را شناسایی کردند و اقدام به ادغام آنها کردند. نتایج بر پایگاه‌داده واقعی نشان داد که ترکیب این رویکردها موجب افزایش دقت پیش‌بینی می‌شود.

پژوهش های دیگر در حوزه پیش‌بینی ریسک، مدل های فرآیندی را در نظر نمی‌گیرند. برای نمونه، Pika و همکاران در پژوهش خود [۳۳] پیش بینی‌ها را در مورد ریسک‌های مربوط به زمان در فرآیندها با تشخیص و بکارگیری نشانگرهای ریسک فرآیندی (به عنوان نمونه، زمان اجرای فعالیت یا تکرار چندباره فعالیت) با اعمال متدهای آماری به سابقه‌رویداد انجام می‌دهند. نشانگرها سپس توسط تابع پیش‌بینی تجمیع می‌شوند که موجب تخمین احتمال تخطی مهلت‌ها می‌شود.

Pika و همکاران [۳۴] نیز کار پیشین خود را با معرفی یک متد برای تنظیم نشانگرهای ریسک فرآیندی توسعه دادند. رویکرد آنها از نتایج پرونده های به اتمام رسیده، مناسب ترین آستانه ها را برای نشانگرهای ریسک فرآیندی فرا می‌گیرد و سپس با احتساب مشخصات فرآیند مفروض، دقت را افزایش می‌دهد.

۲-۱-۳ پیش بینی فعالیت آتی

سومین گروه از پژوهش‌های مرتبط با پیش‌بینی، با فرض وجود فعالیت‌های گذشته، به پیش‌بینی دنباله فعالیت‌های آینده و داده‌های آنها می‌پردازد [۱۴, ۳۵-۳۸]. Polato و همکاران [۱۴] یک رویکرد پیشنهادی برای پیش‌بینی دنباله فعالیت‌های آتی یک پرونده جاری با تکیه بر سیستم گذار حاشیه‌نویسی‌شده مبتنی بر داده که توسعه ای از سیستم گذار حاشیه نویسی شده Aalst و همکاران [۱۳] می‌باشد.

رویکردهای دیگر همانند [۳۵, ۳۶, ۳۸] از شبکه های مصنوعی تکرارشونده با حافظه طولانی کوتاه

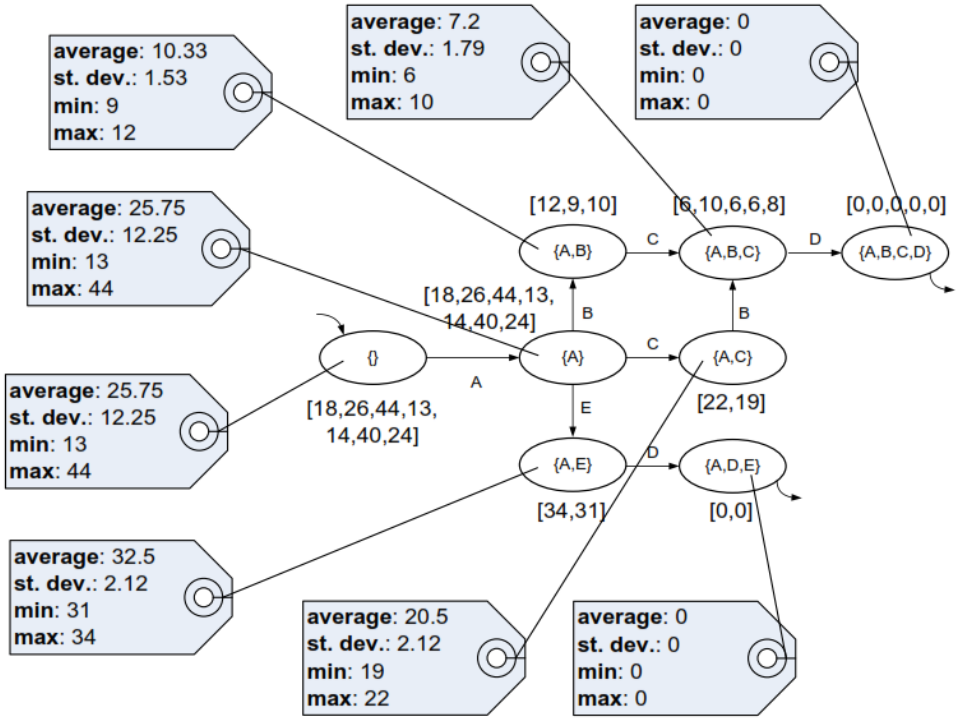
مدت بهره گرفتند. به طور خاص، در پژوهش‌های [۳۶, ۳۸]، از شبکه های مصنوعی تکرارشونده با دو لایه مخفی آموزش داده شده به روش انتشار به عقب استفاده شده است، درحالیکه در پژوهش [۳۵]، یک حافظه طولانی کوتاه مدت به همراه یک کدگذاری برپایه فعالیت‌ها و برچسب‌ها برای انجام پیش‌بینی‌ها و برچسب‌های آنان استفاده می‌شود. در پژوهش [۳۷]، از دانش پیشین a-priori جهت انجام پیش‌بینی دنباله فعالیت‌های آتی بهره می‌برد. بدین جهت از رویکرد حافظه طولانی کوتاه مدت با در نظر گرفتن دانش مفروض در مورد پیشروی یک پرونده جاری در آینده استفاده می‌کنند.

۲-۲ کنترل زمانی

سازمان‌ها برای ارائه به‌موقع خدمات و محصولات، نیاز به مدیریت زمانی فرآیندهای سازمانی خود دارند. بنابراین، آن‌ها وضعیت جاری فرآیند را در نظر می‌گیرند تا زمان وقوع انحرافات نامطلوب را تشخیص دهند و مطابق آن واکنش نشان دهند. مدیریت زمانی نیازمند تخمین طول مدت انجام یک نمونه فرآیند و همچنین وضعیت پرونده‌ها و منابع در یک زمان مشخص در آینده می‌باشد. انگیزه اصلی مدیریت زمانی، تضمین رضایت مشتری با افزایش حجم محصولات و خدمات ارائه‌شده در مهلت‌های زمانی مشخص‌شده می‌باشد. بنابراین، ضرورت وجود متدهایی برای جلوگیری از زمان‌های سرویس‌دهی و زمان‌های انتظار بیش‌ازحد احساس می‌شود. مدیریت و پیش‌بینی را می‌توان در ابعاد مختلفی همانند زمان، هزینه و وضعیت منابع تعریف نمود. برای مثال، مدیریت و پیش‌بینی زمانی رسیدن یک پرونده خاص به یک وضعیت هدف را می‌توان تعریف نمود یا می‌توان محاسبه ریسک تجاوز از مهلت زمانی تعیین‌شده برای یک وضعیت هدف را تعریف نمود.این رساله، یک چارچوب را برای پیش‌بینی و مدیریت زمان اجرای پرونده‌ها در یک یا چند فرآیند با استفاده از اطلاعات به‌دست‌آمده از نظارت بر فرآیند پیشنهاد می‌دهد.

بسیاری از کارهای مرتبط، اساس پیش‌بینی را برپایه مشاهدات پیشین و یا براساس اعتقادات اولیه قراردادند. کارهای مرتبط با پیش‌بینی براساس سری‌های زمانی نیز به صورت گسترده وجود دارد. سری‌های زمانی سنجش‌های پریودیکی از داده‌ها را لحاظ می‌کند. برای مثال، مقدار نهایی روزانه سهام یک شرکت در بازار بورس و یا ارقام فروش هفتگی یک محصول یک سری زمانی می‌باشد. متدهای بسیاری برای ایجاد مدل‌های مناسب برای فرآیند تولیدکننده چنین داده‌هایی وجود دارد. مروری از این روش‌ها در [۳۹] آمده است. آنالیز روندها و نقاط تغییر در سری‌های زمانی در [۴۰] یافت می‌شود اما به حوزه این رساله مربوط نمی‌شود.

یکی از روش‌های مدیریت فرآیندها، مدل‌سازی فرآیندها توسط مدل‌های SPN می‌باشد که آنالیز زمان-گسسته این مدل‌ها توسط van der Aalst [۴۱] و همچنین رساله دکتری Reijers [۴۲] بررسی شده است. در پژوهش دیگری، Van der Aalst و همکاران [۱۳] از اطلاعات موجود در سابقه رویداد برای پیش‌بینی طول مدت زمان باقی‌مانده انجام پرونده بر اساس مدت‌های زمانی مشاهده‌شده در گذشته استفاده می‌کند. آن‌ها با طراحی یک سیستم گذر حالت، مدت زمان باقی‌مانده برای هر حالت ملاقات شده را از دنباله‌های موجود در سابقه رویداد جمع‌آوری می‌کنند. نهایتاً هنگامی که یک پرونده انجام می‌شود، زمان باقی‌مانده در هر حالت، با میانگین‌گیری از طول مدت باقی‌مانده از پرونده‌های پیشین بر اساس سابقه رویداد پیش‌بینی می‌شود. رویکرد ارائه‌شده توسط Rogge-Solti [۳] نیز مشابه آن است با این تفاوت که داده‌ها و منابع را در نظر نگرفته است اما از مدل‌های GDT_SPN به جای سیستم‌های گذر استفاده می‌کند، که باعث دقت بیشتری در هنگام استفاده از کنترل موازی می‌شود. از کارهای ارائه شده در [۱۳] و [۴۳] می‌توان برای محک روش پیشنهادی رساله استفاده نمود.

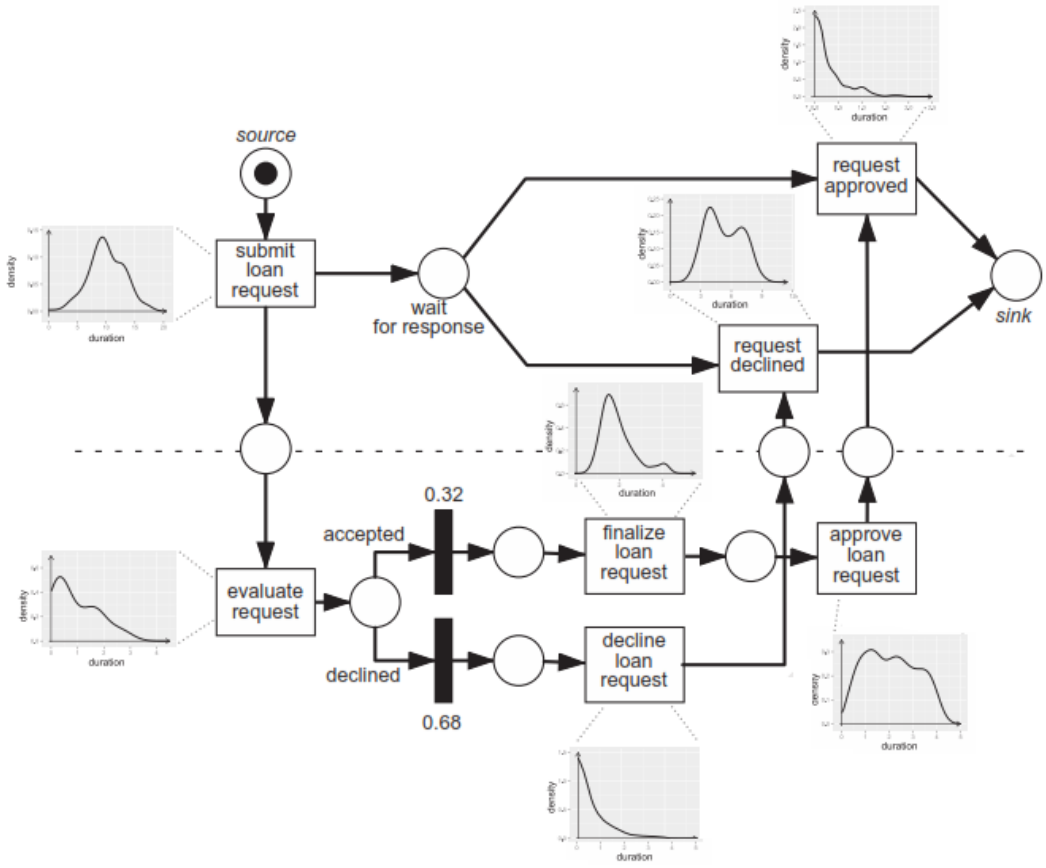


شکل ۲-۱: سیستم گذار حاشیه‌نویسی شده ارائه شده توسط Aalst و همکاران [۱۳]

Folino و همکاران [۱۹]، بهبودی را با توسعه [۱۳] برپایه خوشه‌بندی مبتنی بر پیش‌بینی ارائه دادند. آن‌ها از اطلاعات متنی اضافی یک دنباله (برای مثال، بار کاری جاری سیستم) برای خوشه‌بندی استفاده می‌کنند. ایده پیشنهادی برپایه گروه‌بندی دنباله‌های مشابه می‌باشد و پیش‌بینی را برای دنباله‌های جدید که مشابه با دنباله‌های پیشین می‌باشد، انجام می‌دهد. آن‌ها از پیش‌بینی برای اعلام هشدار در صورت تجاوز از مهلت تعیین‌شده استفاده می‌کنند. ترکیب این روش با روش [۳] نیز می‌تواند در قالب یک رساله مطرح شود.

Hwang و همکاران [۴۴] و همچنین Zheng و همکاران [۴۴] نیز در بحث پیش‌بینی عملکرد، نتایج آزمایش‌ها خود را ارائه دادند. آن‌ها از فرمول‌هایی برای محاسبه کیفیت سرویس‌دهی همانند مدت زمان مورد انتظار در ترکیب سرویس‌ها استفاده می‌کنند. فرض رایج در این کارها آنست که سرویس‌ها از بلوک‌های ساختاری (بلوک‌های تک ورودی-تک خروجی) استفاده می‌کند. این کار را می‌توان با کار ارائه شده توسط Kiepuszewski و همکاران [۴۵] مقایسه نمود. روش‌های

پیشنهادی را می‌توان برای فرآیندهای سازمانی نیز به کار برد. در این رساله با حذف فرض ساختار بلوکی در این کار، مدل‌های پیچیده‌تری را پوشش می‌دهیم و نمونه فرآیندهای در حال اجرا را نیز در نظر می‌گیریم.



شکل ۲-۲: مدل ارائه‌شده توسط Andreas Rogge-soltie **برای حل احتمالاتی فرآیند** [۳]

مقاله دیگر نزدیک به این رساله توسط Wombacher و Lacob ارائه شده است [۴۶]. در آن کار، مؤلفان سابقه رویدادهایی از فرآیندهای غیرساختارمند برای پیش‌بینی مدت زمان فعالیت‌ها استفاده می‌کنند. آن‌ها متوسط مدت زمان فعالیت‌ها را برای پیش‌بینی استفاده می‌کنند، اما آن‌ها از اطلاعات زمانی در زمان اجرا استفاده نمی‌کنند.

با فرض وجود توزیع نرمال برای مدت زمان فعالیت‌ها، Anklesaria و همکاران [۴۷]، زمان اتمام را به کمک شبکه‌های PERT تخمین می‌زنند. شبکه‌های PERT برای مدل کردن پروژه‌هایی با

مجموعه‌ای از فعالیت‌های مورد نیاز که با هم وابستگی دارند، استفاده می‌شود. برای مثال، یک فعالیت تنها در صورتی آغاز می‌شود که دو فعالیت پیشین آن به اتمام رسیده باشد. انشعاب مسیر و حلقه‌ها در این مدل‌ها در نظر گرفته نمی‌شود، گرچه، مدت زمان باقی‌مانده پروژه با طولانی‌ترین مسیر از طریق شبکه تعیین می‌شود. Anklesaria و همکاران تأثیرات همبستگی‌های میان مسیرهای مختلف را برای افزایش دقت خطای پیش‌بینی بررسی می‌کنند.

اخیراً، Sederovich و همکاران داده‌کاوی را بر روی شبکه‌های صف‌بندی بدست آمده از سابقه رویدادها پیشنهاد دادند تا عملیات پیش‌بینی زمان باقی‌مانده را انجام دهند [۴۸]. درصورتی که منابع در مدل در نظر گرفته شود، کار آن‌ها منجر به دقت بالا خواهد شد اما نیازمند اطلاعاتی در مورد زمان انتظار در صف، زمان ورود و زمان خروج از صف می‌باشد. Leoni و همکاران [۴۹]، با ارائه یک رویکرد کلی‌تر در قالب یک چارچوب، مسئله پیش‌بینی در فرآیندهای سازمانی را به یک مسئله داده‌کاوی نگاشت دادند. به صورت دقیق‌تر، آن‌ها هر رویداد یک سابقه رویداد را به جدولی با چند مشخصه مرتبط نگاشت دادند تا روابط و رگرسور ها فراگرفته شود. آن‌ها داده‌های بیشتری را برای امر فراگیری بهتر بکار گرفتند. گرچه، آن‌ها کنترل موازی را در کار خود در نظر نمی‌گیرند. متد ارائه‌شده در [۱۳] نیز نزدیک به این رساله می‌باشد، اما نیاز به مجموعه داده بسیار بزرگی برای افزایش دقت پیش‌بینی زمان باقی‌مانده نیاز دارد. یک شبیه‌سازی نیز توسط Rozinat و همکاران برای تصمیم‌گیری عملیاتی پیشنهاد و بکار گرفته شد [۱۱, ۵۰] . ایده اصلی آن، راه‌اندازی یک محیط شبیه‌سازی که وضعیت جاری را بررسی کند و یک شبیه‌سازی کوتاه‌مدت را با شروع از این وضعیت با پارامترهای شبیه‌سازی مختلف آغاز می‌کند. کاربرد شبیه‌سازی آن‌ها برای پوشش تصمیم عملیاتی می‌باشد و بر روی عملکرد سراسری فرآیندهای سازمانی متمرکز می‌باشد.

حوزه مدیریت فرآیند سازمانی ریسک-آگاه، یک حوزه جذاب برای محققان می‌باشد [51] . یک بررسی جامع اخیر پیرامون این موضوع توسط Suriadi و همکاران [۵۲] فراهم شد. درحالیکه

کارهای پیشین از جنبه کیفی به مسئله می‌نگرد، اخیراً رویکردهای کمی پیشنهاد شده است. کارهای Pika و همکاران [۵۳]، Goluch و همکاران [۵۴] ، و Conforti و همکاران [۵۵] بسیار به رویکرد ما نزدیک هستند.

Pika و همکاران [۵۳]، اندیکاتورهایی را برای ریسک تجاوز از مهلت تعریف می‌کنند. آن‌ها به دنبال الگوهایی همانند مدت زمان غیرنرمال فعالیت، می‌گردند و از آن اطلاعات برای پیش‌بینی تأخیر یک پرونده استفاده می‌کنند. Goluch و همکاران یک نمونه اولیه برای شبیه‌سازی ریسک-آگاه با یک تمرکز خاص بر روی تهدیدهای موجود برای منابع -همانند قطعی برق برای منبع ماشینی ، یا بیماری منبع انسانی- پیشنهاد می‌دهند [۵۴]. در مقابل، ما برروی ریسک تجاوز از مهلت زمانی تمرکز می‌کنیم. علاوه بر این، Kang و همکاران [۵۶] ادعای نظارت بر فرآیند سازمانی را در زمان اجرا دارند. رویکرد آن‌ها برپایه دسته‌بندی دنباله‌های پیشین با تکنیک‌های داده‌کاوی همانند ماشین‌های بردار پشتیبان به دو دسته دنباله‌های درست و نادرست می‌باشد. هدف آن‌ها پیش‌بینی و دسته‌بندی نمونه‌های جاری، برای مثال آن‌هایی که احتمال تجاوز به مهلت زمانی در آن‌ها محتمل می‌باشد. رویکرد آن‌ها تنها فرآیندهای ترتیبی را دریافت می‌کند، گرچه برچسب‌های زمانی رویدادها در پیش‌بینی درنظر گرفته نمی‌شود.

کار ارائه‌شده توسط Leitner و همکاران [۵۷] نمونه‌های جاری را در نظر می‌گیرد. آن‌ها از رگرسیون برای مدت‌زمانی میان دو نقطه در فرآیند استفاده می‌کنند. از پیش‌بینی‌ها سپس برای تشخیص تخطی طبق توافق‌نامه سرویس‌دهی استفاده می‌شود.

مدیریت زمان یکی از مؤلفه‌های مهم در مدیریت فرآیندهای بر پایه گردش کار می‌باشد. جنبه‌های مهم مدیریت زمان شامل برنامه‌ریزی زمانی اجرای فرآیند گردش کار، تخمین مدت اجرای گردش کار، پرهیز از تخطی‌های مهلت زمانی، ارضاء تمام قیود زمانی خارجی همانند قیود تاریخ مشخص، حدود بالا و پایین برای بازه‌های زمانی میان فعالیت‌ها می‌باشد.

مدیریت زمان و قیود زمانی در طراحی و مدیریت فرآیندهای سازمانی اهمیت دارد. درنتیجه، مدیریت زمان باید بخشی از هسته اصلی عملکرد باید که توسط سیستم‌های گردش کار برای کنترل چرخه حیات فرآیندها تأمین شود. در زمان ساخت که طرح‌های گردش کار توسعه داده می‌شوند و تعریف می‌شوند، طراحان مدل گردش کار نیاز به ابزاری برای بازنمایی جنبه‌های وابسته به زمان فرآیندهای سازمانی دارند (مدت‌های زمانی فعالیت، قیود زمانی میان فعالیت‌ها و غیره) و امکان‌سنجی آن را بررسی می‌کند (قیود زمانی که با یکدیگر تناقض نداشته باشد). در زمان اجرا که نمونه‌های گردش کار نمونه‌سازی و اجرا می‌شود، مدیرهای فرآیند نیازمند مکانیزم‌های پیش کنشی برای دریافت هشدارهای مربوط به تخطی‌های قیود زمانی ممکن می‌باشد. مشارکت‌کننده‌های گردش کار نیازمند اطلاعاتی در مورد اولویت‌های وظایف منتسب به آن‌ها برای مدیریت لیست‌های کاری آن‌ها می‌باشد. اگر از یک قید زمانی تخطی شود، سیستم گردش کار باید قادر به راه‌اندازی مقابله با استثناء برای دستیابی مجدد به وضعیت سازگار نمونه گردش کار برسد. مهندسان فرآیندهای سازمانی نیازمند اطلاعاتی در مورد مدت‌زمان واقعی اجرای فرآیندها هستند تا به‌وسیله آن فرآیندهای سازمانی را بهبود دهند. کنترل‌کنندگان و مدیران کیفیت نیازمند اطلاعاتی در مورد زمان شروع فعالیت و مدت‌زمان اجرای آن می‌باشند.

در حال حاضر، مدیریت زمان در سیستم‌های گردش کار محدود به شبیه‌سازی‌های فرآیندی (برای تشخیص گلوگاه‌ها، آنالیز مدت‌های زمانی اجرای فعالیت‌ها، غیره) می‌شود. در زمان ساخت، ما یک مدل گردش کاری را بررسی می‌کنیم تا مشخص شود که آیا برنامه‌ریزی اجرای قیود زمانی تعیین‌شده را نقض می‌کند یا خیر. نتیجه یک گراف فعالیت زمانی است که شامل بازه‌های مهلت زمانی برای هر فعالیت می‌باشد. در زمان نمونه‌سازی فرآیند، گراف فعالیت زمانی را تغییر می‌دهیم تا بازه‌های مهلت‌های زمانی و مشخصات زمانی را شامل شود. در زمان اجرا، ما به صورت پویا گراف زمانی را برای فعالیت‌های باقی‌مانده مجدداً محاسبه می‌کنیم تا نظارت بر روی ارضاء قیود زمانی باقی‌مانده انجام گیرد.

۲-۳ نتیجه گیری

بسیاری از روش‌های موجود، توزیع‌های احتمالاتی مدت زمان انجام کار را که تأثیر بسزایی در دقت پیش‌بینی زمانی دارد، در نظر نگرفته‌اند یا در بهترین حالت یک توزیع نرمال در نظر گرفته‌اند. Rogge-solti که در پایان نامه دکتری خود، توزیع‌های احتمالاتی را در نظر گرفته است، از جنبه‌های دیگر از جمله صف پرونده‌ها، منابع و وابستگی ها ضعف دارد. هدف این رساله، ارائه یک چارچوب آماری فرآیند-آگاه برای پیش‌بینی، کنترل و تنظیم نمونه فرآیندها می‌باشد که مفروضات بیشتری همانند صف، منابع و رانش مفهوم را در نظر می‌گیرد و مسئله را جامع‌تر و دقیق‌تر حل می‌نماید.

فصل ۳ : پیش‌بینی زمان باقیمانده نمونه فرآیندها

۳-۱ مقدمه

در مبحث فرآیندهای سازمانی، پیش‌بینی زمانی نمونه‌فرآیندها مقدمه عملیات‌های پیشگیرانه آینده می‌باشد. لذا هر اندازه میزان دقت پیش‌بینی وضعیت نمونه‌فرآیندها براساس مدل‌های مستخرج از سابقه‌رویداد افزایش یابد، میزان دقت عملیاتی همانند تنظیم همزمانی، پیش‌بینی گلوگاه، پیش‌بینی زمان باقیمانده در حالت رانش مفهوم و دیگر عملیات آینده نیز افزایش می‌یابد.

در روش‌های پیش‌بینی زمانی مرسوم، معمولاً با یک عملیات رگرسیون سراسری، زمان باقیمانده چرخه زمانی هر نمونه‌فرآیند را برای هر دنباله جزئی فرآیند، پیش‌بینی می‌کنند، در این فصل، چهار رویکرد مبتنی بر تجزیه مسیر به همراه یک رویکرد احتمالاتی با احتساب صف وظیفه ها، جهت عملیات پیش‌بینی زمانی پیشنهاد می‌شود. منظور از تجزیه مسیر، تجزیه عملیات پیش‌بینی و انجام آن در سطح فعالیت‌های مسیر آینده نمونه‌فرآیند می‌باشد و سپس تجمیع پاسخ‌های عملیات‌های پیش‌بینی از هرکدام می‌باشد؛ عملیات تجزیه مسیر را می‌توان به رویکرد تقسیم و غلبه تشبیه نمود.

در رویکرد اول، برپایه آنالیز تحلیلی مسیر، برای پیش‌بینی در محل انشعابات و همچنین طول مدت زمان فعالیت‌ها، برای تمام اندازه پیشوندهای دنباله های جزئی، یک مدل یادگیری ماشین در نظر گرفته می‌شود. در رویکرد دوم برپایه آنالیز تحلیلی مسیر، برای پیش‌بینی در محل انشعابات و همچنین طول مدت زمان فعالیت‌ها، برای هر اندازه پیشوند دنباله‌های جزئی، یک مدل یادگیری ماشین مجزا در نظر گرفته می‌شود. در رویکرد سوم برپایه آنالیز تحلیلی مسیر، برای پیش‌بینی در محل انشعابات و همچنین طول مدت زمان فعالیت‌ها، از اعداد ثابت میانگین مقادیر سابقه‌رویداد استفاده می‌کند. در رویکرد چهارم برپایه آنالیز تحلیلی مسیر، از ترکیب مدل‌های یادگیری ماشین و مقادیر میانگین به صورت ترکیبی بکار گرفته می‌شود. در رویکرد پنجم پیشنهادی از توزیع های احتمالاتی و احتساب صف انتظار وظیفه‌ها، برای پیش‌بینی زمان باقیمانده نمونه فرآیند استفاده می‌شود.

در ادامه فصل، در بخش ۳-۲، مفهوم تجزیه مسیر شرح داده می‌شود؛ چهار رویکرد پیشنهادی برپایه

تجزیه مسیر در بخش ۳-۳ ارائه می‌شود؛ در بخش ۳-۴، یک رویکرد احتمالاتی با احتساب صف‌های وظیفه‌ها جهت پیش‌بینی زمان باقیمانده پیشنهاد داده خواهد شد. در بخش ۳-۵، پایگاه داده ها و روش‌های محک برای ارزیابی رویکردهای پیشنهادی ارائه می‌شود. در بخش ۳-۶، نتایج ارزیابی رویکردهای پیشنهادی برروی پایگاه داده‌ها ارائه و تشریح داده خواهد شد. در نهایت، جمع بندی فصل در بخش ۳-۷ انجام خواهد شد.

۳-۲ تجزیه مسیر نمونه‌فرآیند

در تجزیه مسیر نمونه فرآیند، برای تخمین مسیرآینده نمونه فرآیند، مسیر را به فعالیت های موجود درآن مسیر تقسیم می‌کنیم. برای نمونه، با استفاده از تکنیک تجزیه مسیر، می توان چرخه زمانی متوسط برای کل فرآیند را با استفاده از متوسط طول مدت زمان فعالیت‌های موجود در تمام مسیر فرآیند محاسبه نمود. از تکنیک تجزیه مسیر می توان برای محاسبه هزینه متوسط کل نمونه فرآیند با استفاده از هزینه اجرای هر فعالیت محاسبه نمود و همچنین می‌توان میزان خطای کل یک فرآیند را با فرض میزان خطای هر فعالیت آن فرآیند محاسبه نمود [۵۸]. توجه داشته باشید که در تجزیه مسیر، مؤلفه‌های ساختار مدل فرآیندی اهمیت دارند که همان وظیفه ها می‌باشند.

چرخه زمانی هر فعالیت شامل زمان پردازش آن فعالیت، به همراه زمان انتظار پیش از اجرای فعالیت می باشد. زمان پردازش به زمانی اشاره دارد که منابع صرف انجام کار می‌کنند. از طرف دیگر زمان انتظار، نسبتی از چرخه زمانی می‌باشد که هیچ کاری برای پردازش انجام نمی‌شود. این زمان ممکن است شامل زمان انتقال اطلاعات نمونه فرآیند میان منابع فرآیند نیز باشد، برای مثال ممکن است اسناد به صورت فیزیکی توسط اداره پست رد و بدل شود. در بسیاری از فرآیندها، زمان انتظار نسبت قابل توجهی از چرخه زمانی کلی را تشکیل می‌دهند. برای مثال ،این حالت ممکن است زمانی که حجم انجام کارها بسیار است، رخ دهد. برای مثال، اگر در فرآیند برآورد نیازمندی‌ها در یک اداره، مسئول

رسیدگی به پرونده‌های خرید و فروش، تمام درخواست‌ها را به صورت دسته‌جمعی در انتهای یک روز کاری بررسی کند، زمان انتظار نمونه فرآیندها افزایش می‌یابد.

برای درک درست راجع به تکنیک تجزیه مسیر، یک نمونه فرآیند خطی با تعدادی مؤلفه از رویدادها همانند شکل ۳-۱ الف را آغاز می‌کنیم. هر مؤلفه یک چرخه زمانی T_i دارد. از آنجاییکه مؤلفه‌ها یکی پس از دیگر اجرا می‌شوند، به صورت شهودی می‌توان گفت که چرخه زمانی CT از یک فرآیند خطی با N مؤلفه، مجموع چرخه های زمانی مؤلفه‌ها می‌باشد:

$$CT = \sum_{i=1}^{N} T_i \qquad (۱-۳)$$

حال فرض کنید در یک فرآیند، یک نقطه تصمیم میان N مؤلفه دو به دو منحصر به فرد، و با دروازه XOR نمایش داده می شود. در این حالت، متوسط چرخه زمانی فرآیند بدین صورت محاسبه می شود:

$$CT = \sum_{i=1}^{N} p_i . T_i \qquad (۲-۳)$$

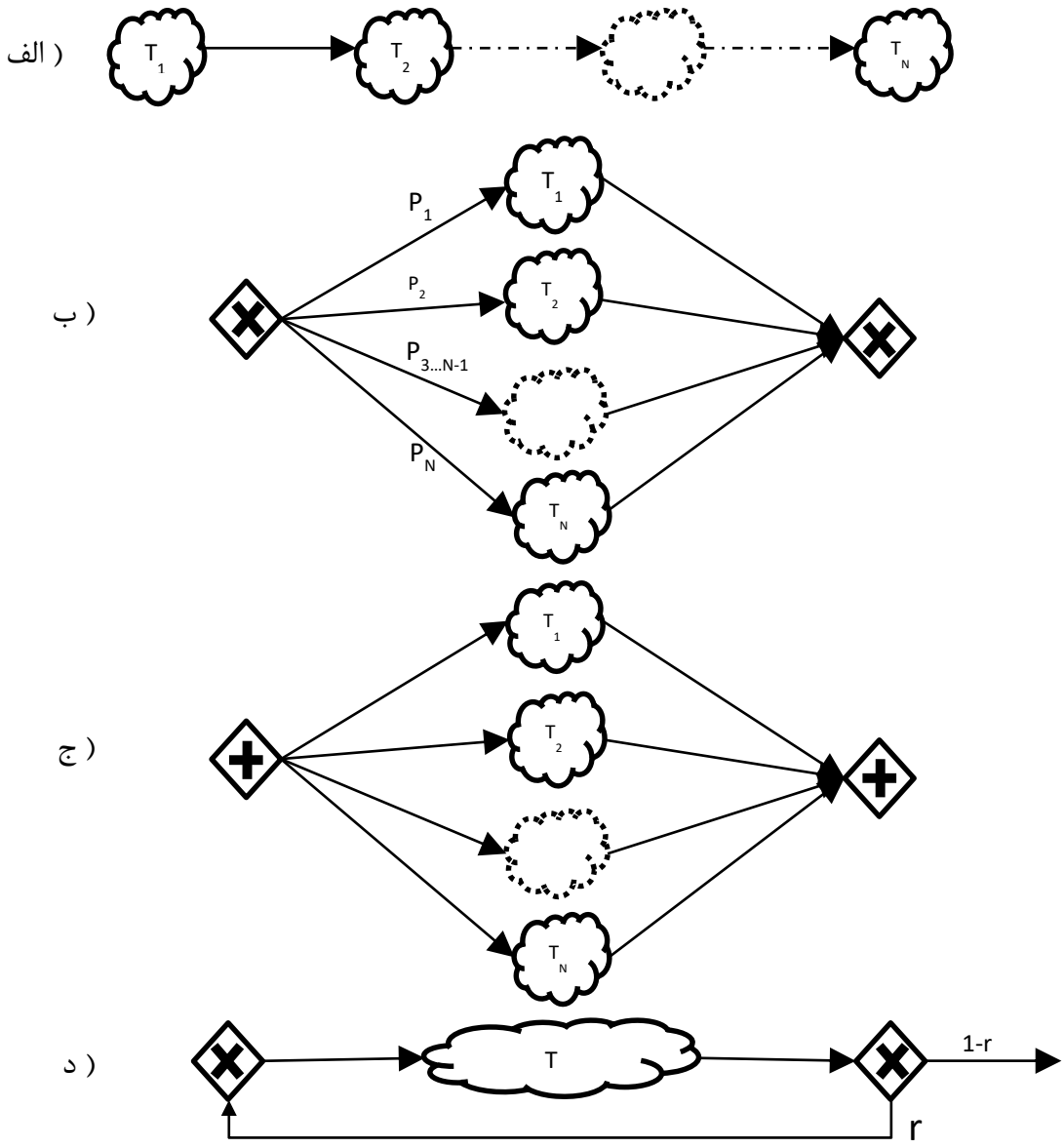
که در آن p_i احتمال هر انشعاب را نمایش می دهد، که برای نمونه فرکانس مربوط به هر انشعاب i از هر دروازه مشخص می‌شود.

در حالت دروازه مسیرهای موازی، که فعالیت ها به صورت همزمان همانند شکل ۳-۱ ج اجرا می‌شوند، چرخه زمانی ترکیبی از آهسته ترین مؤلفه از میان چند مؤلفه بدست می آید که به صورت:

$$CT = \max_{i=1\ldots n} T_i \qquad (۳-۳)$$

توجه داشته باشید که در محاسبه چرخه زمانی متوسط برای وظیفه‌های موازی، چرخه زمانی وظیفه با بیشینه چرخه زمانی متوسط همیشه بزرگتر از چرخه زمانی وظیفه های دیگر با چرخه زمانی متوسط کوچکتر می‌باشد. این فرض در پژوهش [۵۹] و کارهای پیشین دیگری که برروی کیفیت سرویس دهی در سرویس‌های ترکیبی کار می‌کنند، نیز بنا شده است. اگر این فرض در نظر گرفته نشود، باید فرمول‌های مختلفی را براساس توزیع های احتمالاتی وظیفه های موازی بکار گرفته شود. در پژوهش

[۶۰]، فرمول هایی برای محاسبه میانگین بیشینه چند متغیر تصادفی آمده است. اما در عمل، فرمول فوق الذکر تخمین بسیار خوبی برای وظیفه های موازی ارائه خواهد داد.



شکل ۳-۱: الگوهای مرسوم مدل فرآیندی الف) ترتیبی، ب) بلوک XOR، ج) بلوک AND،

در الگوی تکرارشونده، یک وظیفه از یک فرآیند ممکن است چندین بار تکرار گردد؛ این تکرار ممکن است برای مثال به علت مردود شدن و مورد پذیرش نبودن جهت ادامه باشد. این حالت که تکرارکاری نامیده می‌شود، در شکل ۳-۱د نشان داده شده است. وظیفه اصولاً یک بار اجرا می‌شود؛ سپس، آن

وظیفه ممکن است به احتمال r که احتمال تکرارکاری می‌باشد، دفعات دیگر اجرا شود. فرض بر آنست که با هربار تکرار اجرای وظیفه، احتمال r تغییر نمی‌کند و احتمال تکرارکاری برای بار N-ام وابسته به مقدارN نمی‌باشد، تعداد دفعاتی که وظیفه چندبار انجام می‌شود، از توزیع هندسی با مقداری برابر $1/(1-r)$ تبعیت می‌کند. بنابراین، چرخه زمانی متوسط وظیفه در این حالت به صورت ذیل است:

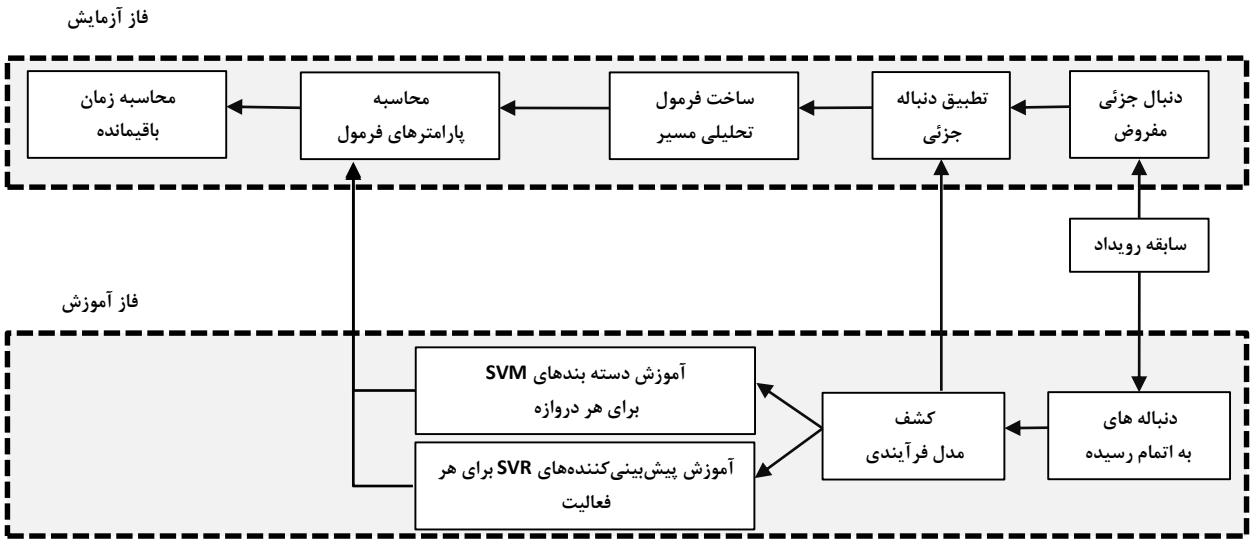
$$CT = \frac{T}{1-r} \qquad (۴-۳)$$

علاوه بر چرخه زمانی، تجزیه مسیر را می توان برای محاسبه معیارهای دیگر نیز به کار برد. برای نمونه، با فرض وجود هزینه هر فعالیت، هزینه فرآیند را می‌توان تقریباً به همان صورتی که چرخه زمانی را محاسبه می‌کنیم، محاسبه نمود. به طور خاص، هزینه یک مجموعه فعالیت به طور سری، مجموعه هزینه های این فعالیت های سری می باشد. تنها تفاوت میان محاسبه چرخه زمانی و محاسبه هزینه مربوط به رویکرد حل دروازه های AND می باشد. هزینه دروازه AND همانند شکل همانطور که در شکل ۳-۱ج نشان داده شده است، بیشینه هزینه انشعابات دروازه AND نمی‌باشد، بلکه مجموع هزینه انشعابات می باشد. علت این کار آنست که پس از آنکه دروازه AND پیمایش می شود، هر انشعاب تا AND الحاقی باید تا انتها اجرا گردد و سپس هزینه های این انشعابات تجمیع می گردد [۵۸].

با کشف ساختار مدل های فرآیندی، هر فرآیند را با چهار نوع شرح داده شده توصیف می‌کنیم و از معادلات فوق الذکر برای تخمین معیار پیش بینی زمانی بهره می‌بریم. گرچه در حالت مدل فرآیندی غیرساختاری یا زمانی که مدل شامل سازه های دیگری علاوه بر AND و XOR داشته باشد، متد محاسبه معیارهای عملکردی پیچیده تر خواهد شد.

۳-۳ رویکردهای پیشنهادی برپایه تجزیه مسیر

در این بخش، رویکردهای پیشنهادی برپایه تجزیه مسیر را جهت پیش‌بینی زمان باقیمانده نمونه فرآیند ارائه می‌کنیم. در این رویکردها، از دنباله‌های موجود مربوط به نمونه‌فرآیندهای گذشته برای کشف مدل فرآیندی استفاده می‌کنیم. هنگامی که مدل فرآیندی استخراج گردید، مجموعه فعالیت‌های فرآیند و همچنین مجموعه نقاط تصمیم فرآیند را مشخص می‌کنیم. سپس، از تکنیک‌های یادگیری ماشین برای



شکل ۳-۲: چارچوب کلی رویکرد پیشنهادی

پیش‌بینی چرخه زمانی هر فعالیت و همچنین پیش‌بینی احتمال انشعابات در هر نقطه تصمیم استفاده می‌کنیم. برای افزایش سرعت در زمان اجرا، این عملیات ها را به حالت برون‌خط انجام می‌دهیم.

زمانی که یک نمونه فرآیند درحال اجرا وجود دارد، دنباله جزئی آن را با مدل فرآیندی کشف شده انطباق می دهیم تا وضعیت جاری نمونه فرآیند در مدل فرآیند تعیین شود. سپس، درخت فرآیند را که در روی مدل فرآیند بدست می‌آید، پیمایش می‌کنیم تا برای مسیر وضعیت جاری تا به انتهای فرآیند، فرمول تحلیلی استخراج نماییم. این فرمول تحلیلی، با استفاده از قواعد فوق الذکر، پیش‌بینی زمانی نمونه‌فرآیند را با استفاده از چرخه های زمانی فعالیت ها و احتمالات انشعابات انجام می‌دهد. مقادیر این مؤلفه‌ها از دسته‌بندهای بردار پشتیبان و پیش‌بینی‌کننده های بردار پشتیبان آموزش حاصل می‌شود. نهایتاً، پس از ارزیابی فرمول، مقدار پیش بینی را از فرمول پارامتریک تحلیلی بدست می‌آوریم.

۳-۳-۱ کشف فرآیند

رویکرد پیشنهادی از مدل فرآیندی به عنوان ورودی استفاده می‌کند. از آنجاییکه مدل فرآیندی همیشه دردسترس نیست ویا ممکن است با سابقه رویداد تطبیق نداشته باشد، نیاز به کشف مدل فرآیندی از سابقه رویداد می باشد. درکشف مدل فرآیند می‌توان از الگوریتم هایی همچون الگوریتم آلفا [61]، الگوریتم اکتشاف‌گر [62]، الگوریتم استخراج‌گر ژنتیک [63] بهره برد. در این رویکرد ما از الگوریتم آلفا برای کشف مدل فرآیند از سابقه‌رویداد استفاده می‌کنیم. مدل ساختاری را می توان به صورت درخت فرآیند نیز بازنمایی کرد. درخت فرآیند، درختی است که هر برگ آن با یک فعالیت برچسب‌گذاری شده است و هر گره داخلی آن یک عملگر جریان کنترلی می‌باشد: توالی، انتخاب انحصاری، انتخاب غیرانحصاری، موازی و یا تکرار.

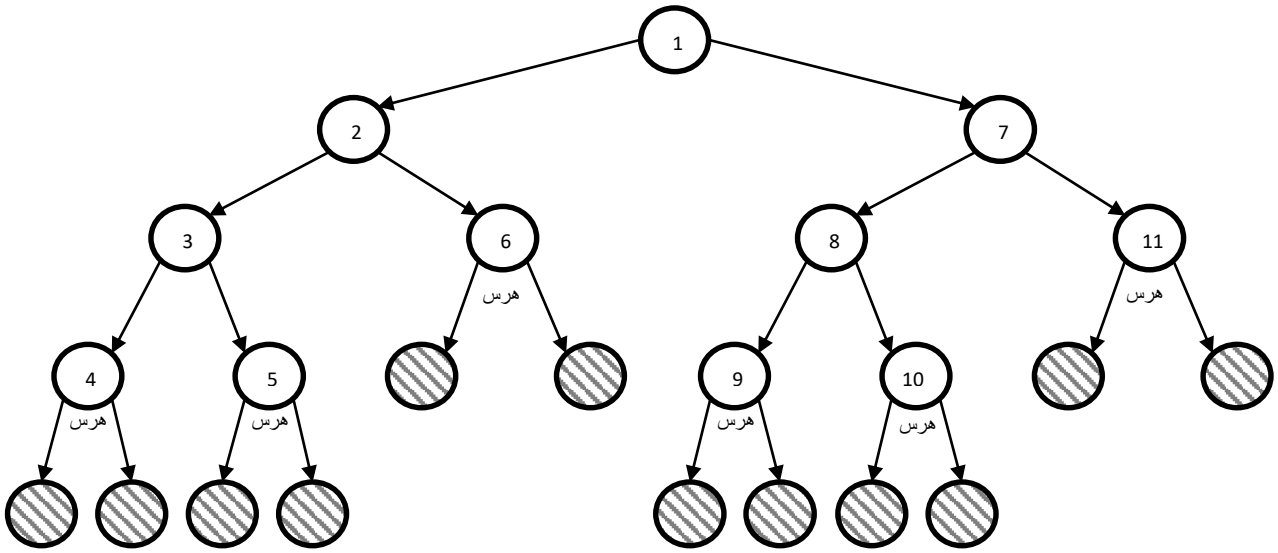
۳-۳-۲ انطباق دنباله های جزئی برروی مدل فرآیند

برای پیش‌بینی زمان باقیمانده چرخه زمانی یک دنباله جزئی مفروض، باید وضعیت جاری دنباله را برروی مدل فرآیندی مشخص نماییم. جهت تعیین وضعیت جاری دنباله جزئی بر روی مدل فرآیندی، باید یک عملیات پیشروی هماهنگ به صورت گام به گام با شروع از ابتدای مدل فرآیندی هماهنگ با ابتدای دنباله جزئی تا زمانی که دیگرگامی برای پیشروی در وظایف دنباله جزئی باقی نماند. عملیات نگاشت اجزای دنباله جزئی برروی اجزای مدل فرآیندی، عملیات تنظیم[۱] مدل فرآیندی و دنباله نامیده می‌شود. پیشروی در مدل فرآیندی باید بحث توکنیزه شدن نمونه فرآیندها را نیز در نظر بگیرد. برای مثال، در یک مدل BPMN، حرکت توکن ورودی برروی دروازه تقسیم XOR منجر به تولید یک توکن می‌شود که در یکی از انشعابات دروازه حرکت می‌کند. در حالت دیگر، حرکت برروی دروازه تقسیم AND، منجر به تولید چند توکن مجزا می شود که در همه انشعاب دروازه حرکت می‌کنند. در هر حرکت پیشروی در عملیات تنظیم، یک اندیس برروی دنباله جزئی پیش می‌رویم و همچنین از طرف

[۱] Alignment

دیگر، یک نشانه گذاری توکن ها را بروی مدل فرآیندی پیش می‌بریم. در رویکرد پیشنهادی از الگوریتم عقب‌گرد، استفاده می‌کنیم که با پیمایش اول-عمق گره‌های درخت فرآیند را پیمایش می‌کند. در این الگوریتم در هر گره، بررسی می‌کنیم که آیا عملیات تنظیم تا آن گره، کافیست و یا اینکه باید به سمت عمق ادامه دهیم. اگر نیاز به ادامه به سمت عمق باشد، مجموعه‌ای از گره های فرزند (حالت های بعدی) از آن گره را تولید می‌کنیم و مسیر را عمقی ادامه می دهیم؛ در غیر اینصورت، در آن گره متوقف می‌شویم، هرس شاخه را از آن گره انجام می دهیم و عملیات عقب گرد را به گره والد انجام می دهیم تا از آنجا شاخه‌های دیگر درخت را پیمایش نماییم. این روش، فضای حالت را مؤثرتر از جستجوی فراگیر[1] جستجو می‌کند.

$Function\ Alignment(\ node\ n)$
$if\ Reject(n)\ then\ return$
$if\ Complete(n) then$
$output(n)$
$for\ n_i : Children(n)\ do\ Explore(n_i)$



شکل ۳-۳: کاربرد الگوریتم عقب گرد در تطبیق دنباله با درخت فرآیندی

[1] Brute Force

در الگوریتم فوق، استخوان بندی کلی روش تنظیم نمایش داده می‌شود. هنگامی که توکن در یک گره فعالیت برروی مدل فرآیندی وجود دارد و همچنین برچسب گره با برچسب رویداد در اندیس جاری دنباله یکسان باشد، آنگاه از تابع children استفاده می‌کنیم چون به احتمال قوی، برچسب گره فرزند با برچسب رویداد در اندیس بعدی یکسان می‌باشد. هنگامی که توکن برروی مدل فرآیندی نیاز به عقب‌گرد داشته باشد و پیمایش های دیگر برروی گره ناکارآمد باشد، از تابع reject برای هرس شاخه درخت فرآیند استفاده می‌کنیم. در نهایت، تابع Complete شرایط حداقلی یک گره را برای عملیات تنظیم قابل قبول بررسی می‌کند. اگر شرایط برقرار باشد پیمایش متوقف می‌شود و بهترین گره باز می‌گردد.

۳-۳-۳ فرمول‌بندی تحلیلی مسیر نمونه فرآیند

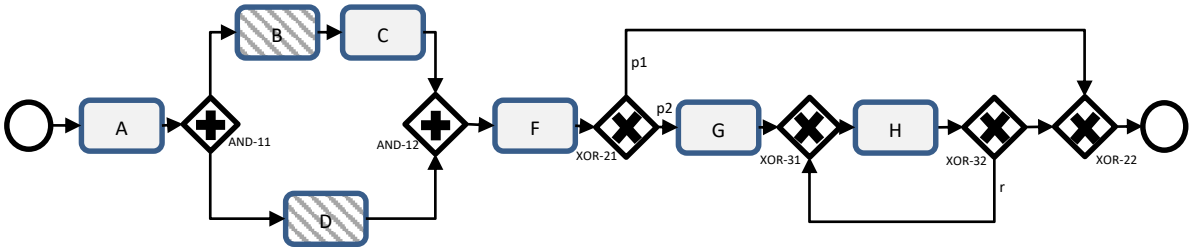
پس از تعیین وضعیت جاری یک نمونه‌فرآیند، مسیرآینده نمونه فرآیند را تا اتمام فرآیند جهت بدست آوردن فرمول های تحلیلی مسیر، پیمایش می کنیم.

به عنوان مثال، مدل فرآیندی ساده ترسیم شده در شکل ۳-۴ را ملاحظه نمایید. با اعمال فرمول‌های تحلیلی بخش ۳-۲، چرخه زمانی متوسط این فرآیند به صورت زیر تجزیه می‌شود:

$$\boldsymbol{CT = T_A + \max(T_B + T_C, T_D) + T_F + p2\left(T_G + \frac{T_H}{1 - r}\right)} \qquad (۵-۳)$$

توجه داشته باشید که یکی از انشعابات دروازه **XOR21** تهی می باشد و بنابراین در چرخه زمانی سهمی ندارد؛ بنابراین، تنها انشعاب احتمال **p2** در معادله آمده است.

اجزای فرمول که شامل چرخه‌های زمانی هر فعالیت و احتمالات انشعاب و تکرارکاری می باشد، را می‌توان به عنوان میانگین مقدار سابقه رویداد در نظر گرفت. گرچه، از آنجاییکه ما نمونه فرآیندهای



شکل ۳-۴: یک نمونه مدل فرآیندی که در آن وضعیت جاری یک نمونه فرآیند، هاشورزده نشان داده شده است.

درحال اجرا را مورد بررسی قرار می‌دهیم، از اطلاعات موجود در پیشوند نمونه فرآیند برای پیش بینی مؤلفه های فوق استفاده می کنیم.

فرض کنید دنباله جزئی $\boldsymbol{hd(\sigma) = < A, D, B >}$ موجود می باشد. با اجرای مجدد دنباله برروی مدل فرآیندی مفروض، به همان صورت که در بخش پیشین توضیح داده شد، وضعیت جاری را که دو توکن در حالت های **B** و **D** در بلوک **AND** وجود دارد، بدست می آوریم. جهت تعیین چرخه زمانی باقیمانده **hd(σ)** با پیمایش مدل از وضعیت جاری تا به نقطه اتمام فرآیند فرمول تحلیلی زیر حاصل می شود:

$$\boldsymbol{CT_{rem} = 0 + \max(0 + T_C, 0) + T_F + p2\left(T_G + \frac{T_H}{1 - r}\right) =} \qquad (۶-۳)$$

$$\boldsymbol{= T_C + T_F + p2(T_G + \frac{T_H}{1 - r})}$$

از آنجاییکه فعالیت های **A**، **B** و **D** تاکنون اجرا شده اند، در چرخه زمانی باقیمانده مشارکت ندارند. بنابراین، آنها بخشی از فرمول تحلیلی نیستند. تمام جمله های ریاضی دیگر با استفاده از داده درون **hd(σ)** پیش‌بینی را انجام می‌دهند.

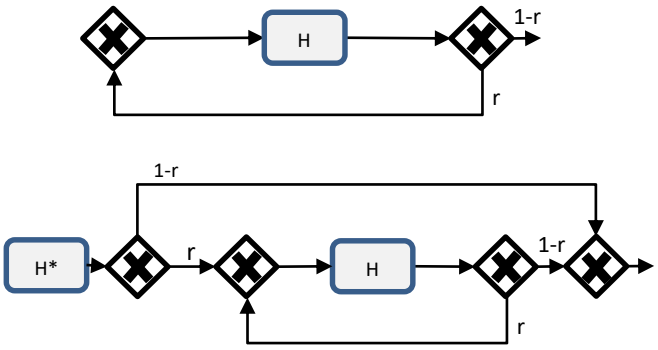
به طور مشابه اگر وضعیت جاری از خود انشعاب **XOR** گذر کند و به تجمیع **XOR** نرسد، احتمالات انشعاب آن نیاز به پیش بینی ندارد. درعمل، احتمال انشعاب تکمیل‌شده را به یک و احتمال دیگر انشعابات را به صفر قرار می‌دهیم.

هنگامی که با حلقه های تکرارکاری در مسیر نمونه فرآیند مواجه می شویم، شرایط پیچیده‌تری رخ می‌دهد. در این حالت، ما دفعاتی از تکرار حلقه را از حلقه خارج می کنیم و به مسیر ترتیبی می افزاییم (همانند شکل ۳-۵). به طور خاص، رخدادهای از پیش انجام شده مؤلفه تکرار کاری را از رخدادهای

آتی جدا می کنیم و رخدادهای پیشین را از حلقه خارج می کنیم. فرض کنید یک دنباله جزئی $hd(\sigma) = < A, D, B, C, F, G, H >$ وجود دارد. با توجه به مدل فرآیندی شکل ۳-۴، وظیفه **H** تاکنون یک بار رخ داده است و به احتمال **r** ممکن است تکرار نیز شود؛ در صورت عدم تکرار از حلقه تکرار کاری خارج می‌شود. برای حفظ عملکرد مدل فرآیند اصلی، میان اولین رخداد **H** از حلقه و رخدادهای آتی یک دروازه **XOR** قرار می‌دهیم. یکی از انشعابات حاوی حلقه تکرار کاری از رخدادهای آینده با احتمال **r** می باشد، درحالیکه انشعاب دیگر گزینه‌ای برای بدون انجام حلقه **H** مسیر را ادامه می دهد. بنابرین چرخه زمانی کل بخش را میتوان به صورت زیر تجزیه کرد:

$$CT_H = T_{H'} + r\frac{T_H}{1-r} \qquad (۷-۳)$$

که در آن $T_{H'}$ به چرخه زمانی رخدادهای پیشین **H** اشاره می کند. ازآنجاییکه این رخدادها در گذشته رخ داده است، زمان واقعی انجام آنها موجود است و نیازی به پیش‌بینی نیست.



شکل ۳-۵: تجزیه دفعات تکرار حلقه تکرار کاری

۳-۳-۴ پیش بینی زمان نمونه فرآیند

ما از فرمول های تجزیه مسیر که در بخش پیشین ارائه گردید، جهت محاسبه زمان باقیمانده چرخه زمانی یک پرونده به شرط وجود: (الف) تخمینی از مدت زمان هر فعالیت از فرآیند و (ب) تخمینی از احتمال هر انشعاب پس از گذرگاه XOR استفاده می‌کنیم.

برای انجام عملیات پیش بینی برپایه تجزیه مسیر چهار رویکرد زیر را پیش می گیریم:

۱. ***رویکرد پیش بینی کننده تک مدله:*** در این رویکرد، از مدل‌های پیش‌بینی برای طول مدت زمان هر فعالیت و برای انشعاب‌های شرطی استفاده نمود. در این رویکرد براساس مدل، یک مدل پیش بینی کننده بردار پشتیبان برای تمام اندازه‌های مختلف پیشوند دنباله ها در نظر گرفته می‌شود.

۲. ***رویکرد پیش بینی کننده چندمدله:*** در این رویکرد، از مدل‌های پیش‌بینی برای طول مدت زمان هر فعالیت و برای انشعاب‌های شرطی استفاده نمود. در این رویکرد نیز پیش‌بینی بر اساس مدل می باشد با این تفاوت که چند مدل پیش بینی کننده بردار پشتیبان در نظرگرفته می‌شود بطوریکه به ازای هر طول مختلف پیشوند دنباله‌ها، یک مدل پیش‌بینی‌کننده در نظر گرفته می‌شود.

۳. ***رویکرد پیش بینی کننده متوسط:*** در این رویکرد از مدت زمان متوسط هر فعالیت و میزان انشعاب متوسط برای انشعاب‌های شرطی استفاده می‌شود. این روش را پیش‌بینی متوسط می‌نامیم.

۴. ***رویکرد پیش بینی کننده ترکیبی:*** با ترکیب رویکرد براساس مدل و رویکرد انتخاب متوسط، زمانی که تعداد موارد آموزشی برای مدل پیش‌بینی یک فعالیت مفروض A کمتر از N_0 باشد، متوسط طول مدت هر فعالیت و متوسط انشعابات شرطی را همانند روش پیش‌بینی متوسط استفاده می‌کنیم. در غیراینصورت، از پیش‌بینی‌کننده‌های براساس مدل استفاده می‌کنیم. در این روش ترکیبی، مقدار N_0 را اهمیت بسیاری دارد و تنظیم آن باید به درستی صورت گیرد.

منطق استفاده از پیش بینی‌کننده‌های ترکیبی، آنست که درصورتیکه ابعاد فضای ویژگی بزرگ باشد، تعداد مشاهدات جهت آموزش یک مدل ممکن است کافی نباشد، بنابراین این پیش‌بینی‌ها ممکن است نامطمئن و ناپایدار باشد [۶۴]. این اتفاق ممکن است زمانی رخ دهد که چرخه‌های زمانی مربوط به فعالیت‌های نادر می‌باشد، یا انشعاب احتمالات دروازه ها به ندرت استفاده شود. علاوه بر

این، اگر پیشوند دنباله جزئی بسیار کوتاه باشد، ممکن است اطلاعات کافی برای پیش بینی چرخه زمانی برخی فعالیت‌ها و احتمالات انشعابات دروازه‌ها وجود نداشته باشد، بخصوص آنهاییکه نزدیک به انتهای فرآیند می باشند. در چنین مواردی، می توان از متوسط چرخه زمانی فعالیت‌ها و متوسط احتمالات به جای مدل پیش‌بینی استفاده نمود.

برای پیش بینی طول مدت زمان هر فعالیت درون مدل فرآیندی، یک مدل رگرسور بردار پشتیبان جداگانه برای آن فعالیت آموزش می‌دهیم و همچنین برای هر انشعاب درون مدل فرآیندی، یک مدل دسته بند بردار پشتیبان جداگانه در نظر می‌گیریم. در آموزش مدل دسته بند بردار پشتیبان، انشعابات را از صفر به بالا برای هر کلاس دسته‌بندی شماره‌گذاری می‌کنیم. مدل‌های پیش‌بینی بردار پشتیبان برای هر فعالیت نیز با پیشوندهای $hd^k(\sigma); 2 \leq k \leq |\sigma|$ از تمام دنباله های σ در مجموعه آموزشی آموزش داده می‌شوند. توجه داشته باشید که آموزش و پیش بینی را دقیقاً پس از اولین رویداد انجام نمی‌دهیم، زیرا هنوز داده‌های ناکافی برای پیش‌بینی در اختیار داریم.

۳-۳-۵ ساخت بردار ویژگی

برای آموزش پیش‌بینی‌کننده‌های بردار پشتیبان، نیاز به ساخت بردار ویژگی به طول ثابت از روی دنباله‌های جزئی مجموعه آموزش می‌باشد. از آنجایی که پیشوندهای دنباله های $\boldsymbol{\sigma}$ در مجموعه آموزشی به صورت $\boldsymbol{hd^k(\sigma); 2 \leq k \leq |\sigma|}$ متغیر می‌باشد، می توان دو رویکرد را برای آموزش در پیش گرفت: (الف) یک مدل پیش بینی کننده بردار پشتیبان برای تمام طول‌های مختلف پیشوند دنباله در نظر گرفت و یا (ب) چند مدل پیش بینی کننده بردار پشتیبان در نظرگرفت بطوریکه به ازای هر طول مختلف پیشوند دنباله، یک مدل پیش‌بینی‌کننده داشته باشیم.

۳-۳-۶ رویکرد پیش بینی‌کننده تک‌مدله

در آموزش یک پیش‌بینی‌کننده، نیاز به مشخصات نمونه‌فرآیندها و همچنین مشخصات تجمیعی رویدادها از پیشوندهای مختلف هر دنباله وجود دارد. توابع تجمیع اطلاعاتی که بر روی مشخصات

رویدادها اعمال می شود، شامل تابع count (برای صفات کیفی) و همچنین تابع mean (برای صفات کمی) می باشد.

طول بردار ویژگی، برای هر سابقه‌رویداد ثابت می باشد و طول آن بستگی به *تعداد مشخصات پرونده*، *تعداد مشخصات کیفی رویداد و تعداد مقادیر ممکن برای هر مشخصه کیفی* و *تعداد مشخصات کمی و تعداد توابع تجمیع برای آنها* دارد. بردار ویژگی از سه مجموعه داده تشکیل می‌شود:

۱. مجموعه مشخصات پرونده: مقادیر این مجموعه در طول فرآیند، تغییر نمی‌کند و ثابت می باشد. این مقادیر، همانند سن صاحب پرونده، جنسیت صاحب پرونده و غیره در گذر از رویدادهای مختلف تغییر نمی‌کند.

۲. مجموعه مشخصات رویداد: این مقادیر دینامیک و پویا هستند، بطوریکه هر رویداد در یک دنباله، مقادیر مشخصات جداگانه‌ای دارد. از آنجاییکه بردار ویژگی طول ثابتی دارد و از طرف دیگر، پیشوندهای مختلف هر دنباله، در هرکدام از مشخصات یک رویداد تأثیر می گذارند، برای انباشت اثر هر کدام باید از توابع تجمیع استفاده کرد. برای مشخصات کمی، توابع تجمیع همانند میانگین، مینیمم، ماکزیمم، مجموع و انحراف معیار می‌باشد. برای مشخصات کیفی، از تابع شمارش استفاده می‌کنیم، که برای مثال، تعداد دفعات اجرای یک فعالیت خاص و یا تعداد فعالیت‌های اجراشده توسط یک منبع خاص را نشان می‌دهد.

۳. مجموعه هدف: علاوه بر مجموعه ویژگی های فوق، از یک متغیر هدف y استفاده می‌کنیم، که مدل پیش‌بینی‌کننده باید آن را بیاموزد. برای مثال، زمانی که قصد پیش‌بینی مدت زمان انجام یک فعالیت را داریم، اختلاف زمانی برچسب زمانی اتمام آن فعالیت و اتمام فعالیت پیشین را به عنوان متغیر هدف در نظر گرفته می‌شود. اگر یک فعالیت در یک پرونده خاص، اجرا نشده باشد، طول مدت زمان آن فعالیت نامعلوم است. بنابراین، می‌توان چنین پرونده هایی را از مجموعه آموزشی خارج نمود. برعکس، اگر یک فعالیت چندین مرتبه در یک پرونده انجام گرفته باشد، میانگین طول مدت زمان آن را در نظر می‌گیریم.

مبحث ساخت بردار ویژگی را با یک مثال پیش می‌بریم. جدول ۳-۱ نتیجه ساخت بردار ویژگی برای پیش بینی مدت زمان فعالیت F را نشان می دهد. برای راحتی کار و خوانابودن، برچسب های زمانی را به مقدار نسبی تغییر داریم (برحسب ثانیه)، که مبدأ زمانی را شروع پرونده در نظر گرفتیم. علاوه براین، تنها یک تابع تجمیع، که مجموع می باشد را برای مشخصات عددی در نظر گرفتیم.

برای ساخت بردار ویژگی برای احتمالات انشعابات، یک برچسب کلاس به هر انشعاب خروجی انتساب داده می‌شود. برای مثال، اگر قصد پیش بینی احتمال دروازه XOR-32 را داریم، مقدار صفر را به انشعاب منجر به تکرارکاری انتساب می دهیم و مقدار یک را به انشعاب دیگر می دهیم. پرواضح است که احتمال کلاس صفر برابر با احتمال تکرارکاری r می باشد. بنابراین اولین پرونده در جدول سابقه رویداد، XOR-32 برابر با یک می باشد در صورتیکه برای پرونده دوم XOR-32 نامشخص می‌باشد، زیرا پرونده به دروازه نرسیده است. درنتیجه، از پرونده دوم نمی‌توان برای مجموعه آموزشی برای پیش‌بینی XOR-32 استفاده نمود. جدول ۳-۱ نتایج ساخت بردار ویژگی مجموعه‌آموزشی را نشان می دهد.

۳-۳-۷ رویکرد پیش بینی کننده چندمدله

در ساخت بردار ویژگی با رویکرد پیش بینی‌کننده چندمدله، نیاز به یک مدل پیش‌بینی‌کننده به ازای هرکدام از اندازه های مختلف پیشوند دنباله‌های جزئی می‌باشد. بردار ویژگی شامل مشخصات پرونده و همچنین برای هر اندیس درون دنباله، رویداد رخ داده در آن اندیس و مقدار هر مشخصه رویداد در آن اندیس باید باشد. به صورت کلی، برای هر پرونده با M مشخصه پرونده $\{s_1, \dots, s_M\}$ شامل k رویداد $\{e_1, \dots, e_k\}$، که هر رویداد داده های مربوط به خود را به طول R دارد $\{d_1^1, \dots, d_1^R\}, \dots \{d_k^1, \dots, d_k^R\}$ که منجر به تولید بردار ویژگی به صورت زیر می‌شود:

$$\vec{X} = (S_1, \dots, S_M, e_1, d_1^1, \dots, d_1^R, \dots, e_k, d_k^1, \dots, d_k^R)$$

که در اینصورت طبق فرمول M+k.R طول بردار ویژگی با افزایش هر رویداد انجام شده e_k افزایش می‌یابد. جدول ۳-۱ مثالی از بردار ویژگی مجموعه آموزشی برای پیش بینی مدت زمان انجام فعالیت F برای پیشوندهای با طول k=3 را نشان می دهد.

به طور مشابه، جدول ۳-۲ نیز مثالی را برای بردار ویژگی مجموعه آموزشی برای پیش بینی احتمالات انشعاب XOR-32 با پیشوندهای بطول k=3 را نشان می‌دهد.

توجه داشته باشید که می توان دو رویکرد برای از پروندهای با طول کمتر از k برای پیش بینی های به طول k در نظر گرفت:

- این پرونده ها را می توان در نظر نگرفت و در مجموعه آموزشی استفاده نکرد.
- این پرونده ها را با شرط به اندازه k رساندن به وسیله تولید داده با صفر قراردادن و یا استفاده از داده های میانگین پیشین از پرونده های با اندازه k استفاده کرد.

در این رساله، ما سعی می کنیم که از داده‌های ساختگی استفاده نکنیم و از پرونده های با طول کوچکتر از k برای آموزش و استفاده در پیش بینی های به اندازه k استفاده نکنیم.

جدول ۳-۱: بردار ویژگی و مجموعه داده آموزشی برای پیش بینی طول مدت زمان فعالیت F (رویکرد تک پیش بینی کننده)

داده جدید	مجموع هزینه	مجموع برچسب زمانی	منبع Kate	منبع Mary	منبع Mark	منبع John	فعالیت H	فعالیت G	فعالیت F	فعالیت D	فعالیت C	فعالیت B	فعالیت A	سن	کانال
۵	۱۵	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۱	۳۷	ایمیل
۵	۴۰	۸۰	۰	۰	۱	۱	۰	۰	۰	۰	۰	۱	۱	۳۷	ایمیل
۵	۵۰	۱۸۰	۰	۱	۱	۱	۰	۰	۰	۱	۰	۱	۱	۳۷	ایمیل
۵	۶۰	۳۰۰	۰	۱	۲	۱	۰	۰	۰	۱	۱	۱	۱	۳۷	ایمیل
۵	۸۰	۳۰۵	۱	۱	۲	۱	۰	۰	۱	۱	۱	۱	۱	۳۷	ایمیل
۵	۱۰۰	۳۵۰	۱	۱	۲	۲	۰	۱	۱	۱	۱	۱	۱	۳۷	ایمیل
۵	۱۱۵	۳۶۰	۲	۱	۲	۲	۱	۱	۱	۱	۱	۱	۱	۳۷	ایمیل
۵۰	۲۵	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۱	۳۷	ایمیل
۵۰	۵۰	۳۰۰	۰	۱	۰	۱	۰	۰	۰	۱	۰	۰	۱	۳۷	ایمیل
۵۰	۶۰	۵۷۹۰۰	۰	۱	۱	۱	۰	۰	۰	۱	۰	۱	۱	۳۷	ایمیل
۵۰	۷۰	۵۷۹۶۰	۰	۱	۲	۱	۰	۰	۰	۱	۱	۱	۱	۵۲	ایمیل
۵۰	۸۵	۵۸۰۱۰	۱	۱	۲	۱	۰	۰	۱	۱	۱	۱	۱	۵۲	ایمیل

جدول ۳-۲: بردار ویژگی و مجموعه داده آموزشی برای پیش بینی احتمال انشعاب XOR-32 (رویکرد تک پیش بینی کننده)

کانال	سن	فعالیت A	فعالیت B	فعالیت C	فعالیت D	فعالیت F	فعالیت G	فعالیت H	منبع John	منبع Mark	منبع Mary	منبع Kate	مجموع برچسب زمانی	مجموع هزینه	داده هدف
ایمیل	۳۷	۱	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۱۵	۵
ایمیل	۳۷	۱	۱	۰	۰	۰	۰	۰	۱	۱	۰	۰	۸۰	۴۰	۵
ایمیل	۳۷	۱	۱	۰	۱	۰	۰	۰	۱	۱	۱	۰	۱۸۰	۵۰	۵
ایمیل	۳۷	۱	۱	۱	۱	۰	۰	۰	۱	۲	۱	۰	۳۰۰	۶۰	۵
ایمیل	۳۷	۱	۱	۱	۱	۱	۰	۰	۱	۲	۱	۱	۳۰۵	۸۰	۵
ایمیل	۳۷	۱	۱	۱	۱	۱	۱	۰	۲	۲	۱	۱	۳۵۰	۱۰۰	۵
ایمیل	۳۷	۱	۱	۱	۱	۱	۱	۱	۲	۲	۱	۲	۳۶۰	۱۱۵	۵

جدول ۳-۳: بردار ویژگی و مجموعه داده آموزشی برای پیش بینی احتمال انشعاب XOR-32 و K=3 (رویکرد چند پیش‌بینی‌کننده)

سن	۳۷
فعالیت ۱	A
برچسب زمانی۱	۰
منبع ۱	John
هزینه ۱	۱۵
فعالیت ۲	B
برچسب زمانی۲	۸۰
منبع ۲	Mark
هزینه ۲	۲۵
فعالیت ۳	D
برچسب زمانی۳	۱۸۰
منبع ۳	Mary
هزینه ۳	۱۰
داده هدف	۵

جدول ۳-۴: بردار ویژگی و مجموعه داده آموزشی برای پیش بینی طول مدت زمان فعالیت F و K=3 (رویکرد چند پیش‌بینی‌کننده)

کانال	ایمیل	ایمیل
سن	۳۷	۵۲
فعالیت ۱	A	A
برچسب زمانی۱	۰	۰
منبع ۱	John	John
هزینه ۱	۱۵	۲۵
فعالیت ۲	B	D
برچسب زمانی۲	۸۰	۳۰۰
منبع ۲	Mark	Mary
هزینه ۲	۲۵	۲۵
فعالیت ۳	D	B
برچسب زمانی۳	۱۸۰	۵۷۹۰۰
منبع ۳	Mary	Mark
هزینه ۳	۱۰	۱۰
داده هدف	۵	۵۰

۳-۴ رویکرد پیش بینی با توزیع‌های احتمالاتی با لحاظ صف

در این بخش، یک رویکرد دیگر با همان تکنیک تجزیه مسیر، اما با تکیه بر آنالیز احتمالاتی و با احتساب صف برای پیش‌بینی مسیر نمونه فرآیند ارائه می‌شود. این رویکرد مستلزم استخراج اطلاعات زمانی از سابقه‌رویداد و غنی‌سازی مدل توسط اطلاعات زمانی برای ارائه تحلیل‌های احتمالاتی زمانی می‌باشد. اطلاعات زمانی که برای هر وظیفه تعریف می‌شود می‌تواند شامل زمان فعال شدن یک واحد کاری، مدت‌زمان انتظار واحد کاری در صف، زمان آغاز سرویس‌دهی به فعالیت توسط یک منبع خاص، زمان پایان سرویس‌دهی به یک فعالیت توسط یک منبع خاص باشد.

برای انجام هر فعالیت، یک توزیع احتمالاتی در نظر گرفته می‌شود. برای پرهیز از پیچیدگی محاسباتی، از متغیرهای تصادفی با تابع توزیع نرمال استفاده می‌کنیم:

$$f_a(t) = \frac{1}{\sigma\sqrt{2\pi}} \times e^{-\frac{(t-\mu)^2}{2\sigma^2}} \qquad (۸-۳)$$

با تابع توزیع تجمعی $F_a(t) = \int_0^t f_a(t)\,dt$ احتمال اتمام فعالیت a را در زمان t نشان می‌دهد. برای فعالیت a با تابع چگالی احتمال $f_a(t)$ ، تابع توزیع تجمعی به صورت زیر خواهد بود:

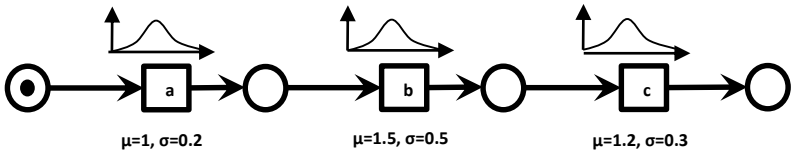
$$F_a(t) = \frac{1}{\sigma\sqrt{2\pi}} \times \int_{-\infty}^{t} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt \qquad (۹-۳)$$

اگر دو فعالیت $a_1, a_2 \in A$ داشته باشیم که توزیع احتمالاتی آنها به ترتیب $\delta(a_1) = \mathcal{N}(\mu_1, \sigma_1^2)$ و $\delta(a_2) = \mathcal{N}(\mu_2, \sigma_2^2)$ باشد، می‌توان ترکیب احتمالاتی اجرای متوالی آن دو فعالیت را به صورت زیر در نظر گرفت:

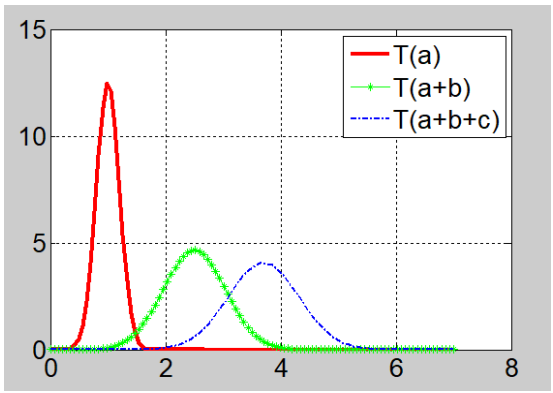
$$\delta(a_1 + a_2) = \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \qquad (۱۰-۳)$$

همانطور که ذکر شد، با فرض وجود تابع چگالی احتمال، تابع توزیع تجمعی را از طریق انتگرال تابع چگالی بدست می‌آید. تابع توزیع تجمعی احتمال P(a)، اتمام فعالیت a را در زمان t نشان می‌دهد. متمم فعالیت a را با $\bar{a}$ نشان می‌دهیم که نشان می دهد که اتمام فعالیت a هنوز اتفاق نیفتاده است و به عبارت دیگر فعالیت هنوز فعال است. $P(\bar{a})$ را با محاسبه $P(\bar{a}) = 1 - P(a)$ بدست می آوریم.

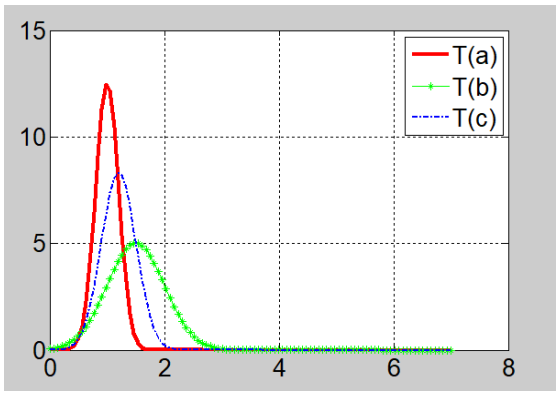
حال فرض کنید، احتمال اجرای نمونه فرآیند در نقطه میان دو فعالیت a و b که a پیش تر از b آمده است، به صورت $P(\bar{b}) - P(\bar{a})$ محاسبه می‌شود.



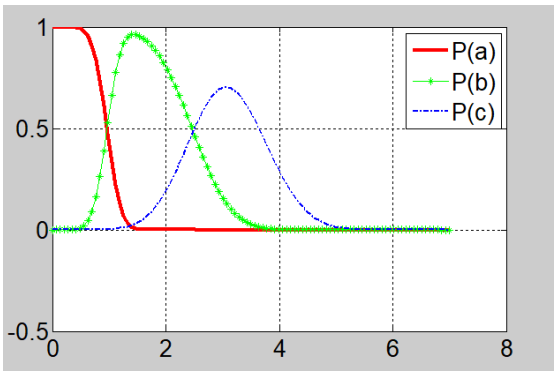
(الف)



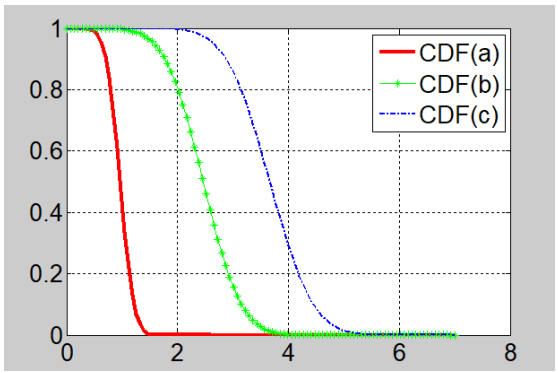
(ج)



(ب)



(ه)

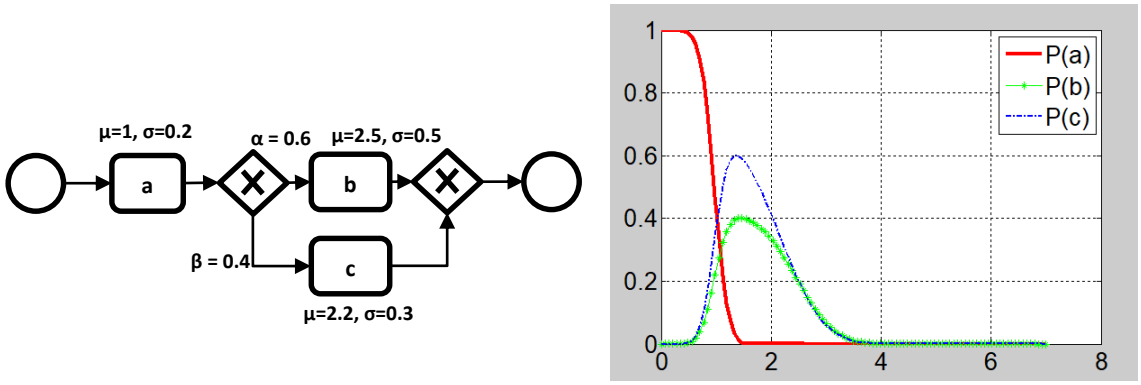


(د)

شکل ۳-۶: توزیع‌های احتمالاتی مربوط به سه وظیفه در فرآیند شکل الف، (ب) به صورت مجزا، (ج) ترکیب احتمالاتی طبق فرآیند، (د) احتمال مستقل "به اتمام نرسیدن" تا لحظه t، (ه) احتمال "به اتمام نرسیدن" تا لحظه t با احتساب وابستگی طبق فرآیند.

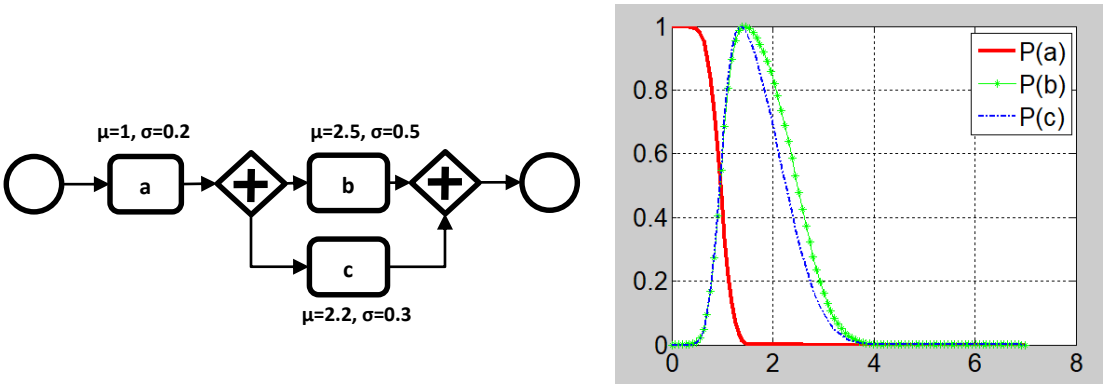
در **شکل ۳-۶** (الف)، مدل فرآیندی، یک فرآیند ترتیبی سه‌وظیفه‌ای را نشان می دهد که با توزیع‌های احتمالاتی نرمال با μ و σ برای هر وظیفه غنی‌سازی شده است. در **شکل ۳-۶** (ب)، توزیع‌های چگالی احتمال هر وظیفه رسم شده است. در **شکل ۳-۶** (ج)، طبق فرمول (۳-۱۱)، ترکیب

احتمالات گوسی وظایف فرآیند را در حالت a+b و a+b+c رسم شده است. در **شکل ۳-۶** (د)، متمم تابع احتمال توزیع تجمعی برای وظایف فرآیند به صورت مستقل ترسیم شده است که نشان دهنده احتمال فعال بودن و به اتمام نرسیدن وظیفه را تا لحظه نشان می دهد. توجه فرمایید که در این نمودار ارتباط ترتیبی وظایف در نظر گرفته نشده است و در نمودار احتمالات مربوط به وظایف فارغ از وابستگی های اجرایی فرآیند رسم شده است. **شکل ۳-۶** (ه)، توزیع های احتمال تجمعی تفریق شده را نشان می دهد که در حقیقت همان احتمالات **شکل ۳-۶** (د) با احتساب وابستگی اجرایی و ترتیب اجرایی فرآیند می باشد. احتمال P(a)، P(b) و P(c) به ترتیب احتمال "فعال بودن" و "به اتمام نرسیدن" فعالیت a، فعالیت b و فعالیت c را در طول زمان t برروی فرآیند را نشان می دهد.



شکل ۳-۷: محاسبه احتمالات در مواجهه با کنترل انتخابی XOR

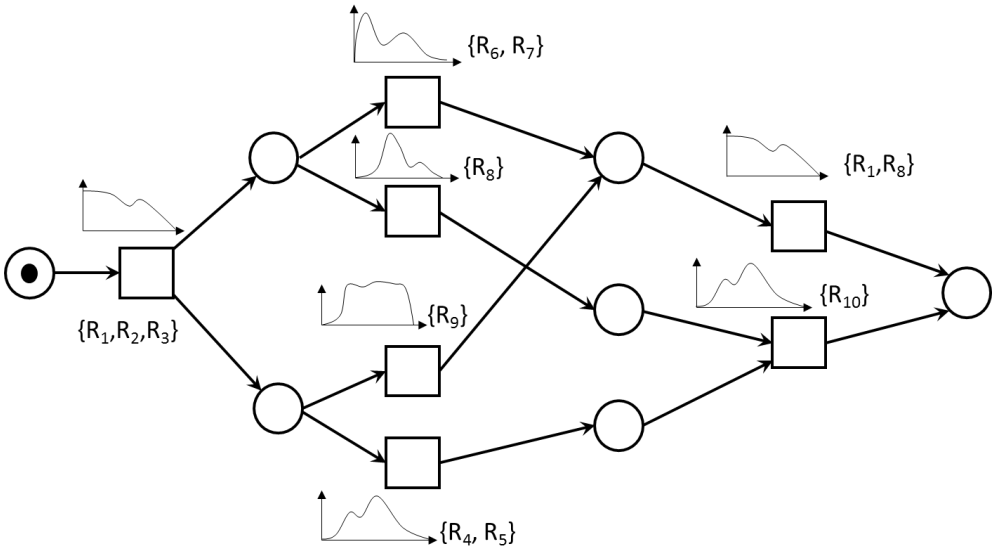
شکل ۳-۷، ترکیب احتمالات وظیفه‌ها را در مواجهه با کنترل جریان انتخابی XOR نشان می دهد، که احتمال انشعاب به وظیفه b و احتمال انشعاب به وظیفه c به ترتیب $\alpha = 0.6$ و $\beta = 0.4$ به صورت ضریبی به احتمال وظایف هر انشعاب اعمال می‌شود. توجه داشته باشید که توزیع احتمالی رسم شده در شکل ۳-۷، توزیع احتمال فعال بودن هر وظیفه در زمان t را طبق ارتباطات درون فرآیند شکل ۳-۷ می باشد. بنابراین طبق فرمول $P(b) = \gamma.(T(a) - T(a+b))$ ضریب احتمال γ همان احتمال انشعاب می باشد.



شکل ۳-۸: محاسبه احتمالات در مواجهه با کنترل AND

شکل ۳-۸، ترکیب احتمالات وظیفه‌ها را در مواجهه با کنترل جریان انتخابی AND نشان می دهد، که هر وظایف هر انشعاب باید پردازش شود. محاسبات احتمالاتی این نوع کنترل جریان AND همانند جریان ترتیبی می باشد. توجه داشته باشید که توزیع احتمالی رسم شده در شکل ۳-۸، توزیع احتمال فعال بودن هر وظیفه در زمان t را طبق ارتباطات درون فرآیند شکل ۳-۸ می باشد.

در شکل ۷، یک نمونه مدل استخراجی، همراه با نام منابع و توزیع‌های احتمالاتی‌های موجود بر روی آن‌ها نشان داده‌شده است. به دلیل کمبود فضا، یک توزیع احتمالی برای هر وظیفه رسم شده است، درحالی‌که هر منبع تعریف‌شده بر روی هر وظیفه، یک توزیع احتمالاتی منحصربه‌فرد دارد.



شکل ۳-۹: فرآیند غنی‌شده با توزیع‌های احتمالاتی

اکثر تکنیک‌های فرآیندکاوی به ساختار مدل فرآیندی و مدت زمان پردازش وظیفه‌ها می‌پردازند، اما

زمان انتظار آیتم‌های کاری برای پردازش در نظر گرفته نمی‌شود. برای استخراج و مانیتورینگ صف از درون سابقه‌رویداد، یک تعریف رسمی برای استخراج صف ارائه می‌شود:

استخراج صف: فرض کنید که یک فرآیند با یک وظیفه وجود دارد، یک مجموعه رویدادهای صف $Q: N^+ \to Q$ به صورت $(t, c, p, a) \in Q$ می‌باشد که در آن:

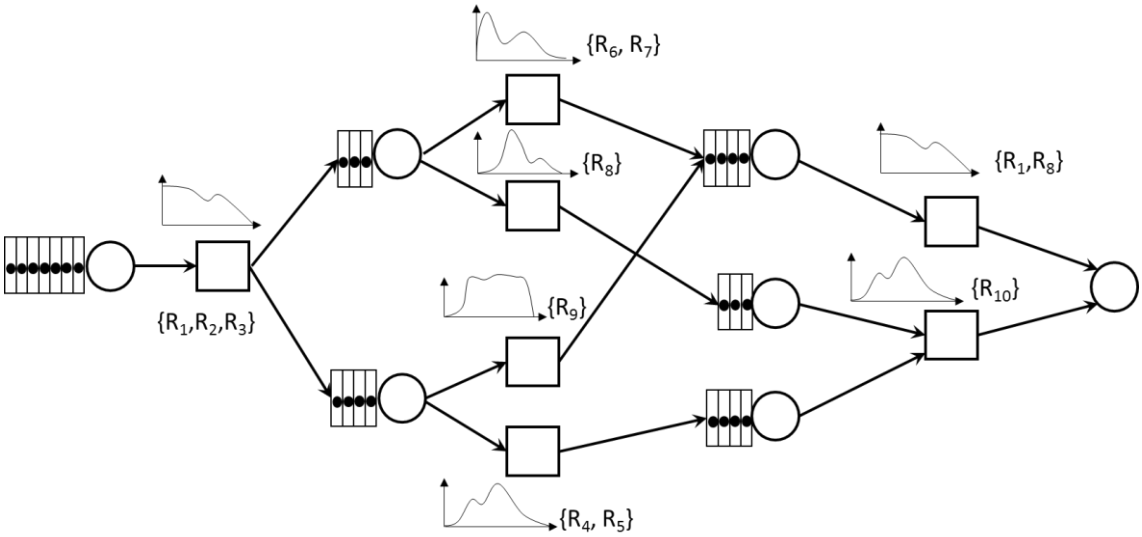
- $t \in N^+$ برچسب زمانی می‌باشد.
- $c \in N^+$ شناسه نمونه‌فرآیند می‌باشد.
- $s \in S = \{qStart, qEnd, sStart, sEnd\}$ نشان دهنده تغییر وضعیت‌ها می‌باشد.
- $a \in A$ نشان دهنده وظیفه مربوطه می‌باشد.

هر رویداد i-ام صف به صورت $Q(i) = (t_i, c_i, s_i, a_i)$ نشان داده می‌شود. رویدادها را به ترتیب برچسب زمانی مرتب می‌کنیم به صورتیکه $t_i < t_j$ و $1 \le i < j \le |Q|$.

در برخی سابقه‌رویدادها، مشخصات زمانی به صورت کامل همانند تعریف فوق الذکر وجود ندارد و همچنین اطلاعات زمانی مربوط به صف وجود ندارد، می‌توان مشخصات $sStart$ و $qStart, qEnd$ را به صورت تقریبی از مشخصه $sEnd$ رویدادهای صف بدست آورد. بدین صورت که اگر بین مشخصه $sEnd$ وظیفه a و وظیفه پیشین a مربوط به دنباله c_i، $sEnd$ مربوط به دنباله‌های دیگر c_k (k متغیر) در همان وظیفه رخ داده باشد، بدین معنی است که دنباله c_i برای پردازش و اجرا در صف قرار داشته است. در ذیل فرمول بندی آن را مشاهده می‌کنید.

$$\begin{cases} Q(i) = (t_i, c_i, \{sEnd\}, prev(a)) \\ Q.new(t_i, c_i, \{qStart\}, a) \end{cases}$$

$$\begin{cases} \begin{cases} QueueList() = \{(k, t_k) \mid t_k < t_M \wedge Q(k) = (t_k, c_k, \{sEnd\}, a) \wedge Q(M) = (t_M, c_i, \{sEnd\}, a)\} \\ k^* = \arg\max_k Queuelist() \end{cases} \\ \begin{cases} Q(j) = (t_j, c_{k^*}, \{sEnd\}, a) \\ Q.new(t_j, c_i, \{qEnd\}, a) \\ Q.new(t_j, c_i, \{sStart\}, a) \end{cases} \end{cases}$$



شکل ۳-۱۰: وضعیت یک فرآیند با احتساب صف های آن در لحظه مشخص t

همانطور که ملاحظه می کنید، در شکل ۳-۱۰، فرآیند شکل ۳-۹ با صف های استخراج شده در لحظه t رسم شده است. جهت کاهش محاسبات در استفاده از توزیع های محاسباتی غیر از توزیع نرمال، یک عدد تصادفی از روی آن توزیع، تولید می‌نماییم و سپس آن عدد ثابت را در محاسبات بکار می‌بریم.

۳-۵ ارزیابی

در این فصل، ما چهار رویکرد پیشنهادی در پیش بینی زمان براساس تجزیه مسیر به همراه یک رویکرد احتمالاتی پیشنهاد نمودیم و حال آنها را با رویکردهای موجود مورد بررسی و ارزیابی قرار می دهیم. از اهداف این ارزیابی، می توان به بررسی کارایی رویکردهای پیشنهادی در پیش بینی زمان باقیمانده نمونه‌فرآیندها نسبت به رویکردهای موجود در ادبیات تحقیق اشاره نمود. همچنین، بررسی کارایی پنج رویکرد پیشنهادی در مقابل یکدیگر می باشد.

۳-۵-۱ پایگاه‌های داده

در این رساله، از ۸ پایگاه داده استفاده نمودیم که ۷ پایگاه داده آن واقعی می‌باشد که می‌توان از طریق

4TU Center for Research Data[1] به آنها دسترسی داشت و یک پایگاه داده نیز شبیه سازی و مصنوعی می‌باشد.

پایگاه داده BPI Challenge 2012. این سابقه رویداد در چالش هوش فرآیند سازمانی که در سال ۲۰۱۲ برگزارشده، مورد آزمایش قرار داده شده است و حاوی داده های مربوط به فرآیند درخواست مالی از یک مؤسسه مالی و اعتباری بزرگ می‌باشد. این فرآیند شامل سه زیرفرآیند می‌باشد: یک زیر فرآیند که وضعیت نرم افزار را پیگیری می کند (BPI Challenge 2012-A)، یک زیرفرآیند که وضعیت پیشنهاد را پیگیری می کند (BPI Challenge 2012-O) و سومی که وضعیت آیتم‌های کاری مربوط به نرم افزار را پیگیری می‌کند (BPI Challenge 2012-W). می توان سه زیرفرآیند را به عنوان پایگاه داده های جداگانه در نظر گرفت. علاوه بر این، زیرفرآیند BPI Challenge 2012-W شامل دنباله هایی از دو یا چند رویداد در یک دنباله از همان فعالیت می باشد؛ به عبارت دیگر، فعالیت ها غالباً چندین بار تکرار می شوند. در بخش معرفی رویکرد پیشنهادی، احتمال تکرار کاری را به صورت ثابت r در نظر گرفتیم. گرچه، در بسیاری فرآیندهای واقعی، مقدار احتمالی برای r پس از هر اجرای حلقه تکرارکاری، کاهش می‌یابد. بنابراین، تعیین مقدار r اهمیت دارد و اگر r به صورت نادقیق پیش‌بینی شود، خطا به صورت سراسری انتشار می‌یابد.

پایگاه داده نیازسنجی اعتبار Credit Requirement. این سابقه رویداد شامل اطلاعاتی در مورد نیازسنجی اعتبار در یک بانک می‌باشد. یک ویژگی این پایگاه داده اینست که از یک مسیر پیروی می کند و بنابراین، مدل فرآیندی ترتیبی می‌باشد و حاوی دروازه های AND وXOR و غیره نمی‌باشد.

پایگاه داده سرویس راهنما Helpdesk. این سابقه رویداد حاوی رویدادهای فرآیند مدیریت تیکت مربوط به یک سرویس راهنما در یک شرکت نرم افزاری ایتالیایی می باشد. هر پرونده این فرآیند، با درج یک تیکت جدید در سیستم مدیریت تیکت، آغاز می شود و با حل شدن مسئله، تیکت بسته

[1] https://researchdata.4tu.nl/en/

می‌شود.

پایگاه داده بیمارستانی. این سابقه رویداد از ماژول های مالی سیستم ERP یک بیمارستان منطقه‌ای بدست آمده است. سابقه رویداد حاوی رویدادهای مربوط به صدور صورتحساب برای خدمات پزشکی ارائه شده توسط بیمارستان می‌باشد. هر دنباله سابقه‌رویداد فعالیت‌های اجراشده برای پکیج خدمات پزشکی را ثبت می‌کند. سابقه رویداد شامل نمونه‌های تصادفی از نمونه‌فرآیندهای ثبت شده در طول سه سال می‌باشد.

پایگاه داده فاکتور. این سابقه رویداد مربوط به فرآیند تأیید فاکتور می باشد که داده های آن مصنوعی و ساخته شده با نرم افزار Minit Process Intelligence می باشد.

پایگاه داده جرایم جاده ای. این سابقه رویداد مربوط به فرآیند مدیریت جرایم جاده ای در یک واحد پلیس در ایتالیا می‌باشد. این سابقه رویداد حاوی رویدادهای مربوط به اعلان های جریمه به همراه بازپرداخت ها می باشد.

تمام این سابقه رویدادها باید پیش پردازش شوند بطوریکه پرونده های به اتمام نرسیده و یا پرونده های از ابتدا ثبت نشده، حذف شوند. علاوه براین، همانطور که اشاره شد تکنیک تجزیه مسیر، نمی‌تواند به طور مستقیم با سابقه‌رویداد غیرساختارمند سروکار داشته باشد. بنابراین، باید مدل فرآیند ساختارمندی را از درون سابقه رویداد استخراج نمود تا تکنیک های تجزیه مسیر را بتوان اعمال نمود. توجه داشته باشید که اجرای همزمان دو فعالیت توسط یک منبع در سابقه‌رویداد توسط این رویکرد کشف فرآیند ارائه شده قابل تشخیص نیست و ممکن است که خطای بکارگیری سابقه رویداد حاوی وظایف همزمان به بخش های پسین در رویکرد پیشنهادی در پیش بینی زمان باقیمانده نمونه فرآیند انتشار یابد و نتایج نادرستی حاصل شود.

۳-۵-۲ ساخت مجموعه آموزشی و مجموعه آزمایشی

در آزمایش‌ها، ابتدا نمونه‌فرآیندها را برحسب زمان ورود نمونه‌فرآیندها به اولین رویداد مرتب

می‌سازیم. سپس، سابقه رویداد را به دو بخش آموزشی و آزمایشی تقسیم می کنیم. ۸۰ درصد اول نمونه فرآیندها که پیش تر از الباقی نمونه فرآیندها پردازش شده را به مجموعه آموزشی اختصاص می دهیم و ۲۰ درصد باقیمانده را جهت ارزیابی پیش‌بینی‌ها به مجموعه آزمایشی اختصاص می دهیم. برای آنکه آموزش ما تنها به سمت مجموعه داده مورد نظر bias نشود، از cross-validation استفاده می‌کنیم.

۳-۵-۳ معیارهای ارزیابی

دو معیار مرسوم در ارزیابی عملیات پیش بینی زمانی، سنجش دقت و تعجیل می‌باشد. در واقع، یک عملیات پیش بینی مناسب، هم دقیق می‌باشد و هم باید زمان مناسب برای واکنش را در اختیار سیستم مدیریت فرآیند سازمانی قرار دهد.

۳-۵-۳-۱ دقت

مشهورترین معیارهای سنجش میزان خطا برای ارزیابی دقت پیش‌بینی متغیرهای پیوسته، میانگین مطلق خطا (MAE)، جذر میانگین مربعات خطا (RMSE) و میانگین درصد خطای مطلق (MAPE) می‌باشد. معیار MAE خطا را به عنوان میانگین محاسباتی خطاهای پیش‌بینی، معیار RMSE خطا را در قالب ریشه میانگین مربعات خطاهای پیش بینی و معیار MAPE خطا را به صورت میانگین درصد خطای بدون علامت اندازه گیری می‌کند. نکته حائز اهمیت در پیش‌بینی زمانی نمونه‌فرآیندها، آنست که مقادیر زمان باقیمانده برای نمونه فرآیندهای مختلف از یک فرآیند، بسیار متغیر است. RMSE بسیار به این نوسانات حساس می باشد. علاوه بر این، زمان باقیمانده ممکن است بسیار به صفر نزدیک باشد، که در این شرایط معیار MAPE نیز به سمت مقادیر بسیار کم گرایش پیدا می کند. بنابراین، از معیار MAE برای اندازه‌گیری خطای پیش‌بینی زمان باقیمانده نمونه فرآیند استفاده نمودیم که به صورت رسمی زیر بیان می‌شود:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y_i}|,$$

که در آن $y_i \in y$ مقدار حقیقی عضو R از تابع برای یک نقطه مفروض می باشد و $\widehat{y_i}$ مقدار

پیش‌بینی‌شده برای آن می‌باشد.

۳-۵-۳-۲ شرط تعجیل در پیش بینی

در بحث عملیات پیش‌بینی زمانی، تعجیل در پیش بینی اهمیت دارد بدین صورت که اگر یک وظیفه مانده به انتهای فرآیند، عملیات پیش بینی را انجام دهیم، زمان کمی برای عملیات های اقداماتی پس از پیش‌بینی وجود دارد و لزوماً بر پیش بینی نمی‌توان به عنوان ابراز پایه برای برنامه ریزی حساب کرد؛ اما اگر پیش بینی را زودتر انجام دهیم که این تعجیل به معنای انتخاب وظیفه های ابتدایی فرآیند به عنوان نقاط مبدأ پیش‌بینی زمانی می‌باشد، می توان زودتر برای اقدامات دیگر همچون همزمان سازی واکنش نشان داد.

با در نظر گرفتن این اهداف، پیش بینی را برای پیشوند دنباله های جزئی $hd^k(\sigma)$ از دنباله های σ در مجموعه آزمایشی با $k=2$ آغاز می‌نماییم. گرچه، استفاده از تمام مقادیر ممکن k چندین مسئله پیش رو دارد. مسئله اول، آنست که تعداد پیشوندهای ممکن در تعداد دنباله های موجود بسیار زیاد می شود که به این ترتیب زمان مرحله آموزش را افزایش می‌دهد. مسئله دوم، در رویکرد پیش‌بینی‌کننده تک مدله، دنباله های بزرگ، پیشوندهای بیشتری نسبت به دنباله های کوچک‌تر تولید می‌کند و بنابراین، مدل پیش‌بینی‌کننده به سمت نمونه فرآیندهای بزرگتر گرایش پیدا می کند [۶۵]. در نهایت، در رویکرد پیش‌بینی‌کننده چندمدله، در پیشوندهای با طول بالا از دنباله‌ها، تعداد پیشوندها بالطبع پایین می‌آید، و در نتیجه مدل به خوبی آموزش نمی‌بیند و نهایتاً مدل پیش‌بینی‌کننده غیرقابل اعتماد می‌شود.

در مجموع، برای کنترل بحث تعجیل، از پیشوندهای حداکثر به طول ۱۰ رویداد را هم در فاز آموزش و هم در فاز آزمایش استفاده می‌کنیم. اگر تعداد رویدادهای یک دنباله کمتر از ۱۰ رویداد باشد، از تمام پیشوندها بجز بزرگ ترین پیشوند استفاده می شود؛ استفاده از بزرگ ترین پیشوند بدلیل اتمام انجام نمونه‌فرآیند، دیگر برای پیش‌بینی بی‌معنا می‌شود. به عبارت دیگر، ورودی آزمایش‌ها از فیلتر $L^* = \{hd^k(\sigma): \sigma \in L, 1 \leq k \leq \min(|\sigma| - 1, 10)\}$ عبور می‌کند.

۳-۵-۴ مراجع ارزیابی

در این ارزیابی، رویکردهای پیشنهادی را با چند رویکرد مرجع مقایسه می‌کنیم. در بحث مقایسه از رویکردهایی که تخمین پیش‌بینی را بدون تجزیه مسیر و با رگرسیون سراسری انجام می دهد و یک مقدار را به عنوان خروجی پیش بینی باز می‌گرداند، استفاده می‌کنیم. رویکرد ارائه شده توسط Leoni و همکاران [۶۶] از نوع رگرسیون سراسری با یک مدل پیش‌بینی‌کننده می‌باشد و می توان آن را به عنوان مرجع مقایسه قرار داد. رویکرد ارائه شده توسط Leontjeva و همکاران [۶۷] از نوع رگرسیون سراسری با چند مدل پیش بینی کننده می باشد و از آن نیز برای مقایسه استفاده خواهیم کرد.

رویکرد ارائه شده توسط Aalst و همکاران [۱۳]، از سیستم گذار برای پیش بینی استفاده کردند که با سه نوع انتزاع شامل مجموعه، کیف و توالی را در آن اعمال کردند. این رویکرد نیز به دلیل پایه ای بودن و مرجع بودن در ارزیابی ها استفاده شده است. رویکرد نهایی در مراجع مقایسه، رویکرد Rogge-solti و همکاران [۳, ۴۳] می‌باشد که از رویکرد تصادفی برپایه شبکه پتری برای پیش بینی زمانی استفاده کردند و قطعاً مقایسه با این رویکرد نیز لازم الاجراست.

۳-۶ ارزیابی نتایج

در جدول ۳-۵، میانگین دقت پیش‌بینی را برای تمام اندازه پیشوندها، که با نسبت تعداد پیشوندها برای اندازه های مختلف وزن دار شده، مشاهده می نمایید. انتساب وزن در محاسبه دقت پیش‌بینی به نسبت تعداد پیشوندها، به این صورت می باشد که به پیشوندهای طولانی‌تر، وزن کمتری انتساب داده می باشد، زیرا تعداد آنها کمتر است و تمام دنباله ها به آن طول نمی‌رسند. همانطور که مشاهده می‌فرمایید، در ۵ پایگاه داده از ۸ پایگاه داده، نتیجه عملکرد رویکردهای پیشنهادی برپایه تجزیه مسیر بهترین نتایج را ارائه کردند. پس از رویکردهای برپایه تجزیه مسیر، رویکردهای برپایه سیستم گذار و همچنین رویکرد با چند رگرسیون سراسری نیز هرکدام در یک پایگاه داده بهترین نتایج را ارائه کردند. شکل ۳-۱۱ تا شکل ۳-۱۳دقت پیش بینی را برای اندازه‌های مختلف پیشوند در قالب MAE ارائه

می‌کند. به جهت خوانایی بهتر نمودارها، از دو رویکرد برپایه سیستم گذار با دو انتزاع، تنها بهترین را با توجه به نتایج جدول ۳-۵ برای نمودارهای شکل ۳-۱۱ تا شکل ۳-۱۳ انتخاب نمودیم. هر نقطه در نمودار که به اندازه خاصی برای پیشوند اشاره می کند، پیشوندهای کوچکتر از آن را دیگر در نظر نمی‌گیرد و تنها پیشوندهای به همان اندازه را در نظر می گیرد. از آنجاییکه تمام نمونه فرآیندها به اتمام نمی‌رسند، با افزایش طول پیشوند، تعداد نمونه فرآیندهای مورد ارزیابی کاهش می‌یابد.

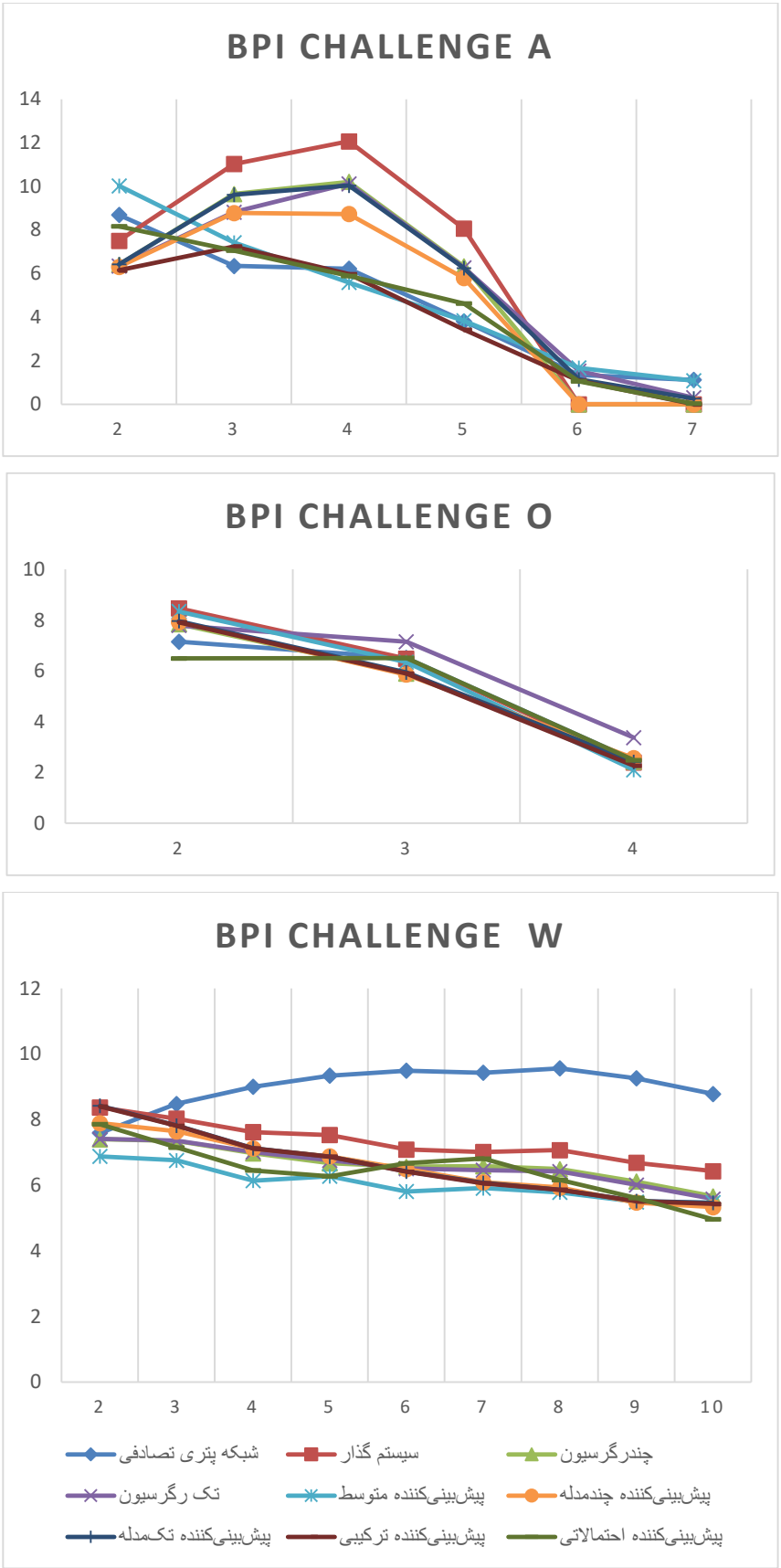
در اکثر پایگاه داده‌ها، مقدار خطای MAE با پیشرفت نمونه‌فرآیندها و افزایش طول دنباله جزئی، کاهش می یابد. طبیعی است هرچقدر به انتهای فرآیند نزدیک‌تر شویم، عملیات پیش‌بینی آسان‌تر می‌شود. توجه داشته باشید، حالتی هم در پایگاه داده BPI Challenge 2012-A وجود دارد که دقت پیش‌بینی‌ها با افزایش طول دنباله جزئی از ۲ به ۴، کاهش می یابد. این پدیده از آنجا نشأت می‌گیرد که دنباله‌های جزئی کوچک، فعالیت هایی پیش رو دارند که پیش‌بینی مدت زمان آنها دشوار می باشد. این دنباله‌های جزئی در ارزیابی دنباله‌های جزئی با طول بزرگتر از ۴ نیامده‌اند، زیرا فعالیت‌هایی از آنها که پیش‌بینی دشواری داشتند، گذرانیده شده‌اند.

نکته قابل ذکر دیگر آنست که در پایگاه داده‌های موجود، اطلاعات زمانی برای صف انتظار وظیفه ها به صورت روشن در پایگاه داده ها ثبت نشده بود و ناچار به استخراج حدودی آن از زمان اتمام فعالیت ها می‌باشیم. اگر پایگاه داده با اطلاعات دقیق زمانی برای صف انتظار وظیفه ها موجود باشد، راندمان و کارایی روش توزیع احتمالاتی با احتساب صف وظیفه ها، قطعاً بیشتر خواهد بود.

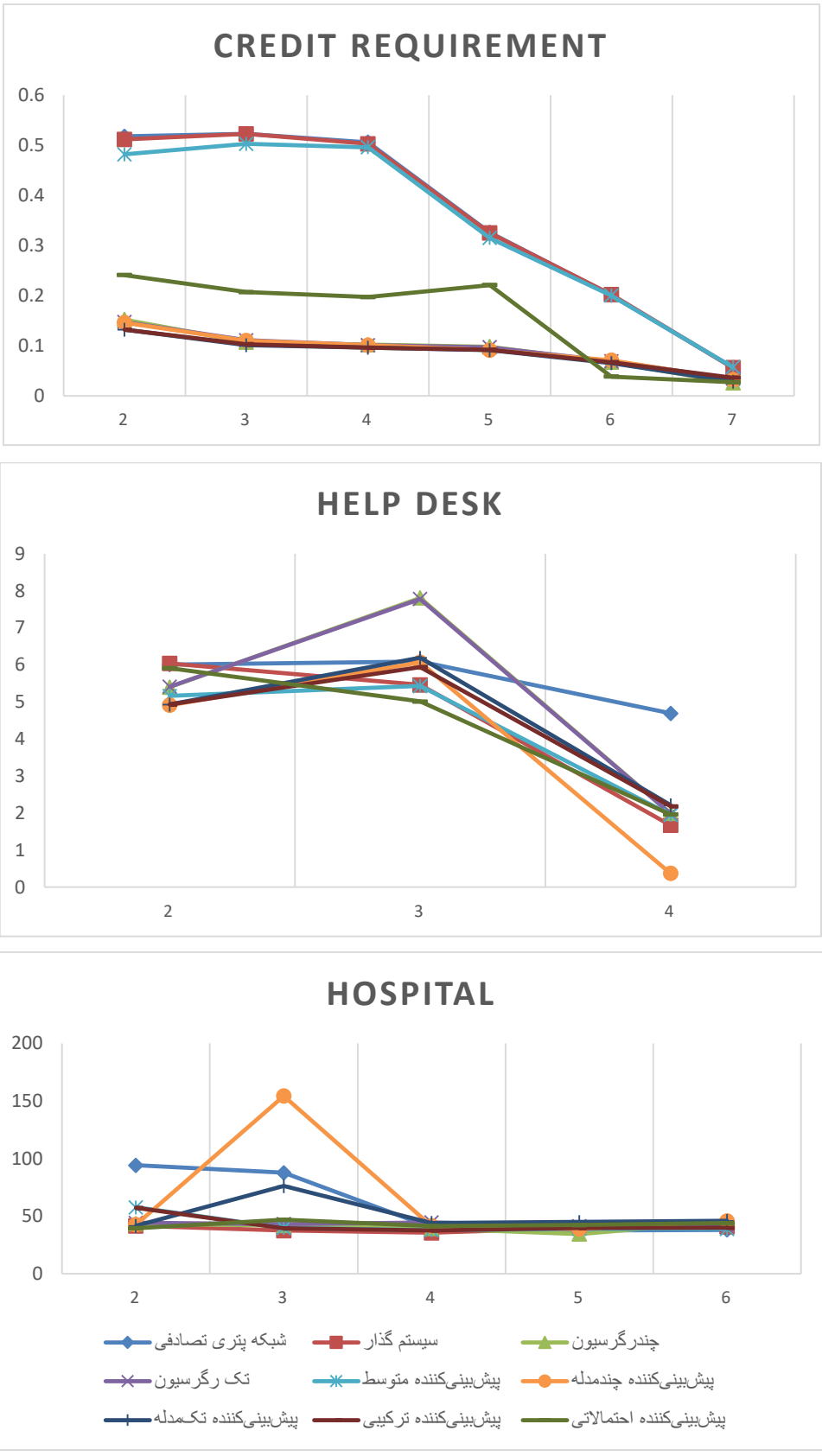
نکته قابل توجه دیگر از نمودارهای شکل ۳-۱۱ تا شکل ۳-۱۳، آنست که آزمایش‌ها و عملکرد رویکرد پیشنهادی بسیار وابسته به مدل فرآیندی درون سابقه رویداد می‌باشد و برخی رویکردها در برخی پایگاه داده ها، عملکرد بهتری دارند. حتی در برخی اندازه پیشوندهای مختلف، دقت پیش بینی برای تمام رویکردها نسبت به اندازه پیشوند کوچکتر، به طور همزمان کاهش می‌یابد و یا به طور همزمان افزایش می‌یابد.

جدول ۳-۵: میانگین خطای وزن‌دار MAE برروی اندازه‌های مختلف دنباله‌های جزئی

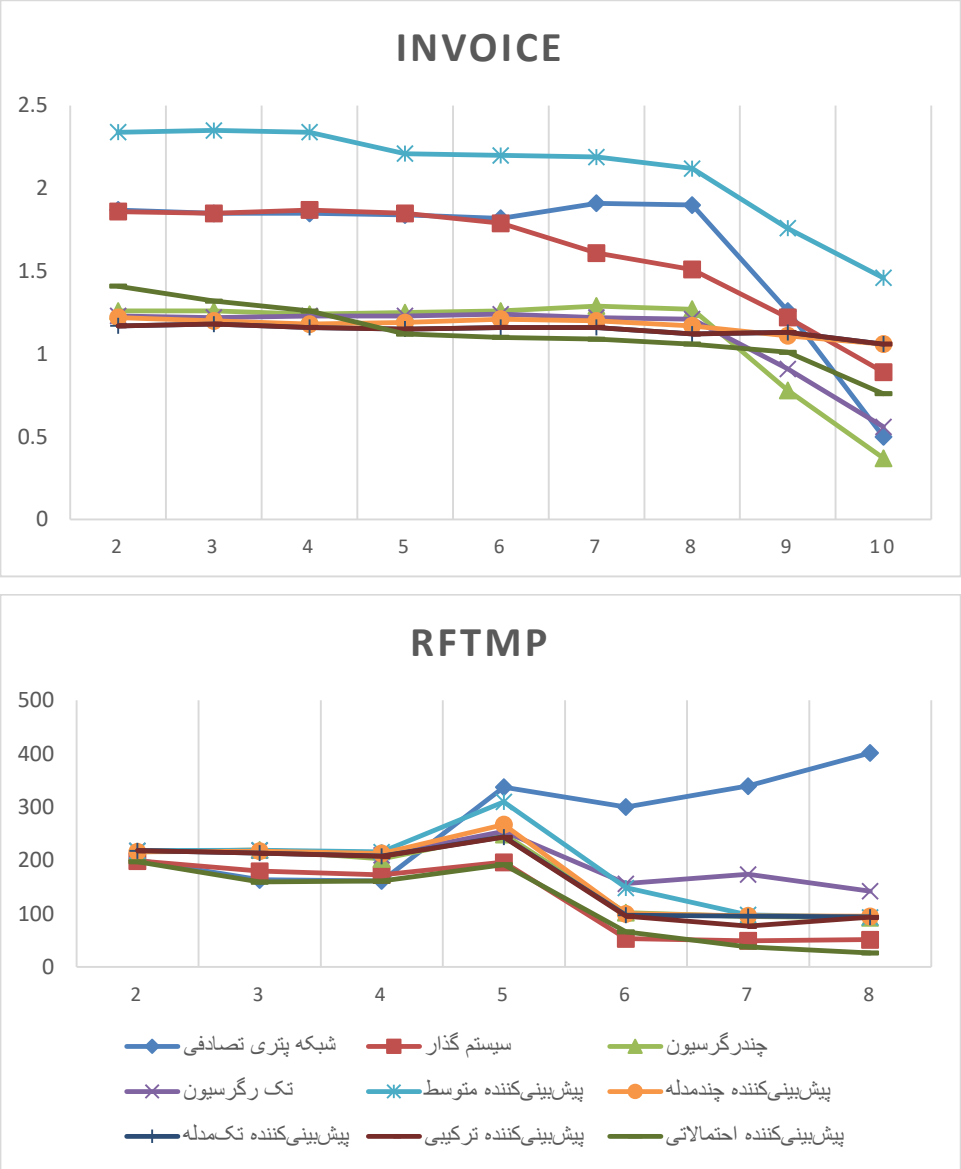
رویکردها	BPI Challenge 2012-A	BPI Challenge 2012-O	BPI Challenge 2012-W	پایگاه داده نیازسنجی اعتبار	پایگاه داده بیمارستانی	پایگاه داده فاکتور	پایگاه داده جرایم جاده ای	پایگاه داده سرویس راهنما
پیش‌بینی کننده تک‌مدله	6.67±3.72	5.95±2.83	6.94±1.05	**0.075±0.039**	51.68±14.94	1.169±0.06	223.50±74.58	5.13±2.09
پیش‌بینی کننده چندمدله	6.83±4.15	6.00±2.64	6.82±0.95	0.087±0.043	72.10±50.77	1.228±0.09	225.46±80.69	**5.04±2.99**
پیش‌بینی کننده متوسط	7.62±3.52	6.24±3.23	**6.19±0.54**	0.339±0.187	42.58±8.62	2.012±0.249	229.33±82.62	5.23±2.02
پیش‌بینی کننده ترکیبی	**6.67±3.72**	5.95±2.83	6.92±1.05	0.078±0.035	42.58±8.62	1.171±0.061	223.51±78.87	5.23±2.02
پیش‌بینی کننده احتمالاتی	7.15±2.32	**5.50±2.34**	7.12±0.37	0.196±0.056	46.74±3.22	1.581±0.135	**172.46±41.23**	5.51±3.73
تک رگرسیون	7.32±4.20	5.75±2.72	6.76±0.6	0.087±0.043	43.23±1.11	1.128±0.242	218.01±38.62	5.80±2.84
چندرگرسیون	7.32±4.56	5.80±2.71	6.80±0.54	0.088±0.847	39.94±4.31	**1.118±0.342**	215.75±70.98	5.74±2.72
سیستم گذار با انتزاع دنباله	8.26±5.23	6.51±3.13	7.61±0.58	0.358±0.201	**36.73±2.41**	1.612±0.375	180.54±71.27	5.98±2.51
سیستم گذار با انتزاع مجموعه	8.26±5.23	6.51±3.13	7.51±0.62	0.358±0.201	**36.74±2.34**	1.612±0.375	180.71±71.89	6.03±0.53
شبکه پتری تصادفی	7.31±1.90	5.94±3.14	8.83±0.64	0.359±0.202	66.37±29.22	1.664±0.48	190.05±94.54	5.96±0.90



شکل ۳-۱۱: دقت پیش‌بینی رویکردها در سه پایگاه داده BPI Challenge 2012 A, B, C با معیار MAE برای اندازه‌های مختلف دنباله‌های جزئی



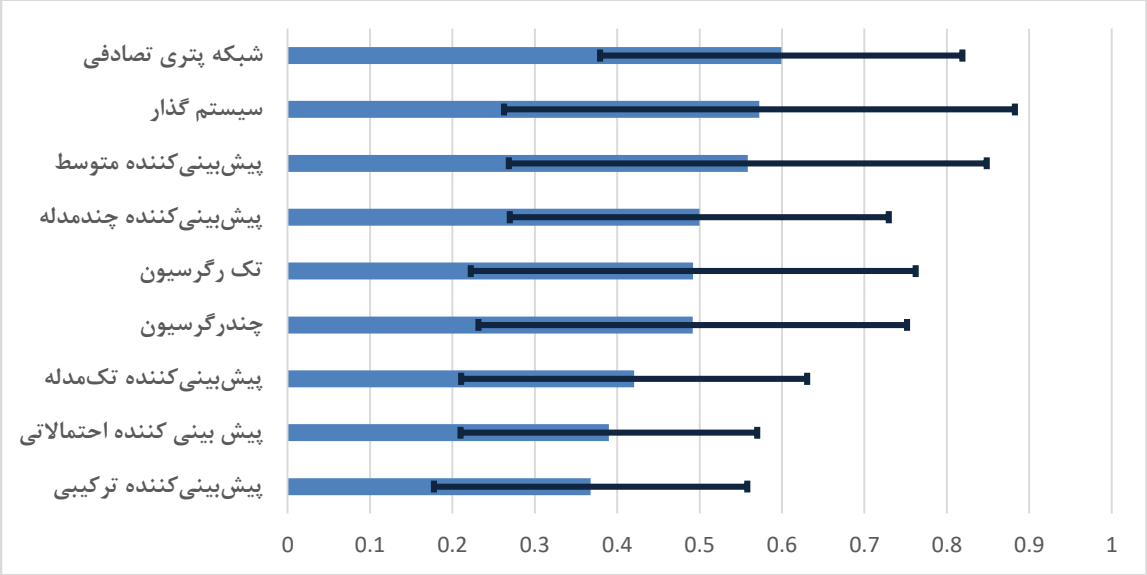
شکل ۳-۱۲: دقت پیش‌بینی رویکردها در سه پایگاه داده CR، HelpDesk و Hospital با معیار MAE برای اندازه‌های مختلف دنباله‌های جزئی



شکل ۳-۱۳: دقت پیش‌بینی رویکردها در سه پایگاه داده RFTMP وINVOICE با معیار MAE برای اندازه‌های مختلف دنباله‌های جزئی

برای جمع‌بندی نتایج بر روی پایگاه‌داده‌های مختلف، مقادیر را با تقسیم بر متوسط چرخه زمانی نمونه فرآیندها، نرمال‌سازی می‌کنیم. برای این کار، میانگین و انحراف معیار گزارش شده برای تمام اندازه‌های مختلف دنباله‌های‌جزئی در جدول ۳-۵ را بر متوسط چرخه زمانی هر فرآیند تقسیم می‌کنیم. مشاهده می‌نمایید که رویکرد پیش بینی ترکیبی، خطای متوسط ۳۶,۷۷ درصد را نسبت به میانگین چرخه زمانی در تمام پایگاه داده‌ها دارد. رویکرد پیش بینی احتمالاتی نیز پس از آن، با خطای **MAE**

متوسط برابر با ۳۸,۹۸ درصد نسبت به چرخه زمانی متوسط در تمام پایگاه داده ها رتبه دوم را کسب کرده است.



شکل ۳-۱۴: خطای میانگین MAE نرمال سازی شده

۳-۷ جمع‌بندی

در این فصل، چهار رویکرد برای پیش‌بینی زمان باقیمانده نمونه فرآیندها با تکنیک تجزیه مسیر و یک رویکرد احتمالاتی با احتساب صف وظیفه ها ارائه گردید که پیش بینی را در سطح فعالیت ها انجام می دهند و لذا توجیه دقیق تری برای تحلیل های فرآیندی دیگر، همانند تشخیص گلوگاهها، انطباق با رانش مفهوم و غیره دارد.

در این فصل، پنج رویکرد پیشنهادی را مورد ارزیابی قرار دادیم. در رویکرد اول، برپایه آنالیز تحلیلی مسیر، برای پیش‌بینی در محل انشعابات و همچنین طول مدت زمان فعالیت‌ها، برای تمام پیشوندهای دنباله‌های جزئی، یک مدل یادگیری ماشین را در نظر گرفته می‌شود. در رویکرد دوم برپایه آنالیز تحلیلی مسیر، برای پیش‌بینی در محل انشعابات و همچنین طول مدت زمان فعالیت‌ها، برای هر پیشوند دنباله‌های جزئی، یک مدل یادگیری ماشین مجزا در نظر گرفته می‌شود. در رویکرد سوم برپایه آنالیز

تحلیلی مسیر، برای پیش‌بینی در محل انشعابات و همچنین طول مدت زمان فعالیت‌ها، از اعداد ثابت میانگین مقادیر سابقه‌رویداد استفاده می‌کند. در رویکرد چهارم برپایه آنالیز تحلیلی مسیر، از ترکیب مدل‌های یادگیری ماشین و مقادیر میانگین به صورت ترکیبی بکار گرفته می‌شود. در رویکرد پنجم پیشنهادی از توزیع های احتمالاتی و احتساب صف انتظار وظیفه‌ها، برای پیش‌بینی زمان باقیمانده نمونه فرآیند استفاده می شود. نتایج تجربی نشان می‌دهد که رویکردهای پیشنهادی، به طور میانگین، دقت پیش‌بینی بالاتری در مراحل مختلف پیشرفت نمونه فرآیندها را ارائه می‌کنند.

فصل ۴ : پیش‌بینی زمانی نمونه فرآیندها با لحاظ رانش مفهوم

۴-۱ مقدمه

یکی از اصلی ترین چالش‌های حوزه فرآیندهای سازمانی، پیش‌بینی زمان باقیمانده نمونه فرآیند می‌باشد، که توجه بسیاری از محققان را نیز به خود جلب کرده است، اما رویکردهای پیش‌بینی در مقابل رانش مفهوم مقاوم نیستند و دچار تزلزل می‌شوند. منظور از رانش مفهوم در فرآیندهای سازمانی، تغییر در رفتار فرآیند به علت تغییر در مدل فرآیندی درون سابقه رویداد می‌باشد. این تغییرات ممکن است سه جنبه کنترلی، داده ای و منابع را تحت تأثیر قرار دهد.

در این فصل، یک رویکرد پیشنهادی، جهت پیش‌بینی زمان باقیمانده نمونه فرآیند به همراه یک روش تطبیق با رانش مفهوم، ارائه می‌شود. در الگوریتم پیش‌بینی زمانی پیشنهادی، یک سیستم گذار ساخته می شود و با احتمالات بدست آمده از ماشین‌های بردار پشتیبان فازی براساس داده نمونه‌فرآیند حاشیه‌نویسی می‌شود. توالی فعالیت‌های آینده نمونه فرآیند با استفاده از الگوریتم کوتاهترین مسیر برروی سیستم گذار حاشیه نویسی شده بدست می آید. پس از آن، مقدار پیش‌بینی زمان باقیمانده نمونه‌فرآیند را با محاسبه مجموع طول مدت فعالیت‌های آینده که با رگرسورهای بردار پشتیبان تخمین زده شده، بدست می‌آید.

به جهت مقاوم سازی الگوریتم پیش بینی زمانی در مقابل پویایی و تغییر مدل فرآیندی، که با نام رانش مفهوم نیز شناخته می‌شود، راهکار پیشنهادی انطباق با رانش مفهوم پیشنهاد ارائه می شود که از مدل‌های یادگیری ماشین چندگانه بهره می‌برد که هرمدل، ازروی بازه‌های زمانی مختلف از سابقه‌رویداد استخراج می شود و به هرکدام وزنی اطلاق می‌شود که ترکیبی از "ضریب زوال" و همچنین "ضریب دوره ای" می‌باشد. ضریب زوال، ضریبی است که با افزایش فاصله زمانی مدل اکتشافی از نمونه فرآیند تحت آزمایش، کاهش می‌یابد. ضریب دوره ای هم، اهمیت داده‌های فصلی دوره‌ای را یکسان سازی می‌کند.

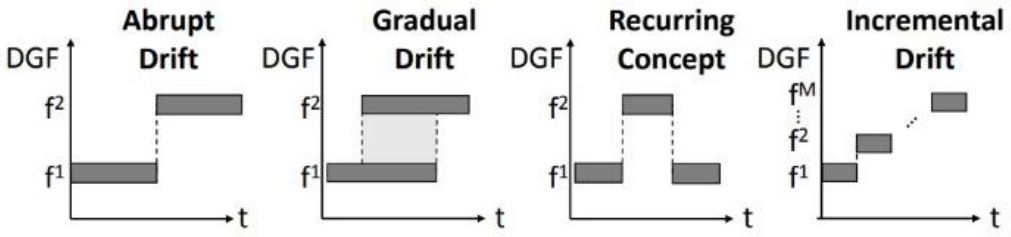
در ادامه فصل، در بخش ۴-۲، رانش مفهوم شرح داده می شود؛ برای درک بهتر رویکرد حل مسئله،

مثالی از یک نمونه سابقه رویداد به همراه سیستم گذار در بخش ۴-۳ ارائه می‌شود؛ در بخش ۴-۴، تکنیک های ساخت سیستم گذار و همچنین ماشین‌های بردار پشتیبان فازی شرح داده خواهد شد. این تکنیک ها در روش پیشنهادی ارائه شده در بخش ۴-۵ استفاده می‌گردد. در بخش ۴-۶، الگوریتم تطبیق رویکرد پیش بینی زمانی با رانش مفهوم توضیح داده می‌شود. در بخش **Error! Reference source not found.**، نتایج آزمایش‌ها رویکرد پیشنهادی را گزارش می شود و در بخش نهایی جمع بندی و نتیجه گیری انجام می‌گیرد.

۴-۲ رانش مفهوم

رانش مفهوم [۶۸] در یادگیری ماشین و داده‌کاوی به شرایطی اطلاق می شود که رابطه میان داده ورودی و متغیر خروجی، که با مدل بازنمایی می شود، در طول زمان تغییر می‌کند. بنابراین، میزان دقت پیش‌بینی‌ها در طول زمان کاهش می یابد. برای جلوگیری از آن، مدل‌های پیش‌بینی‌کننده را باید به صورت برخط با استفاده از داده های جدید تطبیق داد و بروزرسانی نمود. برای مثال، رانش مفهوم در فرآیندهای سازمانی ممکن است به این صورت باشد که در ابتدای سابقه رویداد، دو فعالیت موازی اجرا شوند، در حالیکه در ادامه سابقه رویداد، این فعالیت ها به صورت سری اجرا می‌شوند. فرآیندهای سازمانی ممکن است با توجه به تغییرات فصلی/دوره‌ای تغییر یابد. برای مثال، در ماه اسفند تقاضاها افزایش پیدا می کند یا اینکه در روز چهارشنبه تعداد کارمندها کمتر می‌باشد. در نمونه ای دیگر، زمانی که محیط کاری، رقابتی‌تر می‌شود، تغییرات استراتژیک به فرآیندهای سازمانی اعمال می شود.

رانش مفهوم در فرآیندهای سازمانی انواع مختلفی دارد: رانش ناگهانی، رانش تدریجی، رانش تکرارشونده، رانش افزایشی.



شکل ۴-۱: انواع حالت ها در رانش مفهوم[۶۹]

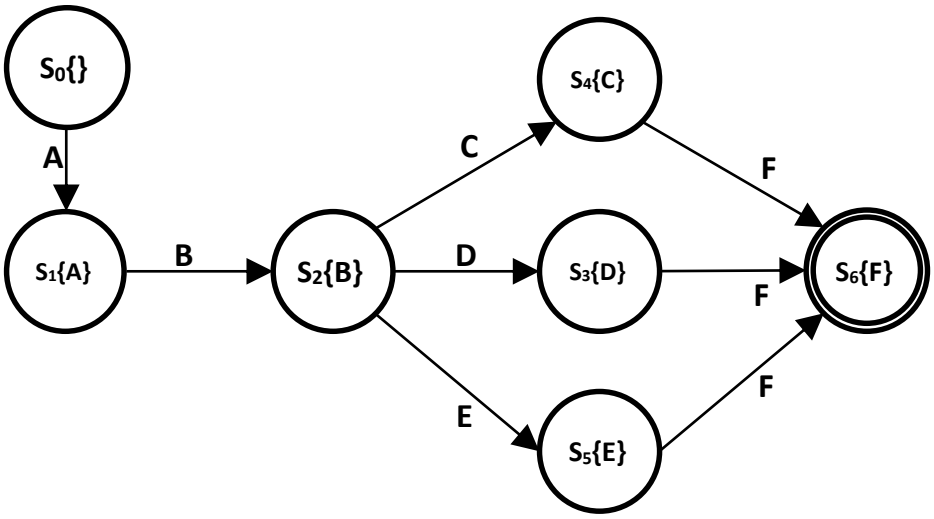
۴-۳ نمونه سابقه رویداد حاوی داده

در جدول ۴-۱، بخشی از سابقه رویداد حاوی داده را مشاهده می نمایید که پس از اجرای A، B و C مشخصات پرونده به "میزان پول=۱۰۰۰ دلار"، "درجه مشتری=طلایی"، "پرداختی موردنیاز=۱۰" تغییرمی یابد. با تمام این اطلاعات (تاریخچه دنباله و مجموعه مشخصات داده) می توان مسیر آینده نمونه فرآیند در حال اجرا را پیش‌بینی نمود.

جدول ۴-۱ : یک مثال از بخشی از سابقه رویداد با رویدادهای مرتب شده توسط برچسب زمانی و گروه بندی شده با شناسه پرونده.

شناسه پرونده	برچسب زمانی	منبع	فعالیت	درجه مشتری	میزان	پرداختی موردنیاز
۶۵۹۲۳	20-02-2002:11:11	Jack	A	--	۱۰۰۰	--
۶۵۹۲۳	20-02-2002:13:31	Jack	B	طلایی	۱۰۰۰	--
۶۵۹۲۳	21-02-2004:08:40	John	C	طلایی	۹۰۰	۱۰
۶۵۹۲۳	22-02-2002:15:51	Joe	F	طلایی	۹۰۰	۱۰
۶۵۹۲۴	19-02-2002:09:10	Jack	A	--	۲۰۰	--
۶۵۹۲۴	19-02-2002:13:22	John	B	استاندارد	۲۰۰	--
۶۵۹۲۴	10-02-2002:17:17	John	D	استاندارد	۲۰۰	۸
۶۵۹۲۴	21-02-2002:10:38	Joe	F	استاندارد	۲۰۰	۸
۶۵۹۲۵	25-02-2002:10:50	Jack	A	--	۸۵۰	--
۶۵۹۲۵	25-02-2002:13:01	John	B	طلایی	۸۵۰	--
۶۵۹۲۵	25-02-2002:16:42	Joe	E	طلایی	۵۰۰	۳
۶۵۹۲۵	26-02-2002:09:30	Joe	F	طلایی	۵۰۰	۳

برای جدول ارائه شده، یک سیستم گذار از سه دنباله <A, B, C, F>، <A, B, D, F> و <A, B, E, F> با تابع بازنمایی رویداد $f^{event}(e) = \pi_a(e)$ و تابع بازنمایی حالت $f^{state}(\sigma) = \{f^{event}(\sigma(|\sigma|))\}$ استخراج شده است. حالت S_0 حالت آغازین و حالت S_6، حالت نهایی پذیرش می باشد. نماد $s_1\{A\}$ بدین معناست که حالت S_1 یک بازنمایی حالت معادل {A} دارد. هر گذار نیز با مقدار بازنمایی رویداد مربوطه برچسب گذاری شده است. شکل ۴-۲، سیستم گذار حاشیه نویسی شده مستخرج از سابقه رویداد جدول ۴-۱ را نشان می‌دهد.



شکل ۴-۲: سیستم گذار حاشیه نویسی شده با توابع بازنمایی حالت و بازنمایی رویداد

۴-۴ تکنیک‌های مورد استفاده در رویکرد پیشنهادی

۴-۴-۱ سیستم گذار

سیستم گذار یک ماشین انتزاع می‌باشد که می تواند در یکی از حالات متناهی تعریف شده در آن قرار داشته باشد. با فعال کردن حالت گذار، ماشین می تواند از یک حالت به حالت دیگر حرکت کند، اما در هر لحظه از زمان، حداکثر می تواند در یک حالت باشد. سیستم گذار را می توان با گراف جهت دار

بازنمایی نمود، که در آن هر گره یک حالت را نشان می دهد و هر یال، گذار میان یک حالت به حالت دیگر را نشان می‌دهد. سیستم گذار را به صورت فرمال در ذیل تعریف می‌کنیم:

تعریف ۱ (سیستم گذار TS). یک سیستم گذار، یک سه تایی TS = (S, A, T) می باشد که در آن S مجموعه ای از حالت‌های ممکن فرآیند می باشد، E مجموعه ای از رویدادها (گذار برچسب گذاری شده) می باشد، و $T \subseteq S \times E \times S$ مجموعه ای از گذار می‌باشد که حرکت سیستم از یک حالت به حالت دیگر را توصیف می‌کند. $S^{start} \subseteq S$ مجموعه‌ای از حالت های آغازین می باشد و $S^{start} \subseteq S$ مجموعه‌ای از حالت‌های پایانی می‌باشد.

سیستم گذار را می توان با استفاده از توابع بازنمایی حالت و بازنمایی رویداد که به انتزاع هم شناخته می شوند، ساخت.

تعریف ۲ (تابع بازنمایی حالت). فرض کنید Rs مجموعه ای از بازنمایی‌های حالت باشد، یک تابع بازنمایی حالت $f^{state} \in \sum \rightarrow R_S$ تابعی است که با فرض وجود دنباله (جزئی) σ یک بازنمایی ازآن ارائه می‌کند (برای مثال دنباله ها، مجموعه ها، مجموعه های چندگانه را برروی برخی مشخصات رویداد تعریف می کنیم) [۱۳].

دراینجا یک دنباله به نمونه فرآیند به اتمام رسیده اشاره می‌کند، درحالیکه یک دنباله جزئی، نمونه فرآیند درحال اجرا را بازنمایی می‌کند. طبق پژوهش [۷۰]، سه انتزاع حالت مرسوم بدین شرح می‌باشد:

- انتزاع دنباله، که تنها ترتیب فعالیت ها در هر حالت را ارائه می‌کند.
- انتزاع مجموعه چندگانه، که شامل تعداد دفعات هر فعالیت را بدون در نظر گرفتن ترتیب آنها را ارائه می کند.
- انتزاع مجموعه، که وجود فعالیت ها را بدون در نظر گرفتن ترتیب فعالیت ها و تعداد دفعات اجرای فعالیت ها ارائه می‌کند.

تعریف 3 (تابع بازنمایی رویداد). فرض کنید R_e مجموعه ای از بازنمایی‌های رویداد باشد، یک تابع بازنمایی رویداد $f^{event} \in E \rightarrow R_e$ تابعی است که با فرض وجود رویداد e یک بازنمایی از آن ارائه می‌کند (توابع تصویر برروی $e \in E$). در رویکرد پیشنهادی، یک سیستم گذار با استفاده از توابع f^{state} و f^{event} تعریف می‌کنیم که در آن حالت های مرتبط با پیشوندهای دنباله‌های سابقه رویداد از طریق f^{event} به بازنمایی رویدادها نگاشت می‌شوند. در شکل ۴-۲، یک سابقه رویداد به همراه یک سیستم گذار مستخرج از آن نشان داده شده است. هر حالت $s_0, \ldots, s_5$ با یک تابع بازنمایی حالت و هر گذار با تابع بازنمایی رویداد برچسب گذاری شده است.

۴-۴-۲ ماشین بردار پشتیبان فازی FSVM

در دسته‌بندی‌کنندهٔ SVM استاندارد، اهمیت میزان خطا (مقدار متغیرهای ϵ_i) به ازای نمونه‌های آموزشی مختلف یکسان است، در حالیکه منطقاً می توان آن را تعمیم داد. با استفاده از منطق فازی، می‌توانیم میزان اهمیت هر نمونه را در فاز آموزش دخالت دهیم. همچنین میتوانیم با استفاده از منطق فازی، در مرحله تصمیم‌گیری به جای یک تصمیم‌گیری خشن (Hard) در SVM استاندارد، یک تصمیم‌گیری نرم را انجام دهیم. در بخش‌های بعدی به شرح چگونگی دو عمل فوق خواهیم پرداخت. در این بخش، برای پرهیز از ذکر موارد پایه و بیسیک، با فرض اینکه خواننده با SVM استاندارد آشنایی دارد، فقط نحوه فرمول بندی FSVM را بیان خواهیم نمود.

۴-۴-۲-۱ دخالت دادن میزان اهمیت نمونه‌ها

SVM استاندارد، نمونه‌های آموزشی را بصورت زوجهای (x_i, y_i) در نظر می گیرد و $y_i \in \{-1,+1\}$. حال برای دخالت دادن میزان اهمیت هر نمونه، نمونه‌ها را بصورت سه‌تایی (x_i, y_i, s_i) در نظر میگیریم s_i .در واقع میزان درجه عضویت نمونه x_i به کلاس خودش است. برای دخالت دادن اهمیت هر نمونه، فرمولاسیون SVM را بصورت زیر تغییر میدهیم [۷۱]:

$$\begin{cases} Minimize_w & \frac{1}{2}\, w.w + C\sum_{i=1}^{N} s_i\varepsilon_i \\ subject\ to & y_i(w.x_i + b) \geq 1 - \ \varepsilon_i \\ & \varepsilon_i \geq 0, for\ i = 1, \dots, N \end{cases} \quad (۱-۴)$$

با این تغییر، مسأله دوگان بهینه‌سازی در SVM بصورت زیر خواهد شد:

$$\begin{cases} Maximize_\alpha & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\, K(x_i, x_j) + \sum_{i=1}^{N}\alpha_i \\ subject\ to & \sum_{i=1}^{N}\alpha_i y_i = 0,\ 0 \ \leq \ \alpha_i \ \leq \ s_i C \\ & , for\ i = 1, \dots, N \end{cases} \quad (۲-۴)$$

در واقع تفاوت SVM فازی با SVM استاندارد در این است حد فوقانی ضریب لاگرانژ α_i برابر s_iC است، درحالیکه در SVM استاندارد این حد برابر C می‌باشد. بردارهای پشتیبان الگوهایی خواهند بود که ضرایب لاگرانژ متناظر آنها در رابطه $0 < \alpha_i \leq s_iC$ صدق کند. تعدادی از بردارهای پشتیبان که ضرایب لاگرانژ متناظر آنها در رابطه $0 < \alpha_i < s_iC$ صدق کند، برای محاسبه b استفاده می‌شود.

حال باید میزان عضویت s_i را به ازای نمونه x_i تعیین کنیم. در مرجع [۷۱] و [۷۲]، s_i براساس نسبت فاصله نمونه از مرکز کلاس به فاصله دورترین نمونه همان کلاس از مرکز کلاس بدست آورده شده است. این روش به داده‌های پرت (Outlier) حساس است و مناسب نمی‌باشد. در اینجا ما s_i را به طریق زیر محاسبه کرده‌ایم:

$$s_i = \exp\left(-\frac{1}{2}(x_i - \mu)\right)^T \sum^{-1} (x_i - \mu)) + \varepsilon \quad (۳-۴)$$

در اینجا ε یک عدد کوچک با مقدار 0.01 قرار داده شده است. μ و Σ به ترتیب بردار میانگین و ماتریس کوواریانس کلاس مربوط به نمونه x_i میباشد. برای سادگی و کاهش محاسبات، می‌توان ماتریس کوواریانس را بصورت قطری فرض کرد. با روش فوق و اعمال s_i در فاز آموزش SVM، در واقع میزان

اهمیت هر نمونه s_i در خطای آن نمونه ε_i ضرب کرده‌ایم و بدین ترتیب اثر داده‌های پرت را کاهش داده‌ایم.

۴-۴-۲-۲ تصمیم‌گیری نرم با استفاده از منطق فازی

در این مقاله، حل مسأله دسته‌بندی در حالت چندکلاسه و بصورت زوج-زوج (Pairwise) مورد نظر است. اگر بردار وزن و مقدار سطح آستانه برای جداسازی داده‌های کلاس i و j پس از آموزش، به ترتیب w_{ij}و b_{ij} باشد، آنگاه تابع تصمیم‌گیری(تابع تمایز) برای جداسازی کلاس i و j چنین خواهد بود:

$$D_{ij}(x) = w_{ij}^T x + b_{ij} \quad , \quad i \neq j \qquad (۴-۴)$$
$$D_{ji}(x) = -D_{ij}(x)$$

در SVM استاندارد، برای جداسازی کلاس i از سایر کلاسها از تابع تمایز زیر استفاده می‌شود:

$$D_i(x) = \sum_{j=1, j\neq i}^{C} sign(D_{ij}(x)) \qquad (۵-۴)$$

C تعداد کلاس هاست. کلاس بردار x چنین تعیین می شود:

$$c(x) = \arg\max_i D_i(x) \qquad (۶-۴)$$

مشکلی که در این توابع تمایز وجود دارد، این است که نواحی غیرقابل دسته بندی (Unclassifiable Regions) وجود دارند.

یک راه برای حل این مشکل این است که تصمیم‌گیری را بصورت نرم و با استفاده از منطق فازی انجام دهیم. برای این کار، هنگام تصمیم‌گیری و تمایز بین دو کلاس i و j ، تابع عضویت $m_{ij}(x)$ را تعریف میکنیم. $m_{ij}(x)$ بدین معنی است که در هنگام دسته‌بندی بردار x ، به چه میزان کلاس i برنده است (و کلاس j بازنده) و میزان تعلق بردار x به کلاس i چقدر است:

$$m_{ij}(x) = \frac{1}{1 + \exp(-D_{ij}(x))} \quad (۷-۴)$$

حال برای آنکه بردار x در تمام زوج های کلاس (i, j) به کلاس i متعلق باشد و کلاس i در تمام این تصمیم گیری های دودویی برنده شود، از AND فازی و اپراتور Min استفاده می کنیم:

$$m_i(x) = \min_{j,\ j \neq i} m_{ij}(x) \quad (۸-۴)$$

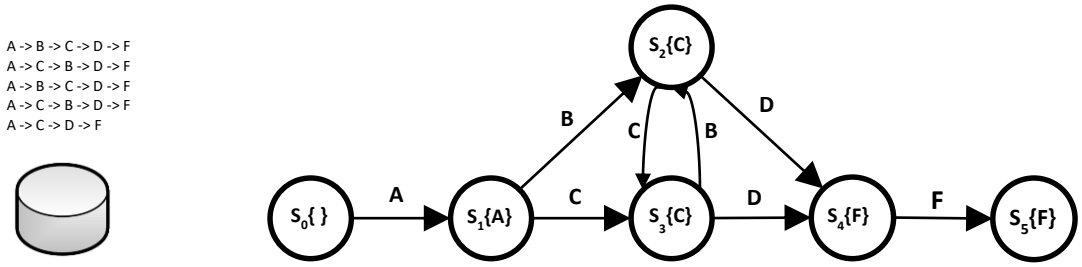
پس از تعیین $m_i(x)$ ها، بردار x را به کلاس $c(x)$ نسبت می دهیم:

$$c(x) = \arg\max_i m_i(x) \quad (۹-۴)$$

با استفاده از منطق فازی و انجام یک تصمیم‌گیری نرم و ملایم، اولاً مشکل نواحی غیرقابل دسته‌بندی برطرف خواهد شد و ثانیاً تصمیم‌گیری و دسته‌بندی به صورت نرم و فازی انجام می‌شود.

۴-۵ پیش بینی مسیر نمونه فرآیند

فرض کنید سابقه‌رویدادی شامل مجموعه ای از دنباله‌های به اتمام رسیده و همچنین یک دنباله جزئی مربوط به نمونه فرآیند در حال پردازش موجود باشد، هدف ما پیش بینی مسیر حرکت نمونه‌فرآیند یا به عبارت دیگر، پیش‌بینی توالی فعالیت‌های آتی دنباله‌جزئی مربوط به نمونه فرآیند در حال پردازش می‌باشد. رویکرد ما، بهبودی برای کار ارائه شده توسط Polato و همکاران [۱۷] محسوب می‌شود که از دسته بند بیز ساده برای تخمین احتمال گذار از حالت جاری به حالت بعدی، استفاده می‌کند. از آنجاییکه در مقاله مذکور، یکی از مفروضات استقلال شرطی پیش‌بینی‌کننده‌ها می باشد که در عمل، همان فرض استقلال مشخصات داده پرونده در فعالیت‌های مختلف می‌باشد. گرچه در فرآیندهای سازمانی واقعی، مشخصات داده معمولاً بهم وابسته‌اند. بنابراین، در این رویکرد، ما از توالی فعالیت ها و داده های هر فعالیت با استفاده از دسته بند SVM فازی برای یافتن مسیر انشعاب استفاده می‌کنیم.

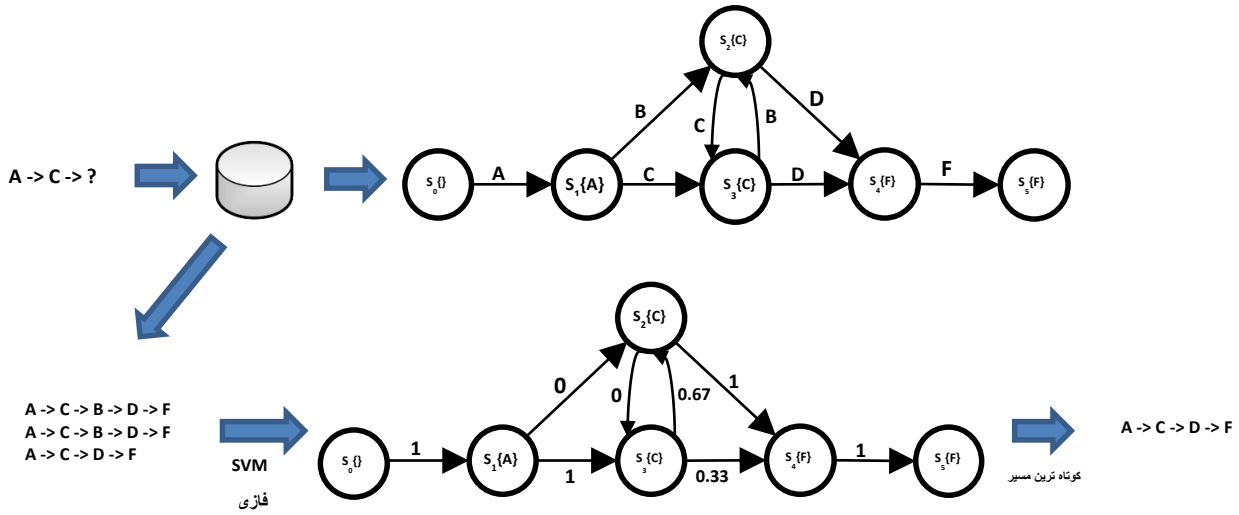


شکل ۴-۳: یک نمونه سیستم گذار مستخرج از سابقه رویداد

ابتدا سیستم گذار را از سابقه‌رویداد طبق الگوریتم ارائه شده توسط Aalst و همکاران [۱۳] استخراج می‌نماییم. برای ساخت سیستم گذار، ما از انتزاع دنباله استفاده می کنیم، که در آن، ترتیب فعالیت‌ها اهمیت دارد. پس از آن، احتمال فعال شدن هر گذار را با دسته بندهای SVM فازی چندکلاسه با استفاده از داده های پیشین موجود که مشابه با نمونه فرآیند جاری مفروض می‌باشد، بدست می آوریم. با در اختیار داشتن احتمالات گذار، محتمل‌ترین توالی از حالت جاری به حالت نهایی فرآیند را بدست می‌آوریم. سپس، دنباله پیش‌بینی‌شده را با دنباله واقعی نمونه‌فرآیند برای ارزیابی رویکرد پیشنهادی استفاده می‌کنیم.

۴-۵-۱ تخمین احتمالات گذار

ایده اصلی رویکرد پیشنهادی، تخمین احتمالات گذار می‌باشد که با دسته‌بندهای SVM چندکلاسه انجام می‌شود و مشخصات نمونه فرآیند جاری را به عنوان ورودی می پذیرد و به جستجوی یک نمونه فرآیند مشابه در مجموعه آموزشی می‌باشد. ما دو بردار ویژگی در عملیات دسته‌بندی استفاده می‌کنیم: بردار ویژگی پیشوند نمونه‌فرآیند و بردار ویژگی مجموعه‌داده‌های نمونه‌فرآیند. بردار ویژگی پیشوند نمونه فرآیند جنبه جریان کنترلی نمونه فرآیند را بازنمایی می‌کند و بردار ویژگی مجموعه‌داده‌های نمونه‌فرآیند جنبه داده‌ای نمونه‌فرآیند را بازنمایی می‌کند.



شکل ۴-۴: رویکرد پیشنهادی برای سیستم گذار حاشیه نویسی شده

۴-۵-۲ پیش بینی محتمل ترین مسیر

حال، با وجود سیستم گذار حاشیه نویسی‌شده با احتمالات گذار، می توان احتمال هر دنباله نمونه فرآیند $S = <s_1, s_2, \dots, s_n>$ را با ضرب احتمالات گذار میان جفت حالات متصل بهم بدست آورد:

$$P_S = \prod_{i=1}^{n-1} p_i \qquad (۱۰-۴)$$

در زمان اجرا، دنباله جزئی یک نمونه‌فرآیند جاری را به حالت جاری سیستم گذار نگاشت می‌کنیم. در رویکرد پیشنهادی، از رویکرد Polato و همکاران [۱۷] برای یافتن محتمل‌ترین توالی دنباله آینده از وضعیت جاری تا زمان اتمام الهام گرفته شده است که مسئله یافتن محتمل ترین دنباله را به مسئله کوتاه‌ترین مسیر نگاشت می‌کند. در واقع سیستم گذار حاشیه نویسی شده را می توان به صورت گراف جهت دار موزون در نظر گرفت که گره‌ها حالت ها را مشخص می‌کنند و یال‌ها گذارهای محتمل را مشخص می‌کنند و وزن‌ها احتمالات گذار را مشخص می کنند. در نتیجه، به دنبال کوتاه‌ترین مسیر را میان گره جاری و گره مقصد می باشیم که معمولا گره نهایی فرآیند می‌باشد. گرچه، در مسئله کوتاهترین مسیر، مجموع وزن‌های یال‌ها را باید کمینه‌سازی نمود. بنابراین، از لگاریتم به جای احتمالات گذار منسوب به وزن‌های یال ها استفاده می‌کنیم. علاوه براین، برای حصول اطمینان از اینکه گذارهای با

احتمال پایین مربوط به یال‌های با وزن‌های بالا می باشد و بالعکس، از مقدار مطلق لگاریتم احتمال گذار استفاده می‌کنیم. سپس باید مجموع ذیل را کمینه نماییم:

$$\sum_{i=1}^{n-1} |log\ p_i| \qquad (۱۱-۴)$$

توجه داشته باشید که در مسئله کوتاهترین مسیر، باید گره جاری و گره مقصد را مشخص کنیم، زیرا برخی فرآیندها ممکن است چندین حالت پایانی داشته باشند. راهکار اینست که یک حالت "نهایی مشترک" مشخص می‌کنیم و تمام گره های نهایی را به حالت نهایی مشترک متصل می کنیم. در نهایت، احتمال گذار از هر حالت نهایی به حالت "خروجی"، احتمال اتمام نمونه فرآیند بوسیله هر حالت نهایی را مشخص می‌کند.

۴-۶ تأثیر رانش مفهوم

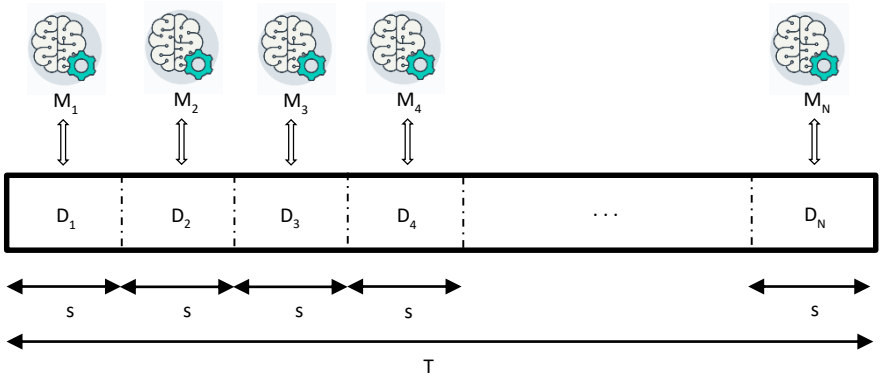
ایده کلی طرح، تقسیم داده های آموزشی به مجموعه های کوچک تر، و همچنین اطلاق وزن به مدل های ساخته شده مربوط هر مجموعه می‌باشد. این رویکرد یک رویکرد کلی می باشد و ما تأثیر آن را در پیش‌بینی زمان باقیمانده نمونه فرآیند بررسی می‌نماییم.

فرض کنید D مجموعه داده باشد و نقاط در D را با d_i نمایش می‌دهیم. در حوزه فرآیندهای سازمانی، d_i پرونده می‌باشد، اطلاعات مربوط به هر پرونده را با $d_i.x$، برچسب زمانی اولین رویداد هر پرونده را با $d_i.time$، زمان پیش بینی را نیز با $d_i.y$ نمایش می‌دهیم.

مجموعه D شامل داده های موجود در بازه زمانی $T = max_{d_i \in D} d_i.time - min_{d_i \in D} d_i.time$ می‌باشد. طول بازه تقسیم بر مجموعه داده های کوچکتر D_j می‌شود، که هرکدام طول S دارند، بطوریکه $d_i \in D_j \iff \left\lfloor \frac{d_i.time}{S} \right\rfloor = j$. در این مقاله، S در طول زمان ثابت می باشد و مقادیر متفاوت S بررسی می‌شوند. انتخاب درست مقدار S بسیار حایز اهمیت می باشد. قاب های زمانی خیلی کوچک، منجر به نقطه داده های با تعداد کم می‌شود و

همچنین قاب های زمانی بزرگ ممکن اســت از رانش مفهوم کوتاه مدت غافل شــود. تقســیم داده در شکل ۴-۵ نمایش داده شده است.

برای هر مجموعه داده D_j یک مدل یادگیری ماشــین آموزش می‌دهیم، که نهایتاً منجر به چندین کلاسه بند M_j می‌شود. ما از یک تابع وزن دهی که در ذیل توضیح داده می شود، برای تعیین اهمیت نسبی کلاسه بندها استفاده می کنیم. در مثال جاری، می توان یک دسته‌بند را به صورت ماهانه (۱ماه = S) آموزش داد که منجر به دوازده مدل در سال می شود، که هرکدام شامل تمام مقالات ارسالی در یک ماه خاص می‌شود.



شکل ٤-٥: تطبیق با رانش مفهوم با رویکرد تقسیم مجموعه داده

در هنگام پیش بینی یک داده آزمایشی، یک وزن براساس میزان قدیمی بودن مدل، به تمام مدل های M_j اطلاق می شــود. ما ترکیبی از دو وزن را پیشنهاد می دهیم. اولین وزن مربوط به تنزل نمایی می‌باشــد و از مثال ارائه شــده در [۷۳] الهام گرفته شــده است. فرض کنید d داده آزمایشی باشد، که برچســب زمانی d.time دارد. همانطور که ذکر شــد، d متعلق به دوره $j = \left\lfloor \frac{d.time}{s} \right\rfloor$ می‌باشــد. قانون وزدهی به این صــورت می باشد که وزن ۱ را به M_{j-1}، وزن $۱۰^{-\beta}$ را به M_{j-2}، وزن $۱۰^{-۲\beta}$ را به M_{j-3} انتساب می‌نماییم $(\beta \in R^+)$ و این روال را ادامه می‌دهیم. به صورت کلی، وزن $10^{-(j-k+1)\beta}$ $(k < j)$ را به مدل M_k انتساب می‌دهیم.

دومین وزن مربوط به اطلاق میزان اهمیت به مدل ها می‌باشــد، که به صــورت فصــلی به یکدیگر شــباهت دارند. ما قصــد داریم تأثیرات رانش مفهوم تکرارشــونده را که هر دوره زمانی $p \in R^+$ اتفاق

می‌افتـد، را کـاهش دهیم. بنـابراین، وزن $\tau \in R^+$ را به مدلهای مربوط به دوره‌های زمانی $d.time-p$، $d.time-2p$ و غیره انتسـاب می‌دهیم. در بسـیاری از شرایط، $p >> S$ می‌باشد و وزن نمایی در مقابل وزن تکرارشونده اهمیت به مراتب کمتری خواهد یافت.

ترکیب مدل ها بدین صـورت خواهد بود: برای داده های در D_j هرکدام از jعدد مدل پیش بینی موجود $(M_0, M_1, \dots, M_{j-1})$، یـک پیش بینی انجـام خواهنـد داد که منجر به تولید jعدد بردار به صـورت $(p_{1,k}, p_{2,k}, \dots, p_{n,k})$، که $p_{l,k}$ احتمال رخداد برچسب l در پیش بینی توسط مدل k می باشد. در اینجا n برچسـب در تمام مجموعه داده D وجود دارد. پیش بینی ترکیبی را برای داده آزمایشی d به صورت مجموع وزن دار زیر محاسبه می کنیم:

$$p_l = \sum_{0 \le k \le j-1} w_k . p_{l,k} \qquad (۱۲-۴)$$

که w_k وزن هر مدل پیش بینی را مشخص می‌کند. برچسب پیش بینی شده نهایی با $\arg\max_l p_l$ بدسـت می آید. برای تطبیق هرچه بهتر با رانش مفهوم، نیاز اسـت تا مقدار متغیرهای β، $\boldsymbol{\tau}$، p و S به بهترین شکل ممکن تنظیم شود.

الگوریتم پیش بینی با استفاده از رویکرد وفقی با رانش مفهوم:

$$j \leftarrow \left\lfloor \frac{d.time}{S} \right\rfloor$$
$$for\ k = 0 \dots j-1\ do$$
$$\quad w_k \leftarrow 10^{-(j-k+1)\beta}$$
$$end$$
$$t \leftarrow d.time - p$$
$$while\ t > 0\ do$$
$$\quad k \leftarrow \left\lfloor \frac{t}{S} \right\rfloor$$
$$\quad w_k \leftarrow w_k + \tau$$
$$\quad t \leftarrow t - p$$
$$end$$

۴-۷ آزمایش‌ها

در این بخش، ما نتایج اعمال روش پیشـــنهادی را به پایگاه داده شـــبیه ســـازی و همچنین پایگاه داده QUT تشریح می نماییم. از آنجاییکه هدف روش پیشنهادی، تطبیق با تأثیرات رانش مفهوم می‌باشد و نه تشخیص آن، امتیاز بدست آمده اهمیت دارد.

۴-۷-۱ پایگاه داده شبیه سازی

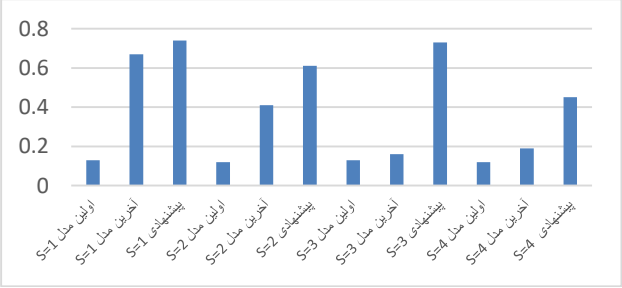
نتایج آزمایش‌ها بر روی پایگاه داده شـــبیه ســـازی شـــده در شـــکل ۴-۶ و جدول ۴-۲ آمده است و نشـــان می دهد که ۱۲ روش اعمال شـــده است. بهترین امتیازات مربوط به روش پیشنهادی در جدول ۴-۲ زیرخط دار شده است و مقادیر بهینه برای β و $\boldsymbol{\tau}$ را مشخص می کند. جدول ۴-۲ همچنین معیار دقت و F_1 را برای همه مقادیر مختلف β و $\boldsymbol{\tau}$ و S نشان می‌دهد.

در شـــکل ۴-۶، از معیار F_1 برای مقایســـه نتایج شبیه سازی استفاده می‌کنیم. روش پیشنهادی در مقابل روش اولین مدل و روش مدل اخیر ارزیابی می‌شـــود. هر روش برای مقادیر مختلفی از S محک زده شده است.

پرواضـح است که تأثیرگذاری روش پیشنهادی برای هر مقدار تعیین شده S بهتر از دو روش دیگر عمل می کند. روش اولین مدل نیز، برای هر مقدار از S ناکارآمد می‌باشد. البته کاملاً هم منطقی است، زیرا داده مربوط به یک ســـال، چهار رانش مفهوم مختلف را در بر دارد. روش اولین مدل، قادر به پیش بینی درســـت برای پایگاه داده ای که برای آن آموزش دیده نیست، چه رسد به حالتی که پایگاه داده، رانش تدریجی و رانش تکرارشونده داشته باشد.

جدول ۴-۲: نتایج آزمایش‌ها برروی پایگاه داده شبیه سازی شده. امتیازات زیرخط دار متعلق به بهترین امتیازات برای هر مقدار S می باشد.

S	τ	β = 0		β = 1		β = 2	
		ACC	F_1	ACC	F_1	ACC	F_1
1	0	0.14	0.109	0.669	0.672	0.667	0.671
	0.01	0.139	0.108	0.670	0.673	0.669	0.673
	0.1	0.154	0.130	0.671	0.674	0.670	0.674
	1	0.263	0.295	0.742	0.653	0.742	0.652
	10	0.649	0.605	0.723	0.642	0.722	0.642
	100	0.709	0.635	0.716	0.639	0.716	0.639
2	0	0.113	0.090	0.344	0.348	0.395	0.396
	0.01	0.114	0.092	0.352	0.355	0.422	0.444
	0.1	0.134	0.118	0.513	0.516	0.528	0.532
	1	0.245	0.279	0.607	0.549	0.605	0.545
	10	0.547	0.515	0.582	0.529	0.582	0.530
	100	0.575	0.526	0.577	0.527	0.577	0.527
3	0	0.123	0.090	0.161	0.171	0.161	0.171
	0.01	0.127	0.096	0.161	0.171	0.161	0.171
	0.1	0.142	0.119	0.162	0.172	0.162	0.172
	1	0.381	0.426	0.731	0.649	0.731	0.648
	10	0.724	0.641	0.734	0.646	0.734	0.646
	100	0.731	0.645	0.734	0.646	0.734	0.646
4	0	0.096	0.075	0.180	0.185	0.186	0.192
	0.01	0.097	0.077	0.185	0.190	0.188	0.194
	0.1	0.112	0.099	0.203	0.207	0.204	0.209
	1	0.215	0.241	0.454	0.430	0.454	0.428
	10	0.450	0.415	0.460	0.420	0.460	0.420
	100	0.461	0.422	0.461	0.422	0.461	0.422



شکل ۴-۶: نتایج پایگاه داده شبیه سازی شده در قالب معیار F1 می باشد. هر سه روش برای مقادیر مختلف S ارزیابی شده اند.

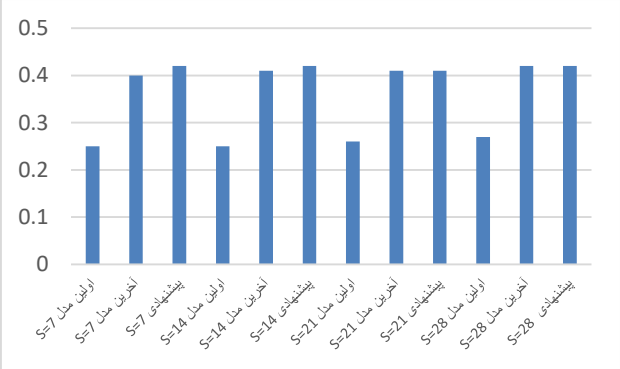
۴-۷-۲ پایگاه داده QUT

نتایج پژوهش پایگاه داده QUT در شکل ۴-۷ و جدول ۴-۳ ارائه شده است. شکل ۴-۷ امتیازات هر ۱۲روش اعمالی را نشان می‌دهد. جدول ۴-۳ نیز مقادیر β و **τ** مرتبط با هر مقدار S را نشان

می‌دهد. این پایگاه داده از تأثیر رانش مفهوم در یک فرآیند ارزیابی وام ساخته شده است. این پایگاه داده نیز در 4TU Center قابل دسترسی می باشد. از آنجاییکه که تأثیر رانش مفهوم تکرارشونده در پایگاه داده QUT وجود ندارد، نتایج روش پیشنهادی برای مقادیر مختلف S نزدیک به مقادیر روش مدل اخیر می باشد.

جدول ۴-۳: مقادیر بهینه برای β و τ و S به همراه دقت و معیار F1 برای پایگاه داده QUT

روش پیشنهادی					روش آخرین مدل		روش اولین مدل	
S	β	τ	Acc	F1	Acc	F1	Acc	F1
7	1	0	0.417	0.418	0.405	0.405	0.251	0.245
14	1	0.1	0.418	0.417	0.412	0.413	0.253	0.245
21	1	0	0.412	0.410	0.405	0.406	0.264	0.258
28	1	1	0.419	0.416	0.415	0.418	0.267	0.259



شکل ۴-۷: نتایج پایگاه داده QUT در قالب معیار F1 می باشد. هر سه روش برای مقادیر مختلف S ارزیابی شده اند.

۴-۸ نتیجه گیری

مسئله مورد بررسی در این فصل، پیش بینی زمانی نمونه‌فرآیندهای سازمانی با لحاظ رانش مفهوم می‌باشد. در شرایط رانش مفهوم و تدریجی، روش های موجود پیش‌بینی زمانی ناکارآمد می باشند. در این مقاله، رویکردی جهت حل مسئله پیش بینی را با تطبیق به شرایط رانش مفهوم ارائه گردید و با مقایسه تأثیرگذاری روش پیشنهادی در مقابل دو روش اولین مدل و آخرین مدل برروی یک پایگاه

داده شـبیه سـازی و یک پایگاه داده QUT، برتری روش پیشـنهادی اثبات گردید. در کارهای آینده، می‌توان مدل فرآیندهای سازمانی را در تطبیق با رانش مفهوم دخالت داد.

فصل ۵ : تنظیم همزمانی روند نمونه فرآیندهای وابسته

۵-۱ مقدمه

مبحث پیش بینی زمانی نمونه‌فرآیندها در نظارت و مانیتورینگ فرآیندها یک مبحث کاربردی پایه محسوب می‌شود. عملیات های پیشگیرانه در فرآیندهای سازمانی در ابتدا نیازمند پیش‌بینی زمانی هستند. یکی از این دسته عملیات‌ها، عملیات تنظیم هم‌زمانی روند پیشرفت نمونه فرآیندها می‌باشد که در درون خود، از عملیات پیش بینی زمانی به عنوان عملیات پایه‌ای بهره می‌برد.

در این فصل، یک رویکرد چهارمرحله‌ای برای مسئله تنظیم هر جفت نمونه فرآیندهای وابسته برای رسیدن به جفت وظایف وابسته متناظر به صورت همزمان یا تقریباً همزمان ارائه می‌شود. در مرحله اول، مدل فرآیندی از درون سابقه‌رویداد استخراج می گردد و با توزیع های احتمالاتی از اطلاعات زمانی غنی‌سازی می‌گردد. در مرحله دوم، وابستگی پردازشی پنهان میان هر جفت نمونه فرآیند وابسته به صورت فرمال تعریف می‌گردد و از درون سابقه‌رویداد استخراج می‌گردد. در مرحله سوم، الگوریتم پیش‌بینی وضعیت نمونه فرآیند برای پیش بینی مسیر آینده نمونه فرآیند ارائه می گردد و سپس زمان باقیمانده نمونه فرآیند تا نقطه مقصد را در مسیر پیش‌بینی شده از نمونه‌فرآیند پیش‌بینی می‌کند. در مرحله چهارم نیز، یک الگوریتم تنظیم همزمانی تکراری[۱] براساس وضعیت آتی پیش‌بینی شده نمونه‌فرآیند ارائه می شود تا هر جفت نمونه فرآیند وابسته را مجبور به رسیدن همزمان یا تقریباً همزمان به جفت وظایف وابسته نماید.

۵-۲ فاز اول: کشف فرآیند

در این مرحله، مدل فرآیند سازمانی از درون سابقه‌رویداد استخراج می‌شود و با توزیع‌های احتمالاتی از اطلاعات زمانی غنی‌سازی می‌شود. در ادبیات تحقیق، الگوریتم‌هایی همانند الگوریتم آلفا [61]، الگوریتم استخراج اکتشافی [62] و الگوریتم استخراج ژنتیک [63] را می‌توان برای کشف اتوماتیک

[۱] Iterative

فرآیند سازمانی نهفته در سابقه‌رویداد استفاده نمود. در این فصل، ما الگوریتم آلفا را برای کشف مدل فرآیندی انتخاب نمودیم. از آنجاییکه قصد داریم تغییراتی در الگوریتم آلفا ایجاد کنیم، نمادها و روال الگوریتم آلفا را در ذیل شرح می دهیم. فرض کنید W یک سابقه‌رویداد گردش کار بر روی T باشد. $\propto (W)$ به صورت زیر تعریف می‌شود.

$$1. T_W = \{t \in T \mid \exists_{\sigma \in W}\, t \in \sigma\}$$

$$2. T_I = \{t \in T \mid \exists_{\sigma \in W}\, t = first(\sigma)\}$$

$$3. T_O = \{t \in T \mid \exists_{\sigma \in W}\, t = last(\sigma)\}$$

$$4. X_W = \{(A,B) \mid A \subseteq T_W \wedge A \neq \emptyset \wedge B \subseteq T_W \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B} a \rightarrow_W b \wedge \forall_{a1,a2 \in A} a_1 \#_W a_2 \wedge \forall_{b1,b2 \in B} b_1 \#_W b_2\}$$

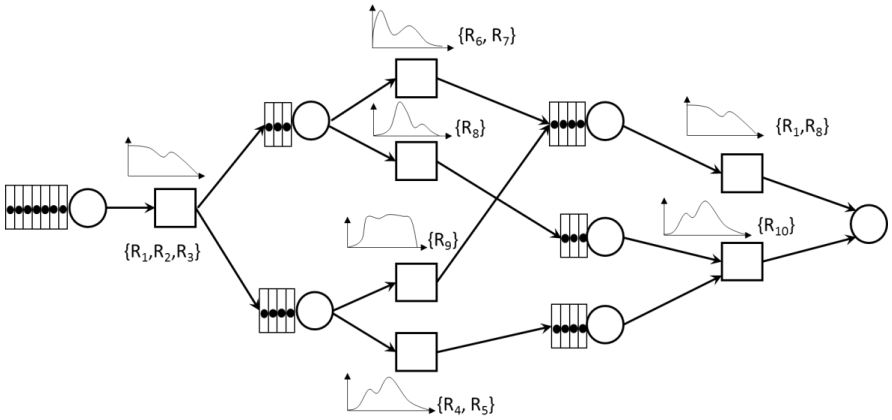
$$5. Y_W = \{(A,B) \in X \mid \forall_{(A',B') \in X} A \subseteq A' \wedge B \subseteq B' \Longrightarrow (A,B) = (A',B')\}$$

در الگوریتم آلفا، رابطه‌های میان وظیفه ها با چهار نماد $\| , \# , \rightarrow , >$ مشخص می‌شود. اگر a وb دو وظیفه باشد و $X \in (T - \{a,b\})^+$ که دنباله ای از وظایف دیگر غیر از a و b باشد. رابطه‌های میان وظیفه ها در الگوریتم آلفا به صورت زیر تعریف می شود.

- $a > b$ اگر وظیفه a در برخی دنباله ها، دقیقاً پیش از b مشاهده شود. به عبارت دیگر، رشته XabX در برخی دنباله ها مشاهده است.
- $a \rightarrow b$ اگر $a > b$ و $b \not> a$ باشد. به عبارت دیگر، رشته $XabX$ در سابقه رویداد مشاهده شده است اما هیچ گاه رشته $XbaX$ مشاهده نشده است.
- $a \# b$ اگر $a \not> b$ و $b \not> a$ باشد. به عبارت دیگر، رشته $XabX$ یا $XbaX$ در هیچ دنباله ای در سابقه رویداد مشاهده نمی‌شود.
- $a \| b$ اگر $a > b$ و $b > a$ باشد. به عبارت دیگر، در سابقه رویداد، هر دو رشته $XabX$ و $XbaX$ مشاهده شده است.

پس از آنکه مدل فرآیندی از درون سابقه رویداد استخراج گردید، مدل فرآیندی را با اطلاعات زمانی غنی‌سازی می‌کنیم؛ بدین معنا که از اطلاعات زمانی درون سابقه همانند طول‌مدت زمان هر فعالیت

استفاده می کنیم و برای هر فعالیت، تمام مشاهدات پیشین را از درون سابقه رویداد استخراج می کنیم و از روی آنها، همانند شکل ذیل یک توزیع احتمالاتی برای طول مدت زمان هر فعالیت بدست می‌آوریم.



شکل ۵-۱: مدل فرآیندی مستخرج غنی سازی شده با توزیع های احتمالاتی

۵-۳ فاز دوم: وابستگی پردازشی پنهان

در هر سابقه‌رویداد، بین وظیفه‌ها معمولاً وابستگی وجود دارد که تمام آنها توسط الگوریتم‌های کشف فرآیند استخراج نمی‌شود. از یک نگاه، وابستگی‌ها را به دو دسته محلی و غیرمحلی طبقه‌بندی می‌کنند. در این مقاله، بخش خاصی از وابستگی‌های غیرمحلی را مورد توجه قرار می دهیم که اجرای یک توکن در یک وظیفه به اجرای توکن متناظر با آن در وظیفه دیگر وابسته است. طبق وابستگی تعریف شده، یک توکن در یک وظیفه در انتظار اجرای توکن متناظر در وظیفه وابسته متناظر خود باشد. انتظار اجرای یک توکن در یک وظیفه برای اجرای توکن متناظر در وظیفه دیگر، منجر به افزایش چرخه زمانی نمونه‌فرآیندها می‌شود، در صورتی‌که می توان توکن‌های متناظر و وابسته را با کاهش تأخیر تقریباً همزمان به وظایف وابسته رساند. یکی از مفروضات این فصل اینست که دو توکن متناظر، مربوط به یک نمونه‌فرآیند یا مربوط به دو نمونه‌فرآیند مختلف می‌باشد.

سابقه‌رویداد L را به صورت زیر در نظر بگیرید.

$$L = \{<A, E, B, F, G, C, D, H>^{70},$$
$$< A, B, E, F, G, C, D, H>^{80},$$

$<A, E, F, B, G, C, D, H>^{70}$,

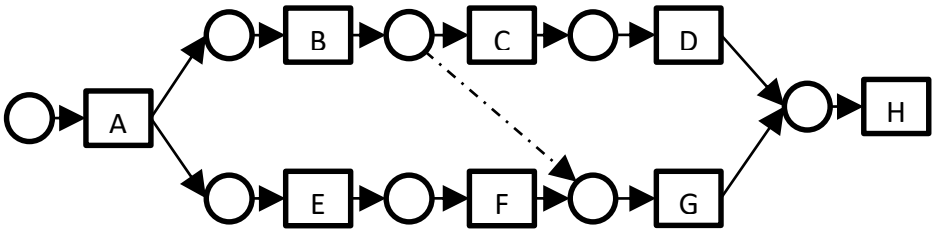
$<A, E, B, F, C, G, D, H>^{40}$,

$< A, B, E, F, C, D, G, H>^{80}$}

با بررسی سابقه‌رویداد L به این نتیجه می‌رسیم که وظیفه B همیشه پیش از وظیفه G مشاهده شده است اما همیشه وظیفه B مستقیماً پیش از G مشاهده نشده است. درصورتیکه در فرآیند شکل ذیل، انشعاب به صورت موازی طراحی شده است و وظایف B و G می توانند با هر ترتیبی اجرا شوند و حتی ممکن است وظایف دیگر بین آنها اجرا شود. وجود یک ترتیب خاص در اجرای وظایف B و G یک قاعده ضمنی در فرآیند را مشخص می کند که با نمادهای رابطه ای موجود در الگوریتم آلفا نمی توان آن را نشان داد. بنابراین نماد رابطه جدیدی را به صورت زیر تعریف می‌نماییم.

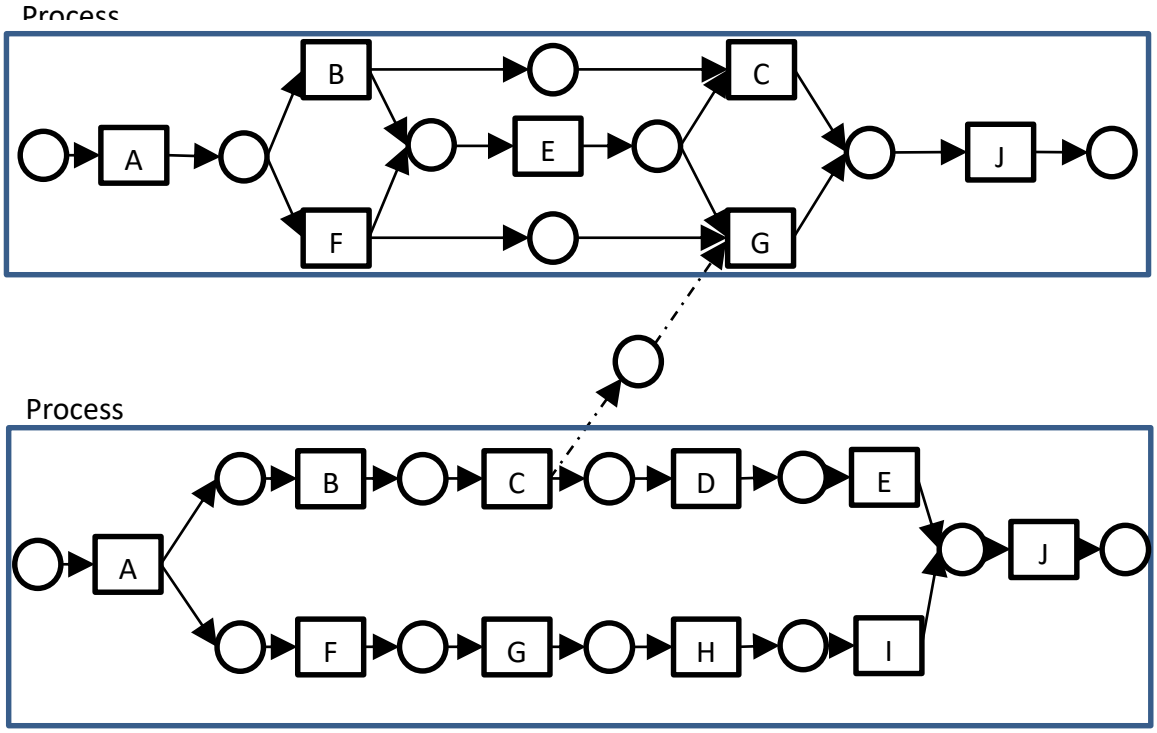
- a ::> b اگر در دنباله مربوط به یک پرونده، وظیفه a در تمام دنباله ها پیش از وظیفه b باشد و در ساختار مدل فرآیندی مستخرج، ترتیب اجرا به صورت مستقیم یا غیرمستقیم مشخص نشده باشد. این رابطه، یک نوع رابطه غیرمحلی می باشد که بین وظایف یک فرآیند تعریف می‌شود. در صورتی که برای دو وظیفه c و d رابطه d ::> c برقرار نباشد، آنرا به صورت c ::$\not>$ d نشان می‌دهیم.

طبق تعریف فوق الذکر، در فرآیند شکل ذیل، G ::> B می‌باشد. چون در سابقه رویداد L، وظیفه B همیشه پیش از وظیفه G آمده است و در ساختار مدل فرآیندی، این ترتیب به صورت مستقیم و یا غیر مستقیم مشخص نشده است. بنابراین، رابطه غیرمحلی ضمنی بین دو وظیفه B و G تعریف می شود. همچنین طبق فوق الذکر، رابطه A ::$\not>$ H برقرار می باشد زیرا در ساختار مدل فرآیند به صورت غیرمستقیم وظیفه A پیش از وظیفه H ترسیم شده است.



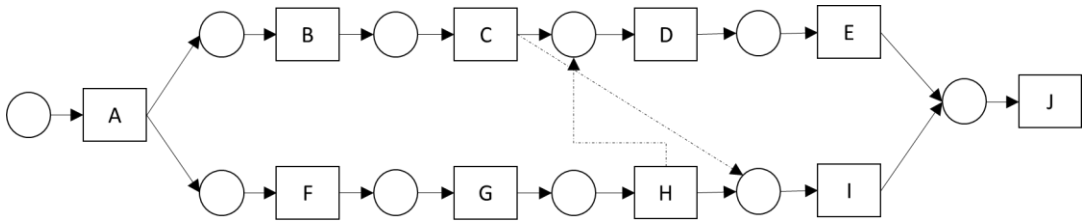
شکل ۵-۲: وابستگی پردازشی پنهان با خط چین نشان داده شده است.

تعریف ارائه شده برای رابطه <:: را می توان نیز میان دو وظیفه از دو فرآیند مختلف تعمیم داد. برای دو وظیفه از دو فرآیند مختلف، نیز می‌توان وابستگی غیر محلی ضمنی تعریف نمود به شرطی که پرونده در یک فرآیند به پرونده در فرآیند دیگر مرتبط باشد و به عبارت دیگر هر پرونده در یک فرآیند یک پرونده متناظر در فرآیند دیگر داشته باشد؛ برای مثال، زمانی‌که صاحب یک پرونده در یک فرآیند با صاحب پرونده در فرآیند دیگر یکسان باشد، ارتباط دو پرونده در دو فرآیند معنادار می‌شود و می توان رابطه <:: را میان وظایف دو فرآیند با بررسی دنباله‌های مرتبط میان آن دو فرآیند، کشف نمود.



شکل ۵-۳: وابستگی پردازشی پنهان میان دو مدل فرآیندی مختلف

در مثالی دیگر از وابستگی پنهان در یک مدل فرآیندی، همانطور که در شکل ذیل مشاهده می‌فرمایید توکن پردازش شده در A، به دو توکن در B و F تقسیم می شود. اما با آنالیز سابقه رویداد، مشخص می شود که اجرای I به اجرای به اجرای C وابسته می باشد و تا زمانی که توکن متناظر آن در مسیر BCDE، اجرای C را به اتمام نرساند، اجرای I برای توکن متناظر در مسیر FGHI آغاز نمی‌شود. از طرف دیگر، اجرای D به اجرای H وابسته می‌باشد و تا زمانی که توکن متناظر آن در مسیر FGHI، اجرای H را به اتمام نرساند، اجرای D برای توکن متناظر در مسیر BCDE آغاز نمی‌شود.



شکل ۵-۴: وابستگی پردازشی پنهان در مسیرهای موازی

۵-۴ فاز سوم: پیش‌بینی زمانی نمونه‌فرآیندها

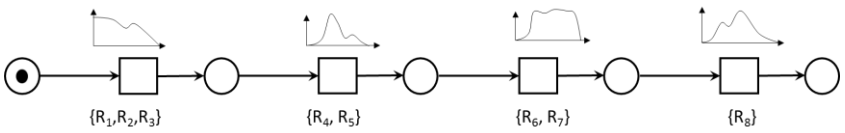
تعیین مسیر آینده نمونه فرآیندها، نیازمند تعیین مسیر در هر انشعاب فرآیند می باشد که ابهام چندمسیری را با پیش‌بینی ادامه مسیر حل نماید. بدین منظور، نمونه فرآیندها را در محل انشعاب ها دسته بندی می نماییم و هر کدام را به یک مسیر انتساب می دهیم. دسته بندی نمونه فرآیندها برای پیش بینی مسیر آینده نمونه فرآیندها با استفاده از ماشین بردار پشتیبان صورت می گیرد. دسته بند ماشین بردار پشتیبان در محل وظایفی که مابعد آن انشعاب در مسیر وجود دارد، در نظر گرفته می‌شود. بنابراین، تعداد دسته بندهای مورد نیاز در یک فرآیند برابر با تعداد وظایفی است که مابعد آن انشعاب وجود دارد. منظور از دسته‌بندی نمونه فرآیندها، در حقیقت دسته‌بندی پاره‌دنباله‌های مربوط به هر نمونه فرآیند می‌باشد. دسته بند ماشین بردار پشتیبان، فعالیت بعدی پرونده‌ها (در حقیقت پاره‌دنباله‌ها) را براساس مجموعه‌داده مربوط به آن مشخص می کند. دسته بند ماشین بردار پشتیبان، در حقیقت یکی از چند فعالیت بعدی را که در مسیر انشعاب ها قرار دارد، پیش بینی می کند. پس از آنکه دسته

بندی در محل انشعاب های موجود تا انتهای فرآیند انجام شد، مسیر آینده نمونه فرآیند مشخص می گردد.

مسیر حرکت هر پرونده در طول مسیر با یک پیمایش جزئی مشخص می شود، از آنجا که آموزش دسته بند، براساس داده های پرونده هاست، بنابراین پیمایش جزئی هر پرونده و آخرین نسخه مجموعه داده هر پرونده را به ورودی هر دسته بند ارائه می دهیم. خروجی دسته بند باید فعالیت بعدی پرونده را مشخص سازد؛ از آنجا که در مرحله آموزش قرار داریم، فعالیت بعدی پرونده از پیش مشخص می باشد و دسته بند را با آن آموزش می دهیم.

$Path \quad PathPredictionAlgorithm(ProcesModel\ P: <>, \quad Case\ case, \quad Activity\ A_C)\{$
$\quad partialTrace \leftarrow < A_C >$
$\quad while\ (A_C\ != P.END)\{$
$\quad\quad if\ ((A_C, A_{next}) \in L)\ and\ (A_{next} \in A))\{$
$\quad\quad\quad partialTrace.push(A_{next})$
$\quad\quad\quad A_C \leftarrow A_{next}$
$\quad\quad \}elseif(\ (A_C, C_{next}) \in L)\ and\ (C_{next} \in C))\{$
$\quad\quad\quad SVMClassifier \leftarrow \partial(C_{next})$
$\quad\quad A_{next} \leftarrow\ SVMClassifier(case, A_C)$
$\quad\quad partialTrace.push(A_{next})$
$\quad\quad A_C \leftarrow A_{next}$
$\quad\quad \}$
$\quad \}$
$\}$

با الگوریتم فوق، مسیر حرکت نمونه‌فرآیند پیش‌بینی می گردد. پیش از این، مدل فرآیند را با اطلاعات زمانی درون سابقه‌رویداد غنی سازی نمودیم. حال مسیر بدست آمده خطی را به صورت یک فرآیند غنی‌سازی شده در نظر می‌گیریم. در این مسیر خطی، توزیع های احتمالاتی مدت زمان انجام پردازش‌ها، در آن تعبیه شده است.



شکل ۵-۵: نمونه ای از مسیر پیش بینی شده و غنی سازی شده با اطلاعات زمانی

۵-۵ فاز چهارم: تنظیم همزمانی نمونه‌فرآیندها

برای حل مسئله افزایش زمان انتظار و افزایش چرخه زمانی برای جفت نمونه فرآیندهای با وابستگی پردازشی پنهان، یک الگوریتم تنظیم همزمانی براساس الگوریتم پیش‌بینی وضعیت فرآیند برای رساندن هر جفت نمونه‌فرآیندهای وابسته به جفت وظایف متناظر به صورت همزمان یا تقریباً همزمان ارائه می‌شود. بنابراین، هدف الگوریتم تنظیم همزمانی، کاهش اختلاف زمانی میان زمان جفت نمونه فرآیندهای وابسته می‌باشد. بدین جهت، الگوریتم تنظیم همزمانی نیازمند تخمین مجدد زمان باقیمانده تا نقطه مقصد همزمان‌سازی در فرآیند برای هردو نمونه فرآیند وابسته می‌باشد. هرزمان که هر نمونه‌فرآیند یک وظیفه به پیش می‌رود، الگوریتم تنظیم همزمانی به صورت تکراری براساس الگوریتم پیش‌بینی انجام می‌گیرد. هدف این الگوریتم تکراری، کاهش تأخیر در زمان رسیدن دو نمونه فرآیند براساس زمان های پیش‌بینی‌شده به دو وظیفه متناظر می‌باشد. کاهش تأخیر همزمان‌سازی، با تنظیم آرایش آیتم‌های کاری در لیست کاری به جهت تنظیم اولویت آیتم‌های کاری برای همزمان‌سازی می‌باشد.

بنابراین برای تنظیم همزمانی دو نمونه فرآیند وابسته، دو مهلت زمانی بدست می‌آوریم. حال باید این حرکت نمونه فرآیندها را در این مهلت‌های زمانی مدیریت کنیم. فرض کنید مهلت زمانی یک پرونده خاص از شروع زمان همزمان‌سازی تا پایان نقطه مقصد همزمان‌سازی 5 روز باشد. مسئله مهلت زمانی در فرآیند معمولاً به صورت یک مسئله سراسری در فرآیند مطرح می‌شود، برای کنترل زمانی دقیق‌تر، مسئله را به چند مسئله با حوزه محلی تبدیل می‌کنیم، به صورتی که در هر مسئله محلی، یک مهلت زمانی خاص برای هر وظیفه مشخص می‌شود. بنابراین، باید مهلت زمانی کلی یا سراسری را میان وظایف در طول مسیر تا نقطه مقصد توزیع نمود و برای هر وظیفه یک مهلت زمانی خاص مشخص نمود.

اگر بیشتر از یک لیست کاری در مسیر همزمان‌سازی برای آیتم کاری وابسته وجود داشته باشد، باید مهلت زمانی همزمان سازی را میان آن ها به عدالت توزیع نمود. مهلت زمانی شامل زمان پردازش

آیتم‌های کاری به همراه زمان انتظار آیتم‌های کاری در لیست‌های کاری پیش رو می‌باشد. لازم به ذکر است که زمان پردازش به ازای هر وظیفه ثابت در نظر گرفته نمی‌شود و در هر بار استفاده، یک عدد تصادفی برحسب توزیع احتمال مدت زمان انجام آن وظیفه، تولید می‌شود. مهلت زمانی هر نمونه‌فرآیند شامل زمان پردازش نمونه‌فرآیند در وظیفه‌های درون مسیر به علاوه زمان انتظار نمونه‌فرآیند در وظیفه‌های درون مسیر می‌باشد. بنابراین، ابتدا مجموع زمان‌های پردازش برای نمونه فرآیند به ازای وظایف در مسیر را از مهلت زمانی تعیین شده تفریق می‌کنیم. سپس، مقدار بدست آمده که مهلت زمانی برای زمان انتظار نمونه فرآیند در تمام وظایف می باشد، را به مجموع زمان‌های انتظار صف‌های لیست‌های کاری پیش رو تقسیم می کنیم؛ زمان انتظار صف هر لیست کاری با مجموع n عدد تصادفی تولید شده برحسب توزیع احتمالاتی مدت زمان انجام وظیفه بدست می آید که n تعداد آیتم های کاری در لیست کاری آن وظیفه می باشد. مقدار بدست آمده را به عنوان FairRatio می شناسیم. با ضرب FairRatio در مجموع n بار تولید عدد تصادفی طبق توزیع احتمالاتی و رند کردن آن، موقعیت جدید برای نمونه فرآیند وابسته بدست می‌آید.

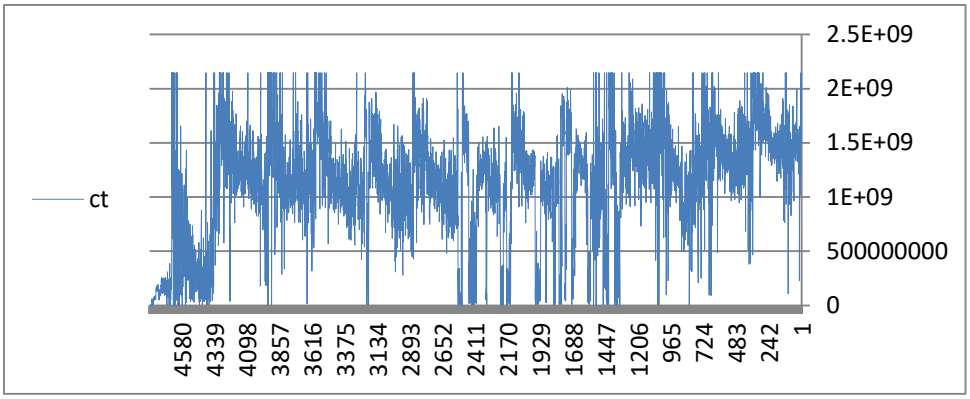
$$FairRatio = \frac{T_{deadline} - \sum_{i \,\epsilon\, \{tasks\ ahead\}} ProcessDistributionRand(i)}{\sum_{i \,\epsilon\, \{tasks\ ahead\}} \sum_{j=number\ of\ items\ in\ worklist\ i} (ProcessDistributionRand(i))} \quad (۱-۵)$$

$$NewPosition(task\ i) = Round(FairRatio \times \sum_{j=number\ of\ items\ in\ worklist\ i} ProcessDistributionRand(i)) \quad (۲-۵)$$

۵-۶ نتایج آزمایش‌ها

در شکل ذیل، چرخه زمانی ۴۷۲۰ نمونه فرآیند از سابقه رویداد *BPI Challenge 2012* رسم شده است. همانطور که مشاهده می‌نمایید، حرکت چرخه های زمانی در طول زمان، به صورت منحنی‌های نرم نمی‌باشد و دارای نوسانات و داده های شبه نویز می باشد. در این فصل، به ویژگی وابستگی پردازشی پنهان نیز اشاره کردیم. یکی از اثرات این وابستگی، افزایش چرخه زمانی برای برخی نمونه فرآیندهای

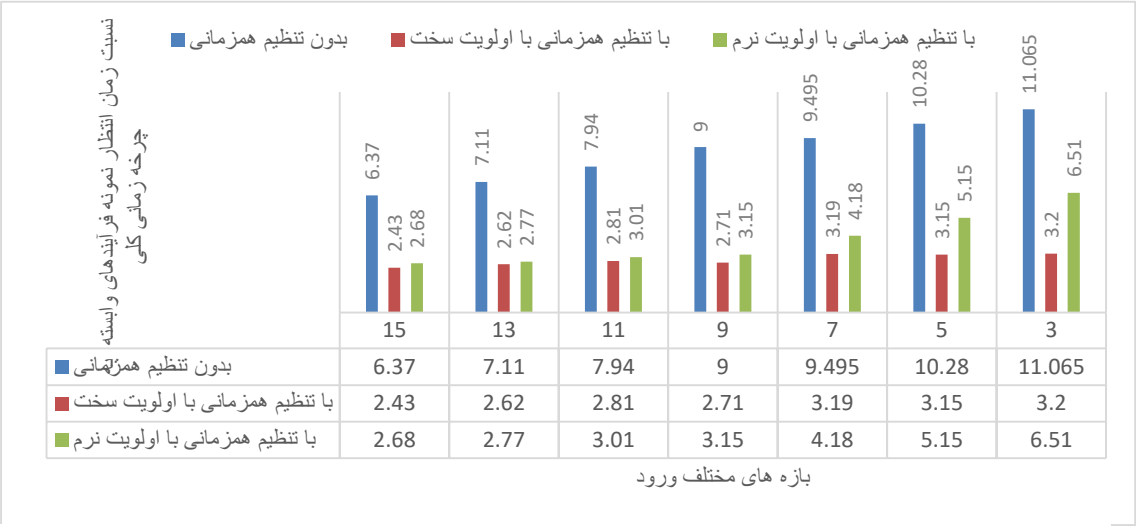
وابسته می‌باشد، زیرا زمان انتظار برای شروع پردازش برای نمونه فرآیند، چرخه زمانی آن را نیز افزایش می دهد. البته شایان ذکر است که تمام داده های شبه نویز بر وابستگی زمانی نمونه فرآیندها دلالت ندارد و نیاز به بررسی دارد.



شکل ۵-۶: نمودار چرخه زمانی نمونه فعالیت های به اتمام رسیده

در شکل ذیل، نسبت زمان انتظار نمونه فرآیندهای وابسته را به چرخه زمانی کلی برای تمام نمونه فرآیندها به ازای نرخ های ورود متفاوت در حالت بدون تنظیم همزمانی و حالت با تنظیم همزمانی نشان می‌دهد. حالت تنظیم همزمانی را با دو پیاده سازی متفاوت اعمال نمودیم. در استراتژی اولویت سخت، اولویت توکن های وابسته را در هر لیست کاری در اولویت اول قرار دادیم و در استراتژی اولویت نرم، اولویت توکن های وابسته را در هر لیست کاری از اولویت اول تا اولویت سوم به صورت تصادفی انتخاب کردیم. این دو استراتژی در پیاده سازی، از آنجا نشأت می گیرد که یک فاکتور انتخابی برای تنظیم اولویت عملیات تنظیم همزمانی براساس تنظیم اهمیت به اولویت بالاتر و اولویت‌های پایین‌تر وجود دارد. جهت آزمایش ازدحام در پردازش نمونه فرآیندها، نرخ های ورود را از ۳ تا ۱۵ تغییر دادیم. همانطور که می دانید، با کاهش نرخ بین هر دو ورود[۱]، تعداد ورود نمونه‌فرآیندها به مدل فرآیندی افزایش می‌یابد و منجر به ازدحام می شود.

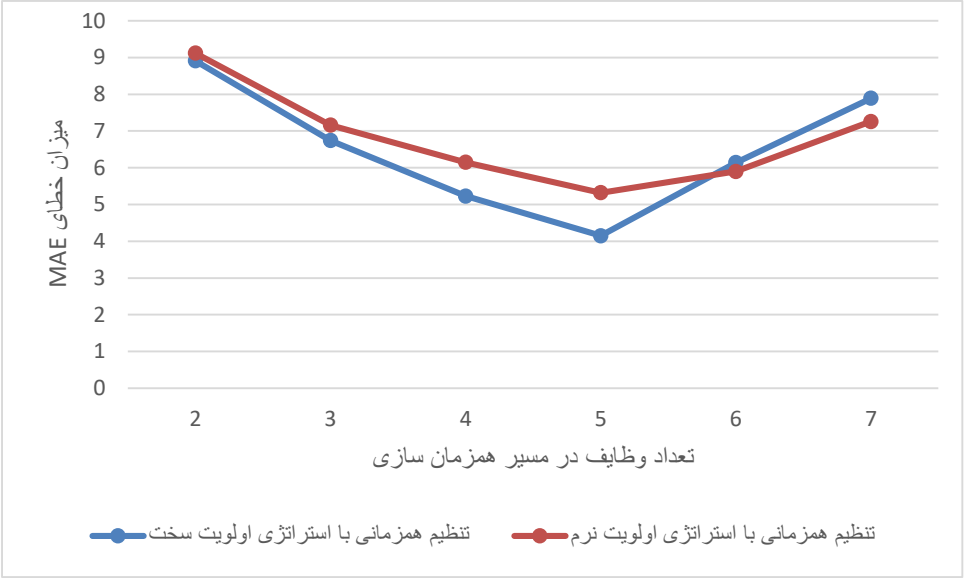
[۱] Inter-arrival rate



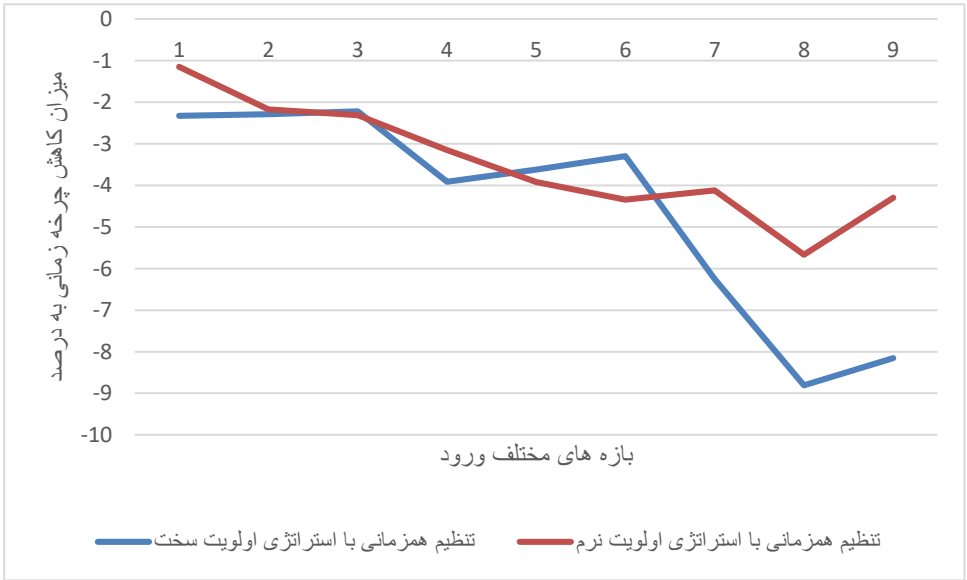
	15	13	11	9	7	5	3
بدون تنظیم همزمانی	6.37	7.11	7.94	9	9.495	10.28	11.065
با تنظیم همزمانی با اولویت سخت	2.43	2.62	2.81	2.71	3.19	3.15	3.2
با تنظیم همزمانی با اولویت نرم	2.68	2.77	3.01	3.15	4.18	5.15	6.51

شکل ۵-۷: نسبت زمان انتظار نمونه فرآیندهای وابسته به چرخه زمانی کلی به ازای بازه های مختلف ورود

از آنجاییکه در عملیات تنظیم همزمانی، تعجیل عامل مهمی محسوب می شود، تعداد وظایف در مسیر از نقطه شروع تنظیم همزمانی نقطه مقصد همزمانی، در عملکرد همزمان سازی تأثیرگذار می‌باشد. در شکل ذیل، میانگین خطای مطلق میان زمان برنامه‌ریزی‌شده رسیدن به مقصد و زمان واقعی رسیدن به مقصد را به ازای طول های مختلف مسیرهای همزمانی برای نمونه‌فرآیندهای وابسته نشان می دهد. همانطور که مشاهده می شود، زمانی که تنها دو وظیفه در طول مسیر همزمان‌سازی وجود دارد، خطای MAE بالا می باشد، زیرا برای عملیات همزمان‌سازی نیاز به اقدام در زمان مناسب دارد و با اقدام دیرتر برای تنظیم همزمانی، زمان انتظار برای توکن‌های وابسته افزایش پیدا کرده است. زمانی که در طول مسیر همزمان سازی، ۵ وظیفه وجود دارد، کمینه خطای MAE بدست می آید و بهترین نتایج را شامل می‌شود. در نهایت، با افزایش وظایف درون مسیر همزمان سازی به ۷ وظیفه، کاهش تأخیر در همزمان‌سازی کار دشواری خواهد بود زیرا همزمان سازی برپایه پیش‌بینی می‌باشد و عملیات پیش‌بینی، برای مسیرهای طولانی تر با کاهش دقت مواجه می‌شود.



شکل ذیل، تأثیر همزمان سازی را برروی کاهش چرخه زمانی کلی برای نرخ های مختلف ورود نشان می‌دهد. کاهش چرخه زمانی کلی دو استراتژی با اولویت سخت و اولویت نرم به ازای نرخ‌های مختلف بین ورود نشان داده شده است. همانطور که نشان داده شده است، استراتژی اولویت سخت از اولویت نرم برای کاهش چرخه زمانی در نرخ‌های ورود بالاتر، افزایش می‌یابد.



شکل ۵-۹: میزان کاهش چرخه زمانی به ازای بازه های مختلف ورود

۵-۷ نتیجه گیری

در این فصل، یکی از انواع وابستگی‌ها را از درون سابقه رویداد کشف نمودیم و از یک رویکرد پیشنهادی جهت تنظیم همزمانی برای هماهنگی و حل و فصل وابستگی استفاده کردیم. با تنظیم همزمانی، نه تنها چرخه زمانی کلی کاهش یافت بلکه نرمال سازی برروی چرخه زمانی نمونه فرآیندها انجام گردید. وابستگی پردازشی پنهان به وضعیتی گفته می شود که نمونه فرآیند تا انجام شدن یک نمونه فرآیند دیگر در وظیفه متناظر، در حال انتظار می‌ماند. رویکرد تنظیم همزمانی در چهار فاز ارائه می شود که در فاز اول، مدل فرآیندی از درون سابقه‌رویداد استخراج می‌گردد و با توزیع‌های احتمالاتی از اطلاعات زمانی غنی‌سازی می‌گردد. در فاز دوم، وابستگی پردازشی پنهان میان هر جفت نمونه فرآیند وابسته به صورت فرمال تعریف می‌گردد و از درون سابقه‌رویداد استخراج می‌گردد. در فاز سوم، الگوریتم پیش‌بینی وضعیت نمونه فرآیند برای پیش بینی مسیر آینده نمونه فرآیند ارائه می گردد و سپس زمان باقیمانده نمونه فرآیند تا نقطه مقصد را در مسیر پیش‌بینی شده از نمونه‌فرآیند پیش‌بینی می‌کند. در فاز چهارم نیز، یک الگوریتم تنظیم همزمانی تکراری[۱] براساس وضعیت آتی پیش‌بینی شده نمونه‌فرآیند ارائه می‌شود تا هر جفت نمونه‌فرآیند وابسته را مجبور به رسیدن همزمان یا تقریباً همزمان به جفت وظایف وابسته متناظر نماید. رویکرد پیشنهادی برروی سابقه رویداد BPI Challenge 2012 منجر به کاهش 4.3% چرخه زمانی کلی و کاهش 39% چرخه زمانی نمونه فرآیندهای وابسته گردید.

[۱] Iterative

فصل ۶ : نتیجه گیری و کارهای آتی

در این رساله، یک چارچوب برای مدیریت، کنترل، برنامه ریزی، پیش‌بینی و تنظیم پیشروی و حرکت نمونه‌فرآیندها در فرآیندهای سازمانی ارائه شده است. در مقوله پیش‌بینی زمانی حرکت نمونه‌فرآیندها در فصل سوم، یک مجموعه رویکردهای برپایه تجزیه مسیر ارائه می‌شود که برخلاف رویکردهای رگرسیون سراسری، راهکاری مشابه رویکرد تقسیم و غلبه ارائه می‌کند و عملیات پیش‌بینی نمونه‌فرآیندهای سازمانی را به پیش‌بینی طول مدت زمان فعالیت‌های در مسیر و همچنین پیش‌بینی احتمال انشعاب‌های در مسیر تقسیم می‌کند. در رویکرد اول، برپایه آنالیز تحلیلی مسیر، برای پیش‌بینی در محل انشعابات و همچنین طول مدت زمان فعالیت‌ها، برای تمام اندازه های مختلف پیشوند دنباله‌های جزئی، یک مدل یادگیری ماشین بردار پشتیبان در نظر گرفته می‌شود. در رویکرد دوم برپایه آنالیز تحلیلی مسیر، برای پیش‌بینی در محل انشعابات و همچنین طول مدت زمان فعالیت‌ها، به ازای هر اندازه پیشوند دنباله‌های جزئی، یک مدل یادگیری ماشین بردار پشتیبان مجزا در نظر گرفته می‌شود. در رویکرد سوم برپایه آنالیز تحلیلی مسیر، برای پیش‌بینی در محل انشعابات و همچنین طول مدت زمان فعالیت‌ها، از اعداد ثابت میانگین مقادیر سابقه‌رویداد استفاده می‌کند. در رویکرد چهارم برپایه آنالیز تحلیلی مسیر، از ترکیب مدل‌های یادگیری ماشین و مقادیر میانگین به صورت ترکیبی بکار گرفته می‌شود. در رویکرد پنجم پیشنهادی از توزیع‌های احتمالاتی و احتساب صف انتظار وظیفه‌ها، برای پیش‌بینی زمان باقیمانده نمونه فرآیند استفاده می شود. آزمایش‌ها نشان می‌دهد رویکرد ترکیبی ارائه شده به نسبت رویکرد رگرسیون‌های محلی در پیش‌بینی چرخه زمانی، خطای MAE نرمال سازی شده را مجموعاً برای تمام ۸ پایگاه‌داده به اندازه ۱۲,۸۳ درصد کاهش می‌دهد.

در چارچوب پیشنهادی ارائه شده در حوزه زمانی فرآیندهای سازمانی، پیش‌بینی دقیق زمان باقیمانده نمونه‌فرآیند بسیار حیاتی می‌باشد، اما رویکردهای پیش‌بینی در مقابل رانش مفهوم مقاوم نیستند و دچار تزلزل می‌شوند. منظور از رانش مفهوم در فرآیندهای سازمانی، تغییر در رفتار فرآیند به علت تغییر در مدل فرآیندی درون سابقه رویداد می‌باشد. در فصل چهارم، یک رویکرد پیشنهادی، جهت پیش‌بینی زمان باقیمانده نمونه فرآیند به همراه یک روش تطبیق با رانش مفهوم، ارائه می‌شود. در الگوریتم

پیش‌بینی زمانی پیشنهادی، یک سیستم گذار ساخته می شود و با احتمالات بدست آمده از ماشین‌های بردار پشتیبان فازی براساس داده نمونه‌فرآیند حاشیه‌نویسی می‌شود. توالی فعالیت‌های آینده نمونه‌فرآیند با استفاده از الگوریتم کوتاهترین مسیر برروی سیستم گذار حاشیه نویسی شده بدست می‌آید. پس از آن، مقدار پیش‌بینی زمان باقیمانده نمونه‌فرآیند را با محاسبه مجموع طول مدت فعالیت‌های آینده که با رگرسورهای بردار پشتیبان تخمین زده شده، بدست می‌آید. به جهت مقاوم سازی الگوریتم پیش بینی زمانی در مقابل پویایی و تغییر مدل فرآیندی، که با نام رانش مفهوم نیز شناخته می‌شود، راهکار پیشنهادی انطباق با رانش مفهوم پیشنهاد ارائه می شود که از مدل‌های یادگیری ماشین چندگانه بهره می‌برد که هر مدل، ازروی بازه‌های زمانی مختلف از سابقه‌رویداد استخراج می‌شود و به هرکدام وزنی اطلاق می‌شود که ترکیبی از "ضریب زوال" و همچنین "ضریب دوره‌ای" می‌باشد. ضریب زوال، ضریبی است که با افزایش فاصله زمانی مدل اکتشافی از نمونه فرآیند تحت آزمایش، کاهش می‌یابد. ضریب دوره ای هم، اهمیت داده‌های فصلی دوره‌ای را یکسان سازی می‌کند. رویکرد پویای چندمدله پیشنهادی در فصل چهارم برای مقابله با رانش مفهوم بهبود ۲۲,۵۶ درصدی در معیار دقت نسبت به رویکرد تک مدله استاتیک مجموعاً بر روی دو پایگاه داده QUT و پایگاه داده شبیه سازی شده را نشان می‌دهد.

مبحث پیش بینی زمانی نمونه‌فرآیندها در نظارت و مانیتورینگ فرآیندها یک مبحث کاربردی پایه در چارچوب مدیریت زمانی فرآیندهای سازمانی محسوب می‌شود. عملیات های پیشگیرانه در فرآیندهای سازمانی در ابتدا نیازمند پیش‌بینی زمانی هستند. یکی از این دسته عملیات‌ها، عملیات تنظیم هم‌زمانی روند پیشرفت نمونه فرآیندها می‌باشد که در درون خود، از عملیات پیش بینی زمانی به عنوان عملیات پایه‌ای بهره می‌برد. در فصل پنجم، یک رویکرد چهارمرحله‌ای برای مسئله تنظیم هر جفت نمونه فرآیندهای وابسته برای رسیدن به جفت وظایف وابسته متناظر به صورت همزمان یا تقریباً همزمان ارائه می‌شود. در مرحله اول، مدل فرآیندی از درون سابقه‌رویداد استخراج می گردد و با توزیع های احتمالاتی از اطلاعات زمانی غنی‌سازی می‌گردد. در مرحله دوم، وابستگی پردازشی پنهان میان هر جفت نمونه

فرآیند وابسته به صورت رسمی تعریف می‌گردد و از درون سابقه‌رویداد استخراج می‌گردد. در مرحله سوم، الگوریتم پیش‌بینی وضعیت نمونه فرآیند برای پیش بینی مسیر آینده نمونه فرآیند ارائه می گردد و سپس زمان باقیمانده نمونه فرآیند تا نقطه مقصد را در مسیر پیش‌بینی شده از نمونه‌فرآیند پیش‌بینی می‌کند. در مرحله چهارم نیز، یک الگوریتم تنظیم همزمانی تکراری[1] براساس وضعیت آتی پیش‌بینی شده نمونه‌فرآیند ارائه می شود تا هر جفت نمونه فرآیند وابسته را مجبور به رسیدن همزمان یا تقریباً همزمان به جفت وظایف وابسته نماید. آزمایش‌ها برروی پایگاه داده BPI Challenge 2012 نشان از کاهش 4.3% چرخه زمانی کلی و کاهش 39% چرخه زمانی نمونه فرآیندهای وابسته نسبت به حالت بدون تنظیم همزمانی دارد.

در کارهای آتی پیرامون این رساله، می‌توان مبحث فرآیندکاوی را با مبحث پردازش متن ادغام کرده و در پیش‌بینی انشعابات از داده‌های متنی در بسته‌داده نمونه‌فرآیندها استفاده نمود. برای نمونه، در فرآیند پروپوزال دانشجویان، با بررسی محتوای متن پروپوزال می‌توان پیش‌بینی چرخه زمانی و مدت زمان انجام پایان‌نامه دانشجویی را با تخمین بهتری انجام داد. در کارهای آتی همچنین می‌توان انحرافات را در پیش‌بینی نمونه‌فرآیندهای منطبق نشده بر روی مدل فرآیندی در فرآیندهای سازمانی در نظر گرفت. پیش‌بینی زمانی را با لحاظ فاکتورهای دیگر همانند هزینه، ریسک، کیفیت سرویس و دیگر فاکتورها به صورت چندهدفه انجام دهیم. پیش بینی گلوگاه را به عنوان یک هدف بررسی و آزمایش و آنالیز نماییم و تأثیر آن را در پیش‌بینی زمانی مورد بررسی قرار دهیم. یکی از علل کاهش در دقت پیش‌بینی چرخه زمانی، دشواری در تخمین زمان حلقه‌های تکرار کاری می باشد که در کارهای آینده لحاظ خواهد شد و احتمالات آن به صورت دقیق‌تر برای هر بار تکرار آن مشخص خواهد شد.

در پیش بینی زمانی با لحاظ رانش مفهوم، تمام انواع رانش مفهوم در مدل فرآیندی شامل رانش مفهوم تکرارشونده، رانش مفهوم افزایشی، رانش مفهوم تدریجی، رانش مفهوم ناگهانی را به صورت جداگانه مورد بررسی قرار دهیم. نقطه دقیق رخداد رانش مفهوم را پیش از انجام انطباق تعیین نماییم

[1] Iterative

و رانش مفهوم را به ابعاد دیگر شامل بعد داده و بعده منابع نیز گسترش دهیم و رانش مفهوم در آن ابعاد را نیز در پیش‌بینی های کمی و کیفی دخیل کنیم.

در مقوله وابستگی‌های بین نمونه فرآیندها، می‌توان وابستگی‌های داده‌ای، معنایی و اجرایی را از درون سابقه‌رویداد کشف کنیم. در کارهای آینده همچنین می‌توان ابتدا موجودیت‌های معنایی، ساختاری از داده‌ها استخراج نمود و براساس ماتریس‌های وابستگی و گراف‌های وابستگی، برنامه‌ریزی برای پیشروی و حرکت نمونه‌فرآیندها انجام داد. کشف وابستگی‌ها، نوعی از کشف قواعد سازمانی نیز می‌باشد؛ با متدهای قاعده کاوی از روی سابقه‌رویداد می‌توان در یافتن نمونه‌فرآیندهای وابسته نیز بهره برد.

محیط‌های بیمارستانی، محیط‌های بسیار پویایی در زمینه فرآیندهای سازمانی می‌باشند و کاوش دقیق مدل فرآیندی و پیش‌بینی گلوگاه در این نوع از محیط‌ها بسیار حیاتی و پیچیده می باشد. پیچیدگی فرآیندهای این سازمان‌ها، ناشی از پیچیدگی تابع رفتار منابع آنها می‌باشد و بررسی این سابقه‌رویدادها، نیازمند تحلیل نوسانات رفتار منابع این محیط‌ها و تشخیص دقیق تابع رفتار آنها می‌باشد. ازاین‌رو در کارهای آینده، نیاز به فرآیندکاوی به صورت برخط با دریافت جریان پویای سابقه‌رویداد می‌باشد که متناسب آن تحلیل و آنالیز داده‌ها را نیز بتوان برخط انجام داد تا نهایتاً بتوان واکنش پویای سریع را به تغییرات سازمانی داشت.

از دیگر مسائل در پژوهش های حوزه فرآیندهای سازمانی، کمبود سابقه‌رویداد با ویژگی‌های کافی همانند اطلاعات زمانی صف‌ها، اطلاعات هزینه‌ای، اطلاعات کیفیت سرویس‌دهی و پشتیبانی از چندفرآیندی به صورت همزمان در یک سابقه‌رویداد می‌باشد. بحث مدیریت چندفرآیندی و کاوش چندفرآیندی یکی از مباحث روز این حوزه است که رویکردهای این رساله را با سابقه رویدادهای چندفرآیندی باید برروی آن آزمایش نمود. ساخت سابقه رویداد با این ویژگی‌ها و انجام آزمایش‌هایی نظیر تنظیم همزمانی روند نمونه‌فرآیندهای وابسته از جمله کارهای پیش رو می‌باشد.

مراجع

[1] W. Van Der Aalst, Process mining: discovery, conformance and enhancement of business processes, Springer Science & Business Media, 2011.
[2] D. Hollingsworth, The workflow reference model: 10 years on, in: Fujitsu Services, UK; Technical Committee Chair of WfMC, Citeseer, 2004.
[3] A. Rogge-Solti, M. Weske, Prediction of business process durations using non-Markovian stochastic Petri nets, Information Systems, (2015).
[4] M. Hammer, J. Champy, Reengineering the Corporation: Manifesto for Business Revolution, A, Zondervan, 2009.
[5] M. Weske, Business process management: concepts, languages, architectures, Springer Science & Business Media, 2012.
[6] W.M. Van Der Aalst, A.H. Ter Hofstede, M. Weske, Business process management: A survey, in: Business process management, Springer, 2003, pp. 1-12.
[7] H.A. Reijers, S.L. Mansar, Best practices in business process redesign: an overview and qualitative evaluation of successful redesign heuristics, Omega, 33 (2005) 283-306.
[8] W.M. Van der Aalst, Verification of workflow nets, in: Application and Theory of Petri Nets 1997, Springer, 1997, pp. 407-426.
[9] M. Jansen-Vullers, M. Netjes, Business process simulation–a tool survey, in: Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools, Aarhus, Denmark, 2006.
[10] R. Lenz, M. Reichert, IT support for healthcare processes–premises, challenges, perspectives, Data & Knowledge Engineering, 61 (2007) 39-58.
[11] A. Rozinat, R.S. Mans, M. Song, W.M. van der Aalst, Discovering simulation models, Information Systems, 34 (2009) 305-327.
[12] W.M. van der Aalst, Process mining and simulation: a match made in heaven!, in: Proceedings of the 50th Computer Simulation Conference, Society for Computer Simulation International, 2018, pp. 4.
[13] W.M. Van der Aalst, M.H. Schonenberg, M. Song, Time prediction based on process mining, Information Systems, 36 (2011) 450-475.
[14] M. Polato, A. Sperduti, A. Burattin, M. de Leoni, Time and activity sequence prediction of business process instances, Computing, 100 (2018) 1005-1031.
[15] M. Dumas, F.M. Maggi, Enabling process innovation via deviance mining and predictive monitoring, in: BPM-Driving Innovation in a Digital World, Springer, 2015, pp. 145-154.
[16] W.M. Van der Aalst, M. Pesic, M. Song, Beyond process mining: from the past to present and future, in: International Conference on Advanced Information Systems Engineering, Springer, 2010, pp. 38-52.
[17] M. Polato, A. Sperduti, A. Burattin, M. de Leoni, Data-aware remaining time prediction of business process instances, in: 2014 International Joint Conference on Neural Networks (IJCNN), IEEE, 2014, pp. 816-823.
[18] F. Folino, M. Guarascio, L. Pontieri, Discovering high-level performance models for ticket resolution processes, in: OTM Confederated International Conferences" On the Move to Meaningful Internet Systems", Springer, 2013, pp. 275-282.
[19] F. Folino, M. Guarascio, L. Pontieri, Discovering context-aware models for predicting business process performances, in: On the Move to Meaningful Internet Systems: OTM 2012, Springer, 2012, pp. 287-304.

[20] M. Ceci, P.F. Lanotte, F. Fumarola, D.P. Cavallo, D. Malerba, Completion time and next activity prediction of processes using sequential pattern mining, in: International Conference on Discovery Science, Springer, 2014, pp. 49-61.
[21] A. Rogge-Solti, M. Weske, Prediction of remaining service execution time using stochastic petri nets with arbitrary firing delays, in: International Conference on Service-Oriented Computing, Springer, 2013, pp. 389-403.
[22] B.F. van Dongen, R.A. Crooy, W.M. van der Aalst, Cycle time prediction: When will this case finally be finished?, in: OTM Confederated International Conferences" On the Move to Meaningful Internet Systems", Springer, 2008, pp. 319-336.
[23] E. Cesario, F. Folino, M. Guarascio, L. Pontieri, A cloud-based prediction framework for analyzing business process performances, in: International Conference on Availability, Reliability, and Security, Springer, 2016, pp. 63-80.
[24] A. Bevacqua, M. Carnuccio, F. Folino, M. Guarascio, L. Pontieri, A Data-adaptive Trace Abstraction Approach to the Prediction of Business Process Performances, in: ICEIS (1), 2013, pp. 56-65.
[25] S. Pandey, S. Nepal, S. Chen, A test-bed for the evaluation of business process prediction techniques, in: 7th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), IEEE, 2011, pp. 382-391.
[26] A. Senderovich, C. Di Francescomarino, C. Ghidini, K. Jorbina, F.M. Maggi, Intra and inter-case features in predictive process monitoring: A tale of two dimensions, in: International Conference on Business Process Management, Springer, 2017, pp. 306-323.
[27] A. Senderovich, M. Weidlich, A. Gal, A. Mandelbaum, Queue mining for delay prediction in multi-class service processes, Information Systems, 53 (2015) 278-295.
[28] T.B.H. Tu, M. Song, Analysis and prediction cost of manufacturing process based on process mining, in: 2016 International Conference on Industrial Engineering, Management Science and Application (ICIMSA), IEEE, 2016, pp. 1-5.
[29] R. Conforti, M. De Leoni, M. La Rosa, W.M. Van Der Aalst, Supporting risk-informed decisions during business process execution, in: International Conference on Advanced Information Systems Engineering, Springer, 2013, pp. 116-132.
[30] R. Conforti, S. Fink, J. Manderscheid, M. Röglinger, PRISM–a predictive risk monitoring approach for business processes, in: International Conference on Business Process Management, Springer, 2016, pp. 383-400.
[31] R. Conforti, M. de Leoni, M. La Rosa, W.M. van der Aalst, A.H. ter Hofstede, A recommendation system for predicting risks across multiple business process instances, Decision Support Systems, 69 (2015) 1-19.
[32] A. Metzger, P. Leitner, D. Ivanović, E. Schmieders, R. Franklin, M. Carro, S. Dustdar, K. Pohl, Comparing and combining predictive business process monitoring techniques, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 45 (2014) 276-290.
[33] A. Pika, W.M. van der Aalst, C.J. Fidge, A.H. ter Hofstede, M.T. Wynn, Predicting deadline transgressions using event logs, in: International Conference on Business Process Management, Springer, 2012, pp. 211-216.
[34] A. Pika, W.M. Van Der Aalst, C.J. Fidge, A.H. Ter Hofstede, M.T. Wynn, Profiling event logs to configure risk indicators for process delays, in: International Conference on Advanced Information Systems Engineering, Springer, 2013, pp. 465-481.
[35] N. Tax, I. Verenich, M. La Rosa, M. Dumas, Predictive business process monitoring with LSTM neural networks, in: International Conference on Advanced Information Systems Engineering, Springer, 2017, pp. 477-492.

[36] J. Evermann, J.-R. Rehse, P. Fettke, Predicting process behaviour using deep learning, Decision Support Systems, 100 (2017) 129-140.
[37] C. Di Francescomarino, C. Ghidini, F.M. Maggi, G. Petrucci, A. Yeshchenko, An eye into the future: leveraging a-priori knowledge in predictive business process monitoring, in: International Conference on Business Process Management, Springer, 2017, pp. 252-268.
[38] J. Evermann, J.-R. Rehse, P. Fettke, A deep learning approach for predicting process behaviour at runtime, in: International Conference on Business Process Management, Springer, 2016, pp. 327-338.
[39] J.G. De Gooijer, R.J. Hyndman, 25 years of time series forecasting, International journal of forecasting, 22 (2006) 443-473.
[40] L. Zeng, C. Lingenfelder, H. Lei, H. Chang, Event-driven quality of service prediction, in: Service-Oriented Computing–ICSOC 2008, Springer, 2008, pp. 147-161.
[41] W.M. Van Der Aalst, K.M. Van Hee, H.A. Reijers, Analysis of discrete-time stochastic petri nets, Statistica Neerlandica, 54 (2000) 237-255.
[42] H.A. Reijers, Design and control of workflow processes: business process management for the service industry, Springer-Verlag, 2003.
[43] A. Rogge-Solti, Probabilistic Estimation of Unobserved Process Events, (2014).
[44] S.-Y. Hwang, H. Wang, J. Tang, J. Srivastava, A probabilistic approach to modeling and estimating the QoS of web-services-based workflows, Information Sciences, 177 (2007) 5484-5503.
[45] B. Kiepuszewski, A.H.M. ter Hofstede, C.J. Bussler, On structured workflow modelling, in: Advanced Information Systems Engineering, Springer, 2000, pp. 431-445.
[46] A. Wombacher, M. Iacob, Estimating the Processing Time of Process Instances in Semi-structured Processes--A Case Study, in: Services Computing (SCC), 2012 IEEE Ninth International Conference on, IEEE, 2012, pp. 368-375.
[47] K.P. Anklesaria, Z. Drezner, A multivariate approach to estimating the completion time for PERT networks, Journal of the Operational Research Society, (1986) 811-815.
[48] A. Senderovich, M. Weidlich, A. Gal, A. Mandelbaum, Queue mining–predicting delays in service processes, in: Advanced Information Systems Engineering, Springer, 2014, pp. 42-57.
[49] M. de Leoni, W.M. van der Aalst, M. Dees, A general framework for correlating business process characteristics, in: Business Process Management, Springer, 2014, pp. 250-266.
[50] A. Rozinat, M.T. Wynn, W.M. van der Aalst, A.H. ter Hofstede, C.J. Fidge, Workflow simulation for operational decision support, Data & Knowledge Engineering, 68 (2009) 834-850.
[51] M. Rosemann, M. zur Muehlen, Integrating risks in business process models, ACIS 2005 Proceedings, (2005) 50.
[52] S. Suriadi, B. Weiß, A. Winkelmann, A.H. ter Hofstede, M. Adams, R. Conforti, C. Fidge, M. La Rosa, C. Ouyang, M. Rosemann, Current research in risk-aware business process management: overview, comparison, and gap analysis, Communications of the Association for Information Systems, 34 (2014) 933-984.
[53] A. Pika, W.M. van der Aalst, C.J. Fidge, A.H. ter Hofstede, M.T. Wynn, Predicting deadline transgressions using event logs, in: Business Process Management Workshops, Springer, 2013, pp. 211-216.
[54] G. Goluch, A. Ekelhart, S. Fenz, S. Jakoubi, S. Tjoa, T. Mück, Integration of an ontological information security concept in risk aware business process management, in:

Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, IEEE, 2008, pp. 377-377.
[55] R. Conforti, M. De Leoni, M. La Rosa, W.M. van der Aalst, Supporting risk-informed decisions during business process execution, in: Advanced Information Systems Engineering, Springer, 2013, pp. 116-132.
[56] B. Kang, D. Kim, S.-H. Kang, Periodic performance prediction for real-time business process monitoring, Industrial Management & Data Systems, 112 (2012) 4-23.
[57] P. Leitner, B. Wetzstein, F. Rosenberg, A. Michlmayr, S. Dustdar, F. Leymann, Runtime prediction of service level agreement violations for composite services, in: Service-Oriented Computing. ICSOC/ServiceWave 2009 Workshops, Springer, 2010, pp. 176-186.
[58] M. Dumas, M. La Rosa, J. Mendling, H.A. Reijers, Fundamentals of business process management, Springer, 2013.
[59] J. Cardoso, A. Sheth, J. Miller, J. Arnold, K. Kochut, Quality of service for workflows and web service processes, Web Semantics: Science, Services and Agents on the World Wide Web, 1 (2004) 281-308.
[60] D. Blumenfeld, Operations research calculations handbook, Crc Press, 2009.
[61] W. Van der Aalst, T. Weijters, L. Maruster, Workflow mining: Discovering process models from event logs, IEEE Transactions on Knowledge and Data Engineering, 16 (2004) 1128-1142.
[62] A. Weijters, W.M. van Der Aalst, A.A. De Medeiros, Process mining with the heuristics miner-algorithm, Technische Universiteit Eindhoven, Tech. Rep. WP, 166 (2006) 1-34.
[63] A.A. De Medeiros, A. Weijters, Genetic process mining, in: Applications and Theory of Petri Nets 2005, Volume 3536 of Lecture Notes in Computer Science, Citeseer, 2005.
[64] S.J. Raudys, A.K. Jain, Small sample size effects in statistical pattern recognition: Recommendations for practitioners, IEEE Transactions on Pattern Analysis & Machine Intelligence, (1991) 252-264.
[65] I. Teinemaa, M. Dumas, M.L. Rosa, F.M. Maggi, Outcome-oriented predictive process monitoring: review and benchmark, ACM Transactions on Knowledge Discovery from Data (TKDD), 13 (2019) 17.
[66] M. De Leoni, W.M. van der Aalst, M. Dees, A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs, Information Systems, 56 (2016) 235-257.
[67] A. Leontjeva, R. Conforti, C. Di Francescomarino, M. Dumas, F.M. Maggi, Complex symbolic sequence encodings for predictive monitoring of business processes, in: International Conference on Business Process Management, Springer, 2016, pp. 297-313.
[68] J.C. Schlimmer, R.H. Granger, Beyond Incremental Processing: Tracking Concept Drift, in: AAAI, 1986, pp. 502-507.
[69] R.J.C. Bose, W.M. Van Der Aalst, I. Žliobaitė, M. Pechenizkiy, Dealing with concept drifts in process mining, IEEE transactions on neural networks and learning systems, 25 (2013) 154-171.
[70] W.M. Van der Aalst, V. Rubin, H. Verbeek, B.F. van Dongen, E. Kindler, C.W. Günther, Process mining: a two-step approach to balance between underfitting and overfitting, Software & Systems Modeling, 9 (2010) 87.
[71] C.-F. Lin, S.-D. Wang, Fuzzy support vector machines, IEEE transactions on neural networks, 13 (2002) 464-471.

[72] Y. Liu, H. Huang, Fuzzy support vector machines for pattern recognition and data mining, International journal of fuzzy systems, 4 (2002) 826-835.
[73] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, A. Bouchachia, A survey on concept drift adaptation, ACM computing surveys (CSUR), 46 (2014) 44.

Abstract

Nowadays, organizations record their activities and events by information systems such as Enterprise Resource Planning (ERP) systems and Business Process Management Systems (BPMS). Process mining is a field wich handles process discovery, process monitoring and improvement of business processes from event logs. Intelligent management of business processes leads to higher efficiency, higher flexibility and quality assurance. Each event in the event log is specified by a set of properties such as time information. Time control of events of process instances using historical information is a challenge for researchers of this domain. Time control of events includes remaining time prediction of process instances, alignment of arrive time of process instances, synchronization of process instances by controlling lateness and earliness.

Investigating state-of-the-art approaches, five approaches are proposed in order to predict remaining cycle time of process instances; four of which are based on path decomposition technique and analytical path formulation which leverages Support Vector Machines. The other is a probabilistic approach taking queues of tasks into consideration. Experiments show that the proposed adaptive approach decreased normalized MAE error by 12.83% in total over 8 datasets compared to multi regression approach. Then, concept drift is also added to assumptions of the problem to make it closer to real-life settings. A statistical approach based on a transition system annotated with Fuzzy SVM probabilities is presented to predict the remaining time of business processes. The proposed dynamic multi-model approach to concept drift adaptation improved the accuracy by 22.56% in total compared to static single model approach over two datasets. Finally, a processing dependency between pairs of dependent process instances is formally defined and extracted from event log. To resolve the issue, time synchronization algorithm is presented to synchronize and control the events of process instances. Experiments on BPI challenge 2012 dataset show 4.3% reduction in overall cycle time and 39% reduction in dependent process instances.

Keywords: Business Processes, Remaining Time Prediction, Concept Drift, Synchronization of Tokens, Dependent Process instances.

Shahrood University of Technology

Faculty of Computer Engineering

Ph.D. Thesis in in Artificial Intelligence

A probabilistic framework to control time of events across business processes with shared resources

By: Iman Firouzian

Supervisor:
Dr. Morteza Zahedi

Advisor:
Dr. Hamid Hassanpour

December, 2019