بسم الله الرحمن الرحیم

باسمه تعالی

دانشکده پردیس بین‌الملل خوارزمی

رساله دکتری هوش مصنوعی

ساده‌سازی فرآیند با استفاده از آنالیز آماری جریان‌های کاری مشابه

نگارنده: طاهره کوهی‌ور دهکردی

استاد راهنما: دکتر مرتضی زاهدی

استاد مشاور: دکتر علی‌اکبر پویان

شهریور ۹۷

تقدیم به:

پدر و مادرم

تقدیر و تشکر

سپاس بی حد خداوند مهربان را سزاست که به انسان توانایی و دانایی بخشید تا به بندگانش شفقت ورزد و مهربانی کند و در حل مشکلاتشان یاری شان نماید. نه فقط شکر خالق که شکر مخلوق هم سزاست چراکه "من لم یشکر المخلوق، لم یشکر الخالق؛ بدین جهت لازم می دانم از استاد محترم جناب آقای دکتر مرتضی زاهدی که زحمت راهنمایی این رساله را بر عهده داشتند و همهٔ عزیزانی که در مراحل مختلف رساله مرا یاری نموده اند، صمیمانه تشکر و قدردانی نمایم و سپاس پایانی را به خانواده عزیزم تقدیم می کنم که همراهان همیشگی ام بودند و رنج سال های دوری و تحصیلم را به جان خریدند.

تعهد نامه

اینجانب طاهره کوهی ور دهکردی دانشجوی دوره دکتری رشته کامپیوتر/هوش مصنوعی دانشکده پردیس/ مجازی دانشگاه صنعتی شاهرود نویسنده رساله ساده‌سازی فرآیند با استفاده از آنالیز آماری جریان‌های کاری مشابه تحت راهنمائی دکتر مرتضی زاهدی متعهد می‌شوم:

- **تحقیقات در این رساله 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 رساله 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 می باشد 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 رساله تأثیرگذار بوده اند در مقالات مستخرج از رساله رعایت می‌گردد.**
- **در کلیه مراحل انجام این رساله، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.**
- **در کلیه مراحل انجام این رساله، 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 رایانه‌ای، نرم افزارها و تجهیزات ساخته شده) 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 رساله بدون ذکر مرجع مجاز نمی باشد.**

چکیده

امروزه جریان‌های‌کاری در آزمایشات علمی و فرآیندهای سازمانی به طور گسترده‌ای استفاده می‌شوند. جریان‌های‌کاری به سرعت جای خود را در بخش عظیمی از فرآیندهای پیچیده صنعتی پیدا کرده‌اند، و همین امر سبب توجه ویژه به استفاده از روش‌های مناسب برای هرچه بهینه‌تر کردن آن‌ها شده است. در این باره تاکنون کارهایی جهت ساده‌سازی جریان‌های‌کاری صورت گرفته‌است. اما روش‌هایی که تاکنون ارائه شده‌اند، انعطاف پذیری لازم را برای پیش‌بینی ساده‌سازی مورد نیاز جریان‌کاری ندارند. زیرا بدون توجه به پویایی و تغییرات جریان‌های‌کاری ، آن‌ها را به صورت ایستا ساده می‌کنند. در حالی که لازم است تغییرات جریان‌های‌کاری اندازه گیری شود، و بر اساس به روزرسانی اندازه گیری‌ها برای ساده‌سازی آن‌ها تصمیم گیری شود. در روش پیشنهادی، ساده‌سازی جریان‌کاری با استفاده از تحلیل آماری پارامترهای نمودار جریان‌کاری و گزارش‌های عملکرد فرآیند، صورت می‌گیرد. در این روش از تحلیل آماری برای دسته‌بندی جریان‌های‌کاری استفاده می‌شود. به این صورت که ابتدا تگ‌های جریان‌کاری را برای یافتن تگ‌هایی که بیشترین کاربرد را در جریان‌های‌کاری دارند استخراج می‌کنیم. این کار با ارزیابی شباهت صورت می‌گیرد. بدین ترتیب که روش ارائه شده با یافتن تگ‌های پرکاربرد از میان تعداد زیادی جریان‌کاری، مدل یادگیری را برای شناخت دامنه-های کاربرد جریان کاری آموزش می‌دهد. براین اساس مدل یادگیر می‌تواند به خوشه‌بندی جریان‌های‌کاری بپردازد. به ازای هرخوشه یک جریان کاری به عنوان نماینده خوشه معرفی می‌شود. در این جریان‌کاری نماینده، پرتکرارترین تگ-ها احتمال حضور بیش‌تری دارند. نتایج به دست آمده از روش ارائه شده نشان می‌دهد که استفاده از تحلیل آماری برای تخمین مسیرهای ساده‌تر در جریان‌کاری موفق بوده است. در صورت استفاده از روش ارائه شده، مقدار پارامتر اتصال جریان کاری به مقدار ۰,۱۸۶ کاهش خواهد یافت. در حالی که بدون استفاده از این روش این مقدار ۰,۵۲۷ می‌باشد. کاهش پارامتر اتصال به میزان حدود ۳۰ درصد، باعث کاهش برخوردهای خطازا در جریان‌کاری می‌شود. مزیت روش ارائه شده این است که این روش به دلیل نیاز نداشتن به اتکا به محدوده خاصی از جریان‌های‌کاری می‌تواند برای ساده‌سازی و بهبود جریان‌های‌کاری مشاهده نشده نیز تصمیم‌گیری‌های قابل قبولی ارائه دهد. به این صورت که در تشکیل پایگاه‌داده که از روی ارزیابی شباهت‌های جریان‌کاری به وجود آمده است، می‌توان داده‌های این جریان‌کاری جدید و مسیرهای جدید پیش‌بینی شده را در پایگاه داده به صورت پویا اضافه کرد، تا امکان ساده‌سازی خودکار جریان‌کاری فراهم شود. هر زمان که با جریان‌کاری جدیدی مواجه شدیم که قبلا مشابه آن دیده نشده باشد، آن را به نماینده خوشه‌ای که بیشترین شباهت را با آن داشته باشد انتساب داده می‌دهیم. بنابراین امکان پیش‌بینی مسیرهای ساده‌سازی برای جریان‌های‌کاری که تاکنون ندیده‌ایم نیز به وجود می‌آید. تخمین داده‌ها به طور میانگین تا ۸۷ درصد منجر به تخمین صحیح دامنه جریان‌کاری می‌شود.

واژه های کلیدی

تحلیل آماری فرآیند، مدیریت جریان‌های‌کاری ، استخراج داده‌های مفهومی، Big Data.

مقالات مستخرج از رساله

[۱] کوهی‌وردهکردی، طاهره؛ زاهدی، مرتضی. تحلیل SWOT نرم افزار «نمآ» برای بهینه سازی طراحی چند-وجهی، هفتمین کنفرانس بین‌المللی مدیریت دانش و اطلاعات، ۱۳۹۴.

[2] T. Koohi, M. Zahedi, Linear Merging Reduction: A Workflow Diagram Simplification Method, 8th International Conference on Information and Knowledge Technology, 2016.

[3] T. Koohi, and M. Zahedi, Scientific Workflow Clustering based On Motif Discovery, International Journal on Computer Science, Engineering and Information Technology (IJCSEIT), 7 (4), 2017.

[4] T. Koohi-Var, and M. Zahedi, Cross-domain graph based similarity measurement of workflows, Journal of Big Data, 5(1):18, 2018.

[5] T. Koohi-Var, and M. Zahedi, A Cloud based Architecture for Working on BIG DATA with Workflow Management, International Journal of Web & Semantic Technology (IJWesT), 9 (2), 2018.

[6] T. Koohi-Var, and M. Zahedi, Cross-Domain Similarity Assessment for Workflow Improvement to Handle Big Data Challenge in Workflow Management, Journal of Big Data, 5(1):26, 2018.

فهرست مطالب

چکیده ..... و

فهرست جداول ..... ک

فهرست شکل‌ها ..... ل

فصل اول: کلیات ۱

۱-۱. مقدمه ۲

۱-۲. تاریخچه توسعه جریان‌های کاری ۴

۱-۳. اهمیت و ضرورت پژوهش ۵

۱-۴. اهداف پژوهش ۶

۱-۵. همبخشی‌ها و نوآوری اصلی ۷

۱-۶. اختصارات رساله ۸

۱-۷. ساختار رساله ۸

فصل دوم: مفاهیم و اصطلاحات ۹

۲-۱. مقدمه ۱۰

۲-۲. فرآیند سازمانی ۱۰

۲-۳. جریان کاری ۱۱

۲-۳-۱. مورد اجرایی (case-trace) ۱۱

۲-۳-۲. وظیفه (task) ۱۱

۲-۳-۳. منبع ۱۲

۲-۳-۴. موتیف ۱۲

۲-۳-۵. روشهای نمایش جریان کاری ۱۲

۲-۴. سامانه‌های مدیریت فرآیندهای سازمانی (BPMSs) ۱۳

۲-۴-۱. سامانه‌های مدیریت جریان کاری ۱۳

۲-۴-۲. سامانه‌های برنامه‌ریزی منبع سازمانی (ERP) ۱۵

۲-۴-۳. سامانه‌های مدیریت ارتباطهای مشتری (CRM) ۱۵

۲-۵. پیچیدگی در جریان‌های کاری ۱۵

۲-۵-۱. پیچیدگی ساختاری ۱۶
۲-۵-۲. پیچیدگی مفهومی ۱۷
۲-۶. ساده‌سازی ۱۸
۲-۶-۱. ساده‌سازی ساختاری ۱۹
۲-۶-۲. ساده‌سازی مفهومی ۲۱
۲-۷. آنالیز آماری ۲۲
۲-۷-۱. تابع وزن PMI ۲۲
۲-۷-۲. متریک آماری انتساب بهینه زیر الگو (OSPA) ۲۳
۲-۸. روش‌های یادگیری ۲۳
۲-۸-۱. روش یادگیری آماری LSA ۲۳
۲-۸-۲. روش یادگیری آماری (SCL) ۲۴
۲-۸-۳. روش یادگیری انتقالی (روش آماری SFA) ۲۴
۲-۸-۴. روش یادگیری نظریه طیفی گراف ۲۵
۲-۸-۵. روش یادگیری عمیق (DNN) ۲۵
۲-۸-۶. متریک ارزیابی روش یادگیری (CA) ۲۶
۲-۹. متریک‌های ارزیابی جریان‌کاری ۲۶
۲-۹-۱ اتصالات (cp) ۲۶
۲-۹-۲ چگالی ۲۶

فصل سوم: کارهای انجام شده قبلی ۲۷

۳-۱. مقدمه ۲۸
۳-۲. رویکردهای برخورد با افزایش پیچیدگی جریانکاری ۲۹
۳-۳. مباحث کارهای مرتبط ۳۵

فصل چهارم: روش پیشنهادی ۴۵

۴-۱. مقدمه ۴۶
۴-۲. محدودیت‌های بهینه سازی ۴۷
۴-۳. بررسی اجمالی روش ۴۸
۴-۴. اعمال الگوریتم SFA در جریان‌های‌کاری ۵۱
۴-۴-۱. تحلیل جریانهای‌کاری در سطح موتیف ۵۱
۴-۴-۲. تنظیم پارامترهای مساله برای اعمال SFA ۵۵
۴-۴-۳. ارزیابی شباهت در جریانهای‌کاری با یادگیری انتقالی ۵۶
۴-۴-۴. توضیحات الگوریتمیک SFA ۶۰
۴-۵. اعمال الگوریتم یادگیری عمیق با DNN ۶۳
۴-۵-۱. تابع هزینه در یادگیری عمیق ۶۴

۴-۶. راهکار به تغییر پویای جریان‌کاری ۶۵

فصل پنجم: نتایج ۶۷

۵-۱. مقدمه ۶۸

۵-۲. مجموعه داده ۶۸

۵-۳. نتایج و تحلیل ۶۹

فصل ششم: نتیجه‌گیری و پیشنهادات ۷۷

۶-۱. نتیجه‌گیری ۷۸

۷- منابع ۷۹

فهرست جداول

جدول ۱-۱ : ویژگی‌های جریان‌کاری خوب ۳

جدول ۱-۲ : خلاصه برخی اختصارات رساله دسته‌های نقض جریان‌های‌کاری [12] ۸

جدول ۳-۱: دسته‌های نقض جریان‌های‌کاری [12] ۳۲

جدول ۳-۲ : بکارگیری ارزیابی شباهت ۳۸

جدول ۳-۳ : متریک‌های ساده که در [31] تعریف شده اند ۴۲

جدول ۵-۱ : اندازه‌های تعلق دامنه متقابل دامنه‌های مختلف جریان‌کاری ۶۹

جدول ۵-۲ : تحلیل بهبود جریان‌کاری ۷۳

جدول ۵-۳ : مقایسه زمان اجرای روش ارائه شده ۷۴

جدول ۵-۴ : خلاصه خصوصیات روش‌های مختلف ۷۵

فهرست شکل‌ها

شکل ۲-۱: رابطه بین اصطلاحات وظیفه، مورد اجرایی، نمونه اجرایی فرآیند و فرآیند [13] ۱۲

شکل ۲-۲: مثال جریان‌کاری در گالاکسی ۱۴

شکل ۲-۳: مثالی از بصری سازی جریان‌کاری در تاورنا [14] ۱۴

شکل ۲-۴: مثالی از نمایش تودرتویی سامانه بزرگ [14] ۱۶

شکل ۲-۵: نمونه جریان‌کاری پیچیده.. ..... ۱۷

شکل ۲-۶: نمودار جریان‌کاری طراحی شده توسط کاربر ۱۷

شکل ۲-۷: نمودار جریان‌کاری تکامل یافته توسط سامانه ۱۸

شکل ۲-۸: چرخه مدیریت فرآیندهای جریان‌کاری ۱۸

شکل ۲-۹: جریان‌کاری پیچیده.. ..... ۲۰

شکل ۲-۱۰: مراحل ساده‌سازی نمودار جریان‌کاری پیچیده ۲۱

شکل ۲-۱۱: مدل طراحی جریان‌کاری ۲۲

شکل ۳-۱- تکنیک ها به تعمیر BPMN با استفاده از رخدادهای ثبت شده [37] ۳۴

شکل ۴-۱–ساده‌سازی جریان‌کاری. الف- نمودار جریان‌کاری. ب- نمودار ساده شده شکل الف. ۴۶

شکل ۴-۲– مراحل ساده‌سازی جریان‌کاری ۴۷

شکل ۴-۳- روند کاری ساده‌سازی جریان‌کاری ۴۸

شکل ۴-۴-روش اعمال شده برای بهبود جریان‌های‌کاری ۴۹

شکل ۴-۵- خوشه‌بندی ۵۲

شکل ۴-۶- تست نمونه ۵۳

شکل ۴-۷- خوشه‌بندی فرآیندها براساس موتیف‌های مشترک ۵۴

شکل ۴-۸- ارزیابی شباهت ۵۸

شکل ۴-۹- تشکیل نماینده خوشه‌ها براساس Eigen Value (ماتریس وزن دار U) ۵۸

شکل ۴-۱۰- مثال دامنه مبدا و دامنه مقصد ۶۲

شکل ۴-۱۱- پویایی ساخت پایگاه‌داده ۶۵

شکل ۵-۱- مقادیر ماتریس U، k=2 ..... ۷۱

شکل ۵-۲- فاصله مبتنی بر OSPA ۷۲

شکل ۵-۳- زمان پردازش روش ارائه شده برای اندازه مجموعه‌داده‌های مختلف ۷۴

فصل اول:

کلیات

۱ ـ۱. مقدمه

فرآیند واژه‌ای عام است که با هر فعالیتی که سلسله مراتبی داشته باشد سروکار دارد. جریان‌های‌کاری ابزارهایی هستند که می‌توانند به توصیف فرآیندها بپردازند. توصیف فرآیندها فقط یکی از کاربردهای جریان‌های‌کاری است. استفاده از جریان‌های‌کاری برای شفاف سازی فرآیندها و نحوه عملکرد عناصر انسانی و نیز مدیریت منابع سازمانی خیلی زود جایگاه ثابتی در جهان امروز پیدا کرده‌است. جریان‌های‌کاری بسته به دامنه‌ها و حوزه کاربرد آن‌ها انواع مختلفی دارند. قدم‌های ابتدایی به منظور یافتن راه‌حل‌هایی برای بهبود جریان‌های‌کاری با به وجود آمدن نوع جدیدی از جریان‌های‌کاری به نام «جریان‌های‌کاری علمی» برداشته شد. جریان‌های‌کاری علمی به کنترل و توصیف فرآیندهای مرتبط با آزمایشات علمی می‌پردازند. به سبب داده های بزرگ و پیچیدگی های مراحل پیاده‌سازی آزمایشات علمی، این نوع از جریان‌های‌کاری بسیار پیچیده هستند و ساده‌سازی آن‌ها می‌تواند به بهبود فرآیند علمی و کاهش هزینه‌ها منجر شود.

امروزه جریان‌های‌کاری حوزه‌ها و دامنه‌ها کاربردی بیشتری را در برمی‌گیرد. از جمله جریان-های‌کاری در حوزه دولت‌ها که به توسعه و مدیریت پروژه‌های کلان کشورها می‌پردازند، می‌توان برای مثال به پروژه‌های مدیریت ساختارها شبکه‌های اتصال قدرت و تاسیسات و پروژه‌ها کلان هوافضایی و پروژه شبکه‌ها ارتباطی اشاره کرد. با ظهور جریان‌های‌کاری دولتی و صرف هزینه‌های کلان، اهمیت مدیریت صحیح جریان‌های‌کاری افزایش می‌یابد. کاربرد جریان‌های‌کاری منجر به ساخته شدن ابزارهای اختصاصی برای مدیریت و کنترل بهینه آن‌ها در حوزه‌های مختلف شده‌است. هم‌چنین تلاش‌ها به سمت خودکارسازی این ابزارها برای مدیریت بهتر منابع رو به افزایش است.

در حال حاضر اغلب فرآیندها با استفاده از تکنیک‌ها و ابزارهای مدیریت فرآیندهای سازمانی (BPM[1]) کنترل می‌شوند. سازمان‌ها با کمک سامانه‌های BPM می‌توانند به کنترل و ذخیره‌سازی جریان داده‌ها بپردازند. این داده‌ها می‌توانند به شکل XML[2] و یا هر پایگاه‌داده دیگری باشند. یکی از مهم‌ترین بخش‌های سامانه‌های BPM بخش مدیریت جریان‌کاری (WM[3]) می‌باشد و بسیاری از توسعه دهندگان این سامانه‌ها با تقاضای روز افزون برای افزایش کیفیت جریان‌های‌کاری مواجه هستند. چراکه محدودیت و افزایش داده‌ها حجم زیادی از پردازش‌ها را به وجود می‌آورد و زمان زیادی برای هر گونه خطایی در مسیرهای جریان‌کاری صرف خواهد شد. تکنیک‌هایی برای ساخت بهینه

[1] - Business Process Management
[2] - eXtended Markup Language
[3] - Workflow Management

جریان‌های‌کاری ارائه شده‌اند که اغلب به دنبال کسب مشخصه‌هایی برای یک جریان‌کاری بهینه هستند. مندلینگ و همکاران [30] برخی از مشخصه‌های یک جریان‌کاری خوب را در جدول ۱-۱ ارائه کرده‌اند.

جدول ۱-۱. ویژگی‌های جریان‌کاری خوب

جریان‌کاری خوب [30]
قطعات کمتر
اتصالات کمینه
دارای رخداد آغاز و پایان
ساختاریافته
کمترین استفاده از عنصر «یا»
دارای برچسب فعالیت فعل-شی
خرد کردن مدل دارای بیش از ۵۰ عنصر

یکی از مسائل مهم در مدیریت و ساده‌سازی جریان‌های‌کاری به دست آوردن مدلی برای معرفی جریان‌کاری بهینه است. در این زمینه می‌توان از روش‌های مختلف مدل‌سازی بهره برد. اما با توجه به این‌که جریان‌های‌کاری پویا و نیازمند اعمال تغییرات گاه به گاه هستند، برای بهینه سازی آن‌ها پردازش داده‌های بزرگ تحمیل می‌شود. لذا استفاده از روش‌های استخراج داده‌های مهم به بهینه-سازی کمک می‌کند.

از میان روش‌هایی که تاکنون برای بهینه کردن و ساده‌سازی جریان‌های‌کاری ارائه شده‌اند، برخی از کارها به کمک استخراج خصوصیات آماری داده‌های شفاف مبتنی بر بردار و برخی با روش‌های دیگر استخراج خصوصیات آماری داده‌های غیرشفاف به بررسی جریان‌های‌کاری در سطوح مختلف پرداخته-اند. اما روش‌هایی که تاکنون ارائه شده‌اند انعطاف‌پذیری لازم را برای اعمال ساده‌سازی مورد نیاز به انواع مدل فرآیندها ندارند. روش‌هایی که تاکنون برای بهبود ارائه شده‌اند اغلب مبتنی بر استفاده مجدد از جریان‌های‌کاری با در نظر گرفتن شباهت‌هایی هستند که محدود به یک دامنه (مثلا سازمانی و یا علمی) می‌شوند. روش‌های مختلفی برای مدل‌سازی جریان‌های‌کاری ارائه شده است. از جمله این‌که یک جریان‌کاری را می‌توان با BPMN[1]، گراف وابستگی[2]، شبکه پتری و غیره مدل کرد. هر

[1] - Business Process Management Network
[2] - Dependency Graph

کدام از این مدل‌سازی‌ها در جای خود مزایای خود را دارند. به عنوان مثال اگر گراف وابستگی جریان-کاری را به عنوان مدل‌سازی انتخاب کرده‌ایم می‌توان از روش‌های ساده‌سازی گراف بهره برد. ولی در صورتی‌که بتوان مفهوم را به نحوی در آن دخیل کرد. به عنوان مثال می‌توان جریان‌کاری را با شبکه گالویس[1] یا شبکه مفهومی نشان داد. پس از در نظر گرفتن مفاهیم در نمایش گراف به عنوان محدودیت‌های ساده‌سازی می‌توان از روش‌هایی کمک گرفت که برای ساده‌سازی گراف ارائه شده‌اند. روش‌های متعددی برای ساده‌سازی گراف ارائه شده‌اند از جمله روش‌های ساده‌سازی گراف به روش کاهش گراف [2] و [4] یا روشی که خانم چوبدار در رساله‌اش معرفی کرده است [52]. مدل ورودی رساله خانم چوبدار گراف شبکه است. در رساله حاضر نیز جریان‌های‌کاری با گراف شبکه مدل شده‌اند. گراف شبکه امکان استخراج ترتیب قطعات جریان‌کاری را فراهم می‌کند. هم‌چنین مرسوم‌ترین روش استخراج مدل جریان‌کاری از گزارش ثبت شده استفاده از گراف شبکه می‌باشد. به طور مثال آقایان یعقوبی و زاهدی در پژوهشی که برای تخصیص منابع با استفاده از فاصله شباهت وظیفه صورت دادند [90] برای مدل‌سازی فرآیندهای کسب‌وکار از گراف شبکه بهره برده‌اند. هر یال موجود در گراف شبکه را معادل ارتباط میان قطعات جریان‌کاری درنظر گرفته‌ایم. هر گره در گراف شبکه را معادل یک قطعه در جریان‌کاری درنظر گرفته‌ایم.

۲.۱ تاریخچه توسعه جریان‌های‌کاری

واژه جریان‌های‌کاری در طی دوره‌هایی از توسعه، کاربردهای مختلفی را در مهندسی پیدا کرده است. قدم ابتدایی در معرفی جریان‌های‌کاری به صورت مفهوم گردش کار و کنترل یک فرآیند توسط تیلور و گانت تا حدودی برداشته شد. تیلور در کتاب خود با نام اصول مدیریت علمی که در سال ۱۹۱۱ چاپ شد به این مفهوم برنامه‌ریزی برای گردش کار اشاره کرده است [88]. گانت نیز در کتاب خود با نام سازمان‌دهی برای کار به اهمیت نمود فرآیندهای کاری در ساختارهای چارت مانند پرداخته است [89]. تیلور و گانت هر دو از دانشمندانی هستند که در ابتدای عصر ظهور جنبش کارآیی به منظم کردن و در نظر گرفتن سلسله مراتب در کارها پرداخته‌اند. اما در دوران حیات آن دو مفهوم جریان‌های‌کاری صرفا در علوم مدیریت مطرح می‌شد. پس از آن در عصر ظهور سیستم‌های پیچیده مفهوم جریان‌های‌کاری در مدیریت جریان اطلاعات فیزیکی سیستم‌های پیچیده استفاده شد. طی

[1] - Galois Lattice

دوران جنگ دوم جهانی آزمایش‌های فضایی و پروژه های سنگین علمی و مهندسی به وجود آمدند که هرکدام حجم بسیاری از اطلاعات و برنامه‌ریزی‌ها در بخش‌های مختلف مالی، علمی و زمان‌بندی دربر داشتند. در نتیجه بتدریج استفاده از جریان‌های‌کاری برای توصیف فرآیندهای پیچیده صنعتی به بخش جدایی ناپذیر از هر پروژه بزرگ تبدیل شد.

پس از دوران جنگ جهانی حدود سال ۱۹۸۰ میلادی در عصر ظهور کیفیت، مشخصه‌های جریان-های‌کاری وارد علم بازمهندسی شد، تا بتوان جریان‌های‌کاری با کیفیت بالاتر برای مدیریت فرآیندها به دست آورد. در همین دوره جریان‌های‌کاری برای فرآیندهای سازمانی نیز به کار گرفته شد. توسعه سیستم‌های مدیریت فرآیندهای کاری به همین دوره برمی‌گردد.

۱ ۳. اهمیت و ضرورت پژوهش

جریان‌های‌کاری در ارتباط با رفتارهای انسانی هستند. رفتارهای انسانی وابسته به عوامل زیادی از جمله تغییرات تکنولوژی، محیط و منابع دائم در حال تغییر است. بنابراین بروز تغییرات امری جدایی ناپذیر از جریان‌های‌کاری می‌باشد. حتی جریان‌های‌کاری که با بررسی‌های علمی و تجربی فراوان توسط افراد خبره تهیه می‌شوند، پس از مدتی چندان مناسب به نظر نمی‌رسند و کاربران به طور دائم ایده‌هایی برای کاراتر شدن آن‌ها در ذهن می‌آورند. در این حالت کنار گذاشتن چنین جریان‌کاری و ساخت جریان‌کاری جدید بدون در نظر گرفتن تجربیات قبلی کاری مقرون به صرفه نیست. در خیلی از مواقع تغییرات جزئی می‌تواند باعث افزایش رضایت‌مندی کاربران شود. گاهی این تغییرات جزئی، خود باعث ایجاد ساختاری ناهمگون در جریان‌کاری می‌شوند. چگونگی اعمال این تغییرات جزئی در جریان‌کاری بسیار حائز اهمیت است. یکی از مهم‌ترین نیازهای سامانه‌های مرتبط با جریان‌های‌کاری این است که این تغییرات به طور کارا در جریان‌کاری اعمال شود. امروزه اهمیت خودکارسازی سامانه-ها برکسی پوشیده نیست. بسیاری از تکنولوژی‌های ارائه شده نظیر اینترنت اشیا[1] و بسترهای ابری در سامانه‌های خودکار مدیریت بهتری خواهند داشت. تحقیقاتی که به خودکار سازی جریان‌های‌کاری می‌پردازند در همین راستا انجام می‌شوند. خودکارسازی جریان‌های‌کاری به نحوی که جریان‌کاری قابلیت خودترمیمی پیدا کند از وظایف مهم توسعه‌دهندگان می‌باشد. ساده‌سازی جریان‌های‌کاری

[1] - Internet of Things

مهم‌ترین بخش سامانه‌های خودکارسازی می‌باشد. با ساده‌سازی جریان‌های‌کاری امکان استفاده مجدد از ساخته‌ها قبلی به وجود می‌آید. ساختن مدل‌های مختلف از جریان‌های‌کاری و مقایسه داده‌های بدست آمده از این مدل‌ها با داده‌های مشاهده شده از جریان‌های‌کاری جدید یکی از روش‌های مناسب تحلیل و تفسیر جریان‌های‌کاری است. روش ارائه شده در این رساله یک روش خودکار برای تفسیر جریان‌های‌کاری می‌باشد. یکی از ویژگی‌های مهم این روش محدود نبودن آن به جریان‌های‌کاری دیده شده است، به گونه‌ای که با این روش می‌توان به تحلیل و ساده‌سازی جریان‌های‌کاری دیده نشده پرداخت.

برخی از مهم‌ترین کاربردهای ساده‌سازی جریان‌های‌کاری در موارد زیر دیده می‌شود:

۱ -کنترل و مدیریت بحران و خطایابی
۲ -مستندسازی و شفاف سازی در مدیریت دانش
۳ -کمک به زمان‌بندی در ساختارها
۴ -فشرده‌سازی

استفاده از جریان‌های‌کاری بهینه شده همواره باعث کاهش هزینه توسعه و ساخت، و خطایابی آزمایشات فرآیندهای پیچیده و هم‌چنین فرآیندهای شهری می‌شود.

۱ ۴. اهداف پژوهش

هدف اصلی پژوهش حاضر معرفی روشی مبتنی بر تحلیل آماری برای ساده‌سازی جریان‌کاری بدون اتکا به جریان‌کاری خاص می‌باشد. بدین مفهوم که به نحوی فرآیند ساده شود و شفاف‌سازی عملکرد صورت بگیرد که مفهوم کلی جریان‌کاری حفظ شود. اهداف فرعی این پژوهش عبارتند از افزایش خوانایی نمودارهای جریان‌کاری، افزایش قابلیت فهم جریان‌کاری، بهبود روند جریان‌کاری. بدین منظور در این نوشتار قصد داریم راه‌حلی با توجه به خصوصیات جریان‌های‌کاری برای کاهش پیچیدگی و افزایش بهینگی آن‌ها ارائه کنیم. بدین منظور به بررسی و تحلیل آماری وضعیت استفاده جریان‌های‌کاری داده شده پرداخته‌ایم. از این تحلیل برای دسته‌بندی جریان‌های‌کاری بهره برده‌ایم. در این صورت کافی است با استفاده از تحلیل آماری گزارش‌های عملکرد یک یا چند فرآیند، برای فرایند مشابه ساده‌سازی ترتیب دهیم. اما با توجه به این که جریان‌های‌کاری از مجموعه ای از زیرجریان‌ها تشکیل شده‌اند، توصیف جریان‌کاری را مبتنی بر مجموعه‌های متناهی قرار داده‌ایم. بدین ترتیب مشکلاتی را نخواهیم داشت که در پی مقایسه برداری جریان‌های‌کاری مثل روش ارائه شده در

[41] با آن مواجه می‌شوند. در مقاله [41] با توصیف برداری جریان‌های‌کاری ماتریس‌های اسپارس[1] حاصل شده است و در نتیجه کیفیت دسته‌بندی جریان‌های‌کاری پایین آمده است. این در صورتی است که اگر از توصیف‌گرهای مبتنی بر مجموعه استفاده کنیم می‌توان اندازه جریان‌های‌کاری را متغیر در نظر گرفت. ضمن این که می‌توان از تئوری مبتنی بر مقایسه مجموعه‌ها برای ارزیابی و تحلیل آن‌ها بهره برد [46]. بدین ترتیب دیگر فرض‌های محاسبات در فضای برداری را نخواهیم داشت.

در روش پیشنهادی این رساله، ساده‌سازی جریان‌های‌کاری به کمک پیاده‌سازی روش‌های یادگیری انتقالی و یادگیری عمیق صورت می‌گیرد. در حین روش یادگیری انتقالی استخراج داده‌ها، خوشه-بندی، و تحلیل آماری جریان‌های‌کاری انجام می‌شود. پس از آن که خوشه‌بندی جریان‌های‌کاری انجام شد، از روش یادگیری عمیق بهره می‌بریم تا از دانش قبلی جریان‌های‌کاری که قبلا در خوشه-های نمونه تهیه شده‌اند، پیشنهاداتی برای ترتیب چینش بهتر قطعات ارائه دهیم. در حین پیاده‌سازی، جریان‌های‌کاری از گزارش عملکرد برداشته می‌شوند. زیرا معمولا سازمان‌ها گزارش‌های عملکرد جریان‌های‌کاری را نیز ثبت می‌کنند. هم‌چنین می‌توان داده‌های جریان‌های‌کاری را از نمودار مدل شده آن اخذ کرد. بررسی روی هر جریان‌کاری ورودی پس از تحلیل آماری و خوشه‌بندی متناسب با دامنه‌ها و حوزه‌های کاربرد مشابه آن صورت می‌گیرد. بنابراین امکان پیش‌بینی مسیرهای ساده‌سازی مشابه برای جریان‌های‌کاری که تاکنون ندیده‌ایم نیز به وجود می‌آید. در نتیجه ساده‌سازی بدون اتکا به حوزه و دامنه جریان‌کاری خاص می‌شود.

۱.۵. هم‌بخشی‌ها و نوآوری اصلی

تا آن‌جا که می‌دانیم این رساله برای اولین بار به بررسی مفهومی جریان‌های‌کاری با یادگیری انتقالی می‌پردازد. برای این کار این رساله روش ارزیابی شباهت SFA[2] را می‌پذیرد. تمرکز این رساله بر این است که با در نظر گرفتن حوزه‌ها و دامنه‌های کاربرد جریان‌های‌کاری، جریان‌های‌کاری ساده‌تر ارائه دهد. نوآوری اصلی این رساله این است که روشی یادگیر برای ساده‌سازی جریان‌های‌کاری ارائه دادیم که بدون اتکا به نوع حوزه و دامنه‌های کاربرد جریان‌کاری ورودی می‌باشد. پویایی تولید پایگاه داده روش یادگیری ارائه شده باعث می‌شود امکان اعمال روش در حین شکل‌گیری جریان‌کاری وجود داشته باشد، و حتی ساده‌سازی جریان‌کاری نادیده داشته باشیم.

[1] - Sparse
[2] - Spectral Feature Alignement

۱ ۶. اختصارات رساله

جدول ۱-۲ اختصارات استفاده شده در رساله را نشان می‌دهد.

جدول ۱-۲. خلاصه برخی اختصارات رساله

توصیف	اختصار
جریان‌کاری	Wf[1]
مجموعه‌های ماژول	MS[2]
مجموعه‌های مسیر	PS[3]
فاصله ویرایش گراف	GED[4]
کیسه کلمات	BW[5]
کیسه تگها	BT[6]
برچسب موتیف	ML[7]
نگاشت موتیف	MM[8]
طول توصیف کمینه[56]	MDL
تحلیل معنایی پنهان	LSA[9]
انتساب خصیصه طیفی	SFA
شبکه‌ها عصبی عمیق	DNN[10]

۱ ۷. ساختار رساله

این رساله مشتمل بر ۶ فصل می‌باشد. فصل اول راجع به کلیاتی درباره جریان‌های‌کاری ضمن ارائه سابقه مختصری از آن‌ها و توضیحاتی درباره مدل‌سازی و استخراج خصیصه و روش ارائه شده می‌باشد. هم‌چنین در این فصل به ضرورت انجام تحقیق و اهداف آن پرداخته شده‌است. فصل ۲ به بیان مفاهیم و اصطلاحات اشاره دارد. سپس در فصل ۳ مروری بر پژوهش‌های مرتبط با موضوع رساله صورت گرفته‌است. روش پیشنهادی در فصل ۴ بیان شده است. فصل ۵ شامل تحلیل و مقایسه نتایج می‌باشد. در نهایت جمع‌بندی و نتیجه‌گیری در فصل ۶ صورت می‌گیرد.

[1] - Workflow
[2] - Module Sets
[3] - Path Sets
[4] - Graph Edit Distance
[5] - Bag of Words
[6] - Bag of Tags
[7] - Motif Label
[8] - Motif Mapping
[9] - Latent Semantic Analysis
[10] -Deep Neural Networks

فصل دوم:

مفاهیم و اصطلاحات

۲ ـ۱. مقدمه

مفاهیم و مبانی مرتبط با داده‌های متنی و نیز شیوه‌های نمایش جریان‌های‌کاری در این پژوهش نیازمند معرفی هستند. نکته‌ای که در نمایش نمودارهای جریان‌های‌کاری وجود دارد این است که گاهی این نمودارها بزرگ، پیچیده و تودرتو می‌شوند که ارزیابی و بازبینی آن‌ها نیازمند معرفی پارامترهای مشخصی می‌باشد. مسیر یک جریان‌کاری در دنیای واقعی یک مسیر کاملا یک جهته خطی نیست. گاهی ممکن است پیوندهای قطعات این جریان‌کاری وجه‌های هندسی زیادی به این جریان‌کاری اضافه کند. از سوی دیگر مکان قرارگیری قطعات نیز می‌تواند باعث افزایش پیچیدگی مسیر عملکردی جریان‌کاری شود. پیچیدگی چنین جریان‌های‌کاری که ناشی از افزایش وجه و قرار-گیری قطعات است، و در برخی موارد اسپاگتی مانند معرفی شده‌اند [28]، باعث می‌شود که خوانایی جریان‌کاری کاهش یابد و فرد با نگاه از هر طرف به آن یک برداشت متفاوت داشته باشد. این پیچیدگی از جنبه‌های مختلف قابل معرفی و اندازه‌گیری می‌باشد. هم‌چنین معیارهایی برای سنجش جریان‌های‌کاری معرفی شده‌اند. از آن جمله می‌توان به چگالی و یا تعداد قطعات جریان‌کاری اشاره کرد، که این معیارها برای پیش‌بینی کیفیت جریان‌های‌کاری می‌توانند به کار آیند. در این فصل به بیان این مفاهیم و اصطلاحات مرتبط می‌پردازیم.

۲ ـ۲. فرآیند سازمانی

دسته‌ای خاص از موردهای اجرایی مدل‌های جریان‌کاری می‌باشد. فرآیند نشان می‌دهد که لازم است کدام وظایف انجام شوند. هم‌چنین ترتیبی که باید انجام شود را نیز نشان می‌دهد. می‌توان فرآیند سازمانی را به عنوان زیربرنامه‌ای برای یک نوع مورد اجرایی خاص از جریان‌کاری درنظر گرفت. به‌طور کلی، فرآیندهای زیادی با به کارگیری یک فرآیند پایه می‌توانند انجام شوند.

دوره زندگی یک مورد اجرایی هم ارز با فرآیند تعریف شده است. زیرا هر مورد اجرایی یک دوره زندگی متناهی با یک آغاز و پایان شفاف دارد، و مهم است که فرآیند نیز با این دوره تطابق[1] داشته باشد. بنابراین هر فرآیند نیز یک آغاز و یک پایان دارد، که به ترتیب ظهور[2] و تکمیل یک مورد اجرایی را نشان می‌دهد.

[1] - Conform
[2] - Appearance

۲ ۳. جریان‌کاری

بسیاری از فرآیندها با جریان‌کاری مدل می‌شوند. جریان‌کاری شامل الگوهای تکرارپذیر فرآیندهای سازمانی در شبکه‌ای می‌باشد که براساس کنترل و وابستگی‌ها داده ساختاریافته شده است. بنابراین جریان‌کاری به شیوه‌ای بازگشتی تعریف می‌شود. یعنی یک کلاس جریان‌کاری از یک یا چند کلاس جریان‌کاری مرتبط تشکیل شده است [7].

۲-۳-۱. مورد اجرایی (case-trace)

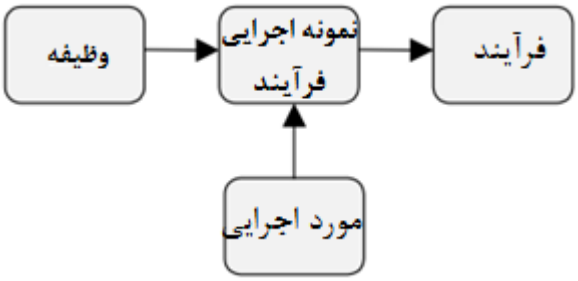
یک اجرا از یک نمونه مدل جریان‌کاری، در حوزه استخراج فرآیند مورد اجرایی نامیده می‌شود و دنباله‌ای از رخدادهاست. ممکن است چند مورد اجرایی از یک مدل جریان‌کاری بخصوص وجود داشته باشد که به طور همزمان اجرا می‌شوند. اما معمولا فرض می‌شود که اجرای مستقل داشته باشند، و معمولا بدون ارجاع به یکدیگر اجرا می‌شوند.

۲-۳-۲. وظیفه (task)

در جریان‌های‌کاری از مفهوم وظیفه به طور گسترده‌ای استفاده می‌شود. یک وظیفه یک واحد منطقی از کار است. اگر هر چیزی در خلال کارآیی یک وظیفه نادرست دربیاید، آن‌گاه باید به آغاز کل وظیفه برگردیم؛ یعنی عقب‌گرد[1] می‌کنیم.

وظیفه به مفهوم عمومی کار اشاره می‌کند و به یک مورد اجرایی خاص اشاره نمی‌کند. یک نمونه اجرایی فرآیند، ترکیبی از مورد اجرایی و وظیفه (وظیفه+مورد اجرایی) است که فقط لازم است انجام شود. یک نمونه اجرایی فرآیند با اجازه یک مورد اجرایی، ایجاد می‌شود. بنابراین می‌شود یک نمونه اجرایی فرآیند را بخشی واقعی از کار دانست که ممکن است انجام شود. اصطلاح فرآیند به کارآیی واقعی یک نمونه اجرایی فرآیند اشاره می‌کند. به محض این که کاری بر نمونه اجرایی فرآیند آغاز شود، آن کار با فرآیند برابر می‌شود. توجه شود که برخلاف وظیفه، هردوی نمونه اجرایی فرآیند و فرآیند با یک مورد اجرایی مشخص به هم متصل شده‌اند. شکل ۲-۱ این نکته را نشان می‌دهد:

[1] - Rollback



شکل ۲-۱- رابطه بین اصطلاحات وظیفه، مورد اجرایی، نمونه اجرایی فرآیند و فرآیند [13]

۲-۳-۳. منبع

اصطلاحی عمومی برای اشاره به هرکس یا هرچیز درگیر در کارایی یک فعالیت فرآیند می‌باشد. منابع در پیشبرد اهداف سازمان و اجرای کارهای تعریف شده در آن سازمان موثر هستند. هر منبع می‌تواند:

- یک شرکت‌کننده در فرآیند، یعنی فردی مثل کارمند باشد.
- یک سامانه نرم‌افزاری باشد.
- تجهیزاتی مانند پرینتر یا کارخانه تولید باشد.

دسته منبع در ارتباط با نقش‌ها و گروه‌ها مشخص می‌شود.

۲-۳-۴. موتیف

به طور کلی به الگویی در جریان‌کاری موتیف گفته می‌شود که به صورت بخشی از جریان‌کاری تعریف می‌شود که بیش‌ترین تکرار را در ساختار جریان‌کاری دارد. هر موتیف ساختاری منحصر به فرد است که در عین ارتباط با سایر بخش‌های جریان‌کاری می‌تواند به صورت مستقل کارآیی داشته باشد.

۲ ۳ ۵. روش‌های نمایش جریان‌کاری

جریان‌های‌کاری را می‌توان به دو شیوه استفاده از گزارش و نمایش شبکه‌ای نشان داد. در حین استفاده از گزارش عملکرد[1]، برای نشان دادن جریان‌های‌کاری می‌توان از روش‌های ثبت گزارش به عنوان مثال مکانیسم ثبت رخداد مبتنی بر XML در فرمت XES [74] استفاده کرد.

در روش نمایش شبکه‌ای می‌توان از توصیف‌گرهای شبکه‌ای استفاده کرد. از جمله این توصیف‌گرها می‌توان به شبکه های پتری و شبکه گالویس اشاره کرد که در ادامه تعریفی برای آن‌ها بیان می‌کنیم.

شبکه‌های پتری: شبکه پتری کلاسیک یک گراف دوبخشی جهت‌دار با دو نوع گره با نام‌های مکان[2] و انتقال[3] است [3]. این گره‌ها از طریق کمان‌های جهت‌دار به هم متصل‌اند. اتصالات بین دو

[1] - Event Log
[2] - Places
[3] - Transitions

گره مشابه اجازه داده نشده است. در شبکه پتری مکان ها با گره‌ها نشـان داده مـی شـوند و انتقـال‌هـا توسط مستطیل‌ها نشان داده می‌شوند.

شبکه گالویس: گرافی است که امکان نمایش همه وابسته‌سازی‌های[1] ممکن بین مجموعه اشیا، یا مشاهدات، و خصیصه‌های توصیف کننده‌شان را می‌دهد.

۲ ۴. سامانه‌های مدیریت فرآیندهای سازمانی (BPMSs[2])

سامانه‌هایی هستند که به پشتیبانی، کنترل و نظارت فرآیندهای سازمانی می‌پردازنـد. سـامانه‌هـای مدیریت فرآیندهای سازمانی گستره وسیعی از ابزارها، تکنیک‌ها و قوانین مـدیریتی را بـا هـدف انجـام بهتر فرآیندها [24] ارائه می‌دهد. ایـن سـامانه‌هـا بـرای ایجـاد و بهبـود طیـف وسـیعی از فرآینـدهای مدیریتی به سازمان کمک می‌کنند. به طور کلی هدف اصلی این سامانه‌ها بهبود فرآینـدهای سـازمانی (BPI[3]) می‌باشد. BPI رهیافتی است برای کمک به یک سازمان تا فرآینـدهای زیرسـاختی‌اش را بـرای نیل به نتایج کارها بهینه کند. به عنوان برخی از این سامانه‌ها می‌توان به سامانه‌های مـدیریت جریـان-کاری (WfMS[4])، سامانه‌های برنامه‌ریزی منبع سازمانی (ERP[5]) و سامانه‌های مدیریت رابطه مشـتری (CRM[6]) اشاره کرد. از این میان سامانه‌های مدیریت جریان‌کـاری کـاربرد بـیش‌تـری دارنـد. زیـرا بـه سازمان‌ها اجازه می‌دهند فرآیندهای سازمانی را خودکار کرده و این فرآینـدهای سـازمانی را مهندسـی مجدد کنند؛ به علاوه این سامانه ها کارایی را افزایش می‌دهند و هزینه‌ها را کاهش می‌دهند.

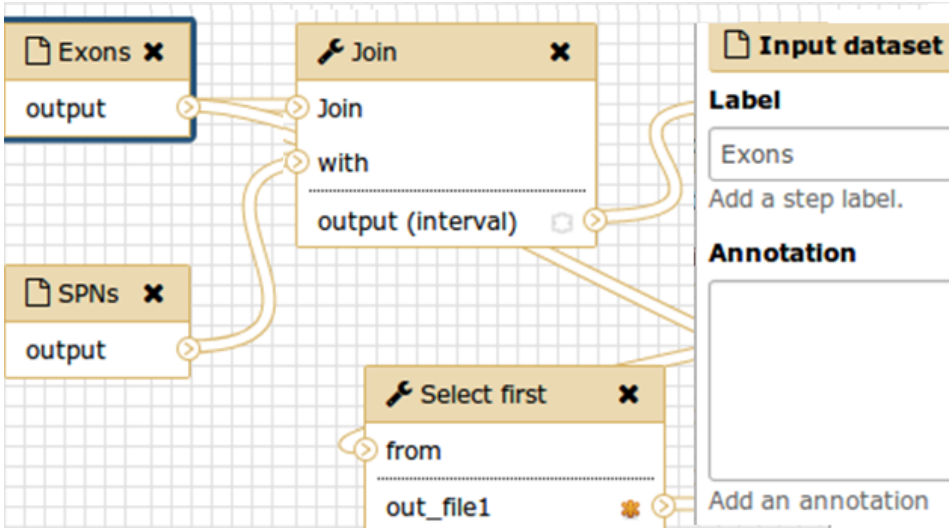
۲-۴-۱. سامانه‌های مدیریت جریان‌کاری

در سامانه‌های مدیریت جریـان‌کـاری، کـاربر مـی‌توانـد بـدون آن کـه درگیـر کدنویسـی شـود بـه ساختاربندی فرآیندها بپردازد. به عبارتی این سامانه‌ها مدلی ساده برای برنامه‌نویسی فراهم مـی‌کننـد، که در آن دنباله‌ای از فرآیندها با اتصال خروجی‌ها از یکی به ورودی دیگری تشکیل شده است.

به عنوان مثالی از سامانه‌هـای مـدیریت جریـان‌کـاری، مـی‌تـوان بـه گالاکسـی[7] [5] اشـاره کـرد. جریان‌کاری نمونه در گالاکسی را در نظر بگیرید (شکل ۲-۲).

---

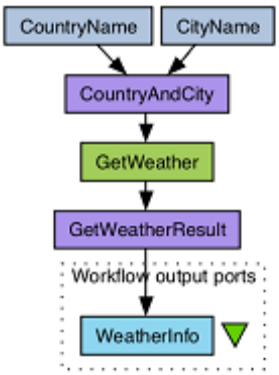
[1] - Associations
[2] - Business Process Management Systems
[3] - Business Process Improvement
[4] - Workflow Management Systems
[5] - Enterprise Resource Planning
[6] - Customer Relationship Management
[7] - Galaxy



شکل ۲-۲- مثال جریان‌کاری در گالاکسی

گراف جریان‌کاری شکل ۲-۲، مثالی از شیوه ساخت جریان‌های‌کاری در گالاکسی را نشان می‌دهد. اتصال‌های گراف گام‌های تحلیل را از ورودی (گوشه سمت چپ-بالا) به خروجی (چپ-پایین) نشان می‌دهد که جریان‌داده‌ها را نشان می‌دهد. فایل‌های داده ورودی و خروجی در بدنه آیکون‌های گره‌های گراف نشان داده شده‌اند. اگر شما خبره‌ای در این جریان‌کاری نباشید، دشوار است که بین عناصر نمایش دهنده گره‌های داده (یعنی ورودی‌ها/ خروجی‌های جریان‌کاری) تمایز قائل شوید.

سامانه مدیریت جریان‌کاری معروف دیگر تاورنا[1]ست [6]. این سامانه ترکیبی از ابزارها برای طراحی و اجرای جریان‌های‌کاری است و اجازه می‌دهد کاربران از خدمات وب بهره ببرند. مثالی از یک جریان‌کاری ساده در شکل ۲-۳ آمده است که پیش بینی آب و هوا را برای یک شهر مشخص بازیابی می‌کند. در تاورنا یک جریان‌کاری به صورت گراف‌هایی ارائه می‌شود که با استفاده از زبان بصری تاورنا ساخته شده باشد. گراف‌ها در آن هم‌چنین جریان داده را از بالا به پایین نشان می‌دهند (شکل ۲-۳).



شکل ۲-۳- مثالی از بصری‌سازی جریان‌کاری در تاورنا [14]

[1] - Taverna

در تاورنا رنگ‌ها برای نشان دادن طبیعت عناصر گراف استفاده می‌شوند. اگرچه الگوی رنگ واقعا محسوس نیست، اما بین عناصر داده و خدمات تاورنا می‌توان به وضوح تمایز قائل شد. در تاورنا، جریان‌های‌کاری نیز می‌توانند به سایر جریان‌های‌کاری تودرتو شوند که جریان‌های‌کاری را قابل اجرای مجدد می‌کند و برای نگهداری ساده‌تر است.

بجز سامانه‌های مدیریت جریان‌کاری به سامانه‌های دیگری در این نوشتار اشاره شده است که در ادامه آن‌ها را تعریف می‌کنیم:

۲-۴-۲. سامانه‌های برنامه‌ریزی منبع سازمانی (ERP[1])

این سامانه‌ها چهارچوب نرم‌افزاری برای مدیریت منابع، درخواست یا سایر نیازمندی‌های یک شرکت را فراهم می‌کند. این سامانه‌ها معمولا براساس مجموعه‌ای از قوانین و ابزارهای از پیش تعیین شده عمل می‌کنند.

۲-۴-۳. سامانه‌های مدیریت ارتباط‌های مشتری (CRM[2])

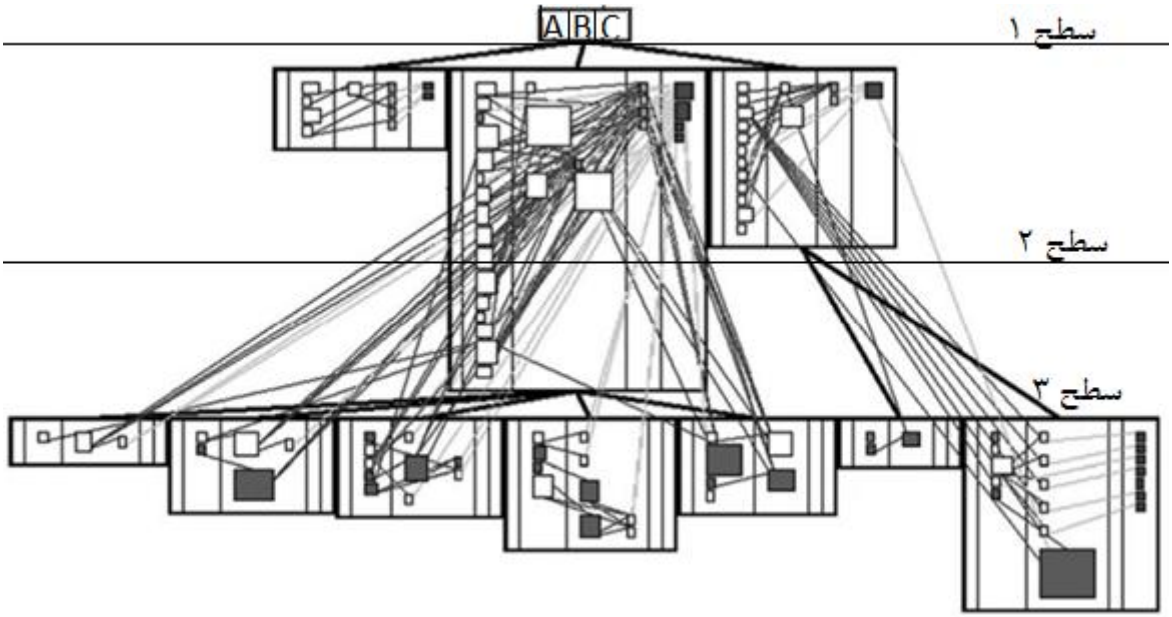
این سامانه‌ها سفارش‌های مشتری و منابع مالی صندوق پول را مدیریت می‌کنند. در این سامانه‌ها معمولا ابزارهایی برای مدیریت نیازمندی‌های مشتری و پیگیری علاقه‌مندی‌های مشتری‌ها وجود دارد. هم‌چنین این سامانه‌ها امکان مدیریت فرد به فرد مشتری‌ها و تنظیم فرآیند خرید آنها را فراهم می‌کند. مهم‌ترین مزیت این سامانه‌ها برای سازمان‌ها امکان تهیه گزارش‌گیری و مستندات با استفاده از ابزارهای تحلیلی می‌باشد.

۲ ۵. پیچیدگی در جریان‌های‌کاری

جریان‌های‌کاری می‌توانند یک فرآیند سازمانی را توصیف کنند. آن‌ها می‌توانند به صورت مدل فرآیند دیده شوند که قطعات در آن جزئی از یک فرآیند را نمایش می‌دهند، و اتصال‌ها جریان داده‌ای بین اجزای فرآیند را به دست می‌گیرند. قطعات امکان نمایش منابع و داده‌ها و سایر اجزای کنترلی یک فرآیند را فراهم می‌کنند. اتصال‌ها نیز امکان نمایش حرکت اسناد و روابط میان اجزا را فراهم می‌کنند. به این صورت مدل فرآیند می‌تواند فرآیندکاری برای یک محصول یا سرویس خاص را نشان دهد. فرآیندهای‌کاری می‌توانند خیلی پیچیده باشند. بنابراین نمایش بصری این که چگونه وظیفه باید تکمیل شود، کمک می‌کند بتوان فرآیند و کار را به صورت کاراتر فهمید. ساده‌سازی در نمایش بصری فرآیندهای‌کاری نیز اثرگذار است. به عنوان مثالی از نمایش تودرتویی سامانه بزرگ می‌توان جریان‌کاری نمایش داده شده در شکل ۲-۴ را نشان داد. در شکل ۲-۴ کاربر می‌تواند خلاصه‌های

[1] - Enterprise Resource Planning
[2] - Customer Relationship Management

تقسیم و انشعاب را بلافاصله ببیند، که به صورت مستطیل‌های بزرگتر با یا بدون رنگ نشان داده شده‌اند. همین روش می‌تواند برای نمایش جریان‌های‌کاری بزرگی اعمال شود که از جریان‌های‌کاری تودر-توی کوچکتر زیادی تشکیل می‌شوند.



شکل ۲-۴- مثالی از نمایش تودرتویی سامانه بزرگ [14]

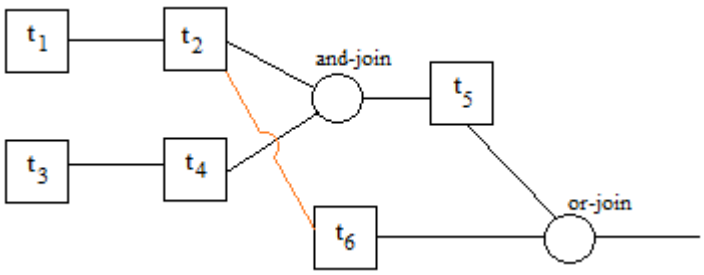
در شکل ۲-۴ در سطح ۱ می‌بینیم که جریان‌کاری کلی از سه بخش A، B و C تشکیل شده‌است. در سطح ۲ مشاهده می‌شود که هرکدام از بخش‌های A، B و C خودشان به نوعی تعدادی جریان‌کاری هستند. در سطح ۳ میزان تودرتویی بیشتر شده و مشاهده می‌شود که هرکدام از بخش‌های سطح ۲ نیز خودشان با هم در ارتباط هستند.

هنگام طراحی نمودار جریان‌کاری، یک رویکرد این است که از ابتدا اجازه طراحی جریان‌های‌کاری دلخواه را ندهیم [10]. رویکرد دیگر این است که اجازه بدهیم جریان‌کاری شکل بگیرد. سپس آن را به جریان‌کاری معقول، خوش‌رفتار و ساختاریافته تبدیل کنیم [23]. در طراحی جریان‌های‌کاری توسط کاربر، افزایش پیچیدگی یک جریان‌کاری از چند جنبه ممکن است به وجود آید:

۲-۵-۱. پیچیدگی ساختاری

جریان‌های‌کاری نزدیک به واقعیت معمولا به شکل جریان‌کاری خطی دیده می‌شوند [48]. در این نوع پیچیدگی بخش‌های پیچیده کننده مثل قطعه مرکب، قطعه بی‌فایده، اتصالات چندلایه و حلقه مطرح هستند. اتصال چند لایه در دنیای واقعی به معنای انجام کارهای موازی است. به عنوان مثال برای دریافت مجوز اکران کنسرت دو راه ممکن است: یک راه این است که از طریق وزارت ارشاد مراحلی را طی کنیم و مجوز بگیریم. راه دیگر این است که از طریق استانداری مراحلی را برای مجوز گرفتن طی کنیم. حال اگر بین یکی از مراحل استانداری و ارشاد اتصالی وجود داشته باشد، نمودار

جریان‌کاری آن دو لایه شده است و پیچیدگی افزایش یافته است. به عنوان مثال در شکل ۲-۵ می‌توان این افزایش پیچیدگی را ملاحظه کرد:

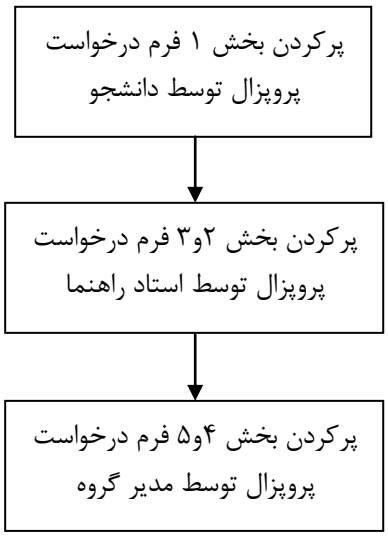


شکل ۲-۵- نمونه جریان‌کاری پیچیده

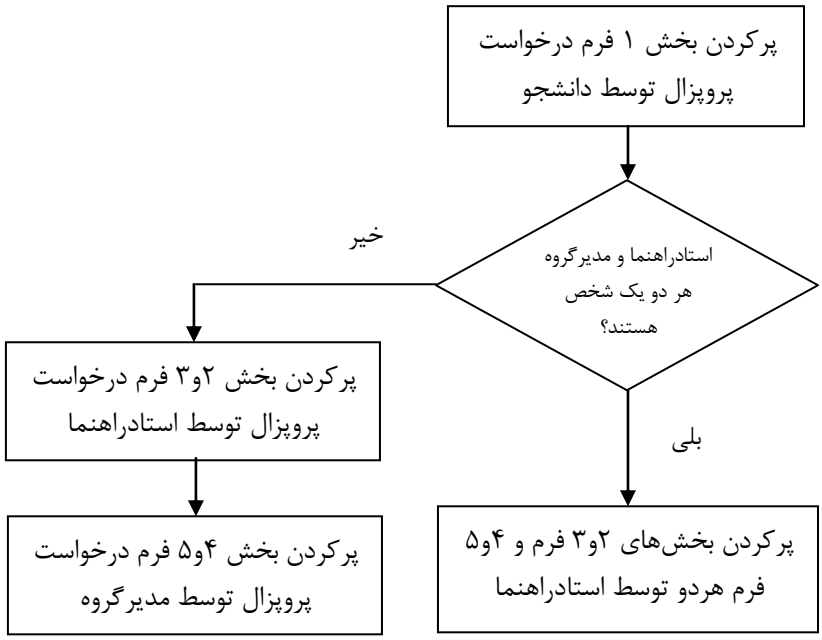
در حالتی که در شکل ۲-۵ رخ داده، کاربر وظیفه t_6 را اضافه کرده است و یال مربوط به آن وظیفه باعث شده است که نمایش جریان‌کاری نتواند در یک لایه صورت بگیرد و یک لایه به نمودار جریان‌کاری اضافه شود. جریان‌کاری پیچیده در مقابل جریان‌کاری خوب قرار دارد. در جدول ۱-۱ ویژگی‌های یک جریان‌کاری خوب آورده شده است.

۲-۵-۲. پیچیدگی مفهومی

گاهی افزایش پیچیدگی جریان‌کاری با تفکر ترتیبی کاربر صورت می‌گیرد. مثل آن که کاربر، ابتدا وظایف را متوالی فرض کرده است و منتظر اتمام یک وظیفه قبل از دیگری شده است. در صورتی که امکان اجرای موازی وظیفه‌ها وجود داشته است [25]. مثال دیگر دادن وظیفه‌ای بیهوده به کاربر است. فرض کنید کاربر ابتدا برای گرفتن تایید نهایی منتظر گرفتن تایید استاد مشاور و سپس استاد راهنما باشد. در صورتی که ممکن است لزوما نیازی به این ترتیب نباشد. مثلا ممکن است بتوان این تایید گرفتن را به صورت موازی پیاده کرد و به نوعی ارتباطات را در آن کاهش داد [25].



شکل ۲-۶– نمودار جریان‌کاری طراحی شده توسط کاربر

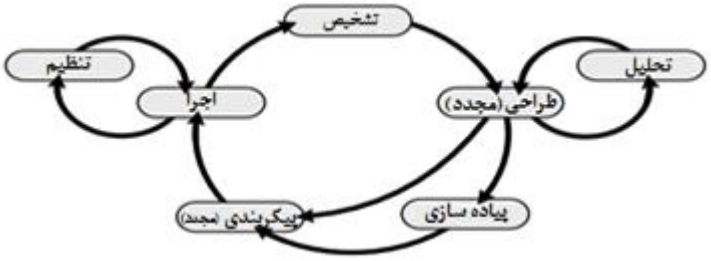


شکل ۲-۷ – نمودار جریان‌کاری تکامل یافته توسط سامانه

نمودارهای اولیه جریان‌کاری و حالت ساده شده آن را در شکل های ۲-۶ و ۲-۷ ملاحظه می‌کنید. شکل پیشنهاد شده توسط سامانه که جریان‌کاری آن در عمل ساده تر شده است، در شکل ۲-۷ آمده است. در واقع این مثالی از ساده‌سازی ساختاری جریان‌کاری با توجه به منابع مشترک است. بدین ترتیب در عمل، جریان‌کاری کاهش پیچیدگی یافته است.

۲.۶ ساده‌سازی

در ساده‌سازی جریان‌های‌کاری اقداماتی ممکن است صورت بگیرد. از جمله مرتب‌سازی، مسیریابی و غیره. ساده‌سازی در مراحل مختلفی ممکن است انجام شود. به طور مثال می‌توان ساده‌سازی را در هر یک از مراحل طراحی، پیاده‌سازی، پیکربندی، اجرا و عیب یابی صورت داد. شکل ۲-۸ چرخه این مراحل را نشان می‌دهد.



شکل ۲-۸- چرخه مدیریت فرآیندهای سازمانی

در شکل ۲-۸ در ارتباط با طراحی (مجدد) تحلیلی صورت می‌گیرد که سازگاری فرآیند سازمانی را از نظر صحت داشتن مورد بررسی قرار می‌دهد. به طور کلی ساده‌سازی را می‌توان به دو دسته ساده‌سازی ساختاری و ساده‌سازی مفهومی تقسیم کرد.

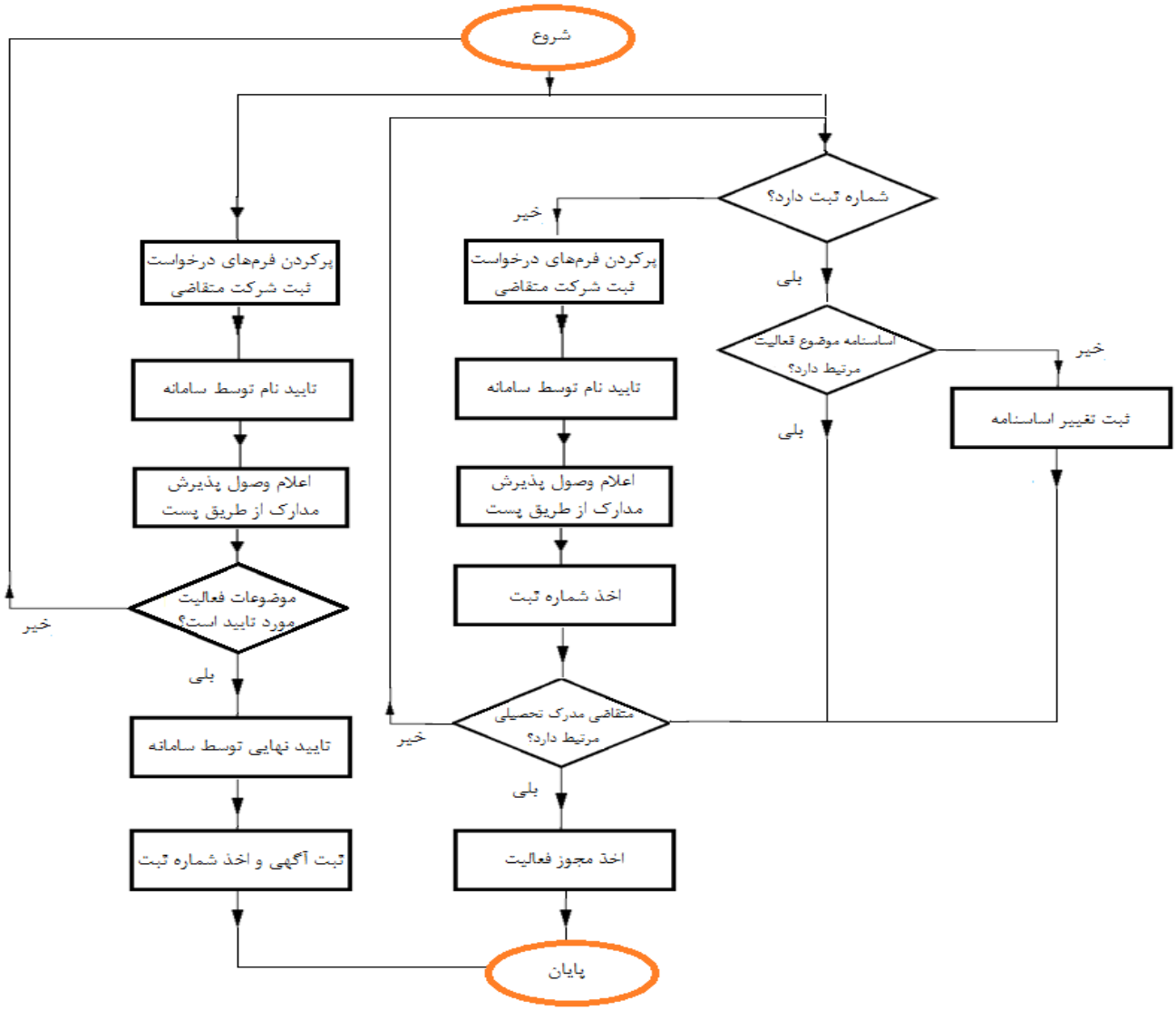
۲-۶-۱. ساده‌سازی ساختاری

در ساده‌سازی ساختاری به بهبود چینش اجزای ساختار جریان‌کاری می‌پردازند. این اجزا می‌توانند شامل گره‌های خاص مثل گره‌های XOR، Or-Join، یا And-Split باشد. در این حالت می‌توان نقش افراد و نیز ارتباطات آن‌ها را در کاهش پیچیدگی درنظر گرفت [87]. به طور کلی برای بررسی صحت ساختاری جریان کاری لازم است در نظر بگیریم که هر عنصر یا قطعه در جریان‌کاری شامل پارامترهای زیر باشد:

- نوع قطعه که می‌تواند آغازی، پایانی، ساده و یا مرکب باشد،
- تاخیر قطعه که مقداری است مثبت
- نوع ورودی که می‌تواند ترتیبی، و-الحاق و یا-الحاق باشد
- منبع که می‌تواند ماشین، ابزار، و یا پرسنل باشد
- وضعیت می‌تواند اختصاص داده شده باشد

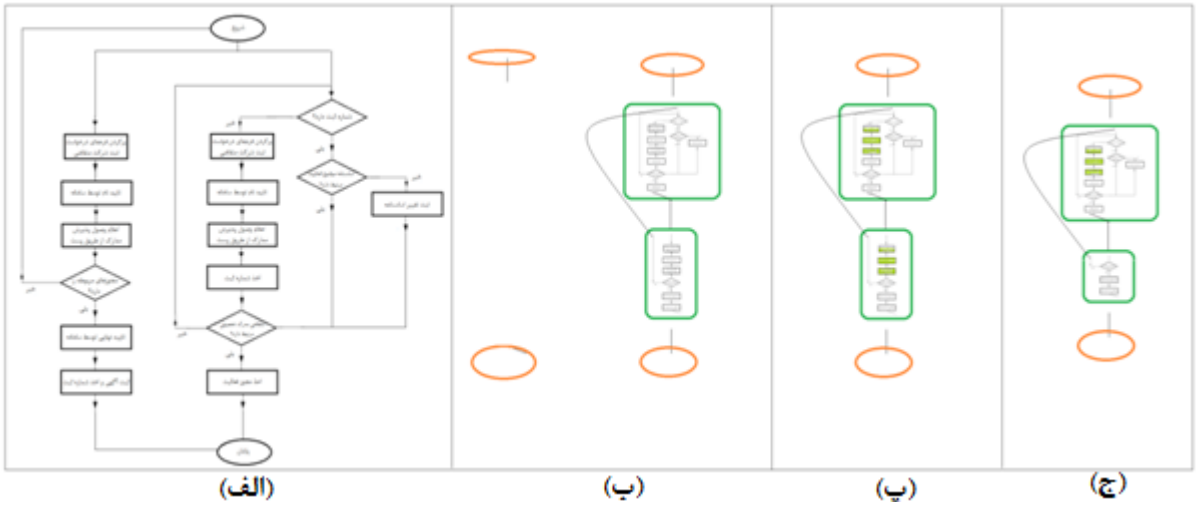
راهکارها به کاهش پیچیدگی ساختاری می‌تواند با بررسی‌های متفاوتی صورت گیرد. به عنوان مثال در صورتی که بررسی تکرارهای اتصال‌ها را داشته باشیم تا بتوان میزان خوانایی را افزایش داد و یا بهبود فهم جریان‌کاری داشت، در این صورت کاهش تعداد اتصالات جریان‌کاری به عنوان پارامتری تعیین کننده در این رابطه مطرح می‌شود [48]. جاگذاری قطعات[1] نیز در بررسی توپولوژیکی مطرح است که به چه صورت قرار بگیرند. به عنوان مثال در طراحی خودکار مدارهای الکترونیکی قبل از مسیریابی[2] معمولا جاگذاری قطعات مطرح می‌شود [8]. این‌که قطعات مدار با کمترین سیم‌بندی در فضای برد مدار چاپی قرار بگیرند مساله‌ای است که برای کاهش پیچیدگی خوانایی و طراحی مدارهای الکترونیکی تحت عنوان جاگذاری قطعات مطرح می‌شود. به عنوان مثالی از ساده‌سازی ساختاری مورد استفاده سازمان‌ها می‌توان مورد ارائه شده در شکل ۲-۹ را مثال زد.

[1] - Component Placement
[2] - Routing



شکل ۲-۹- جریان‌کاری پیچیده

پس از بررسی و تحلیل جریان‌کاری شکل ۲-۹ متوجه می‌شویم که در گره شروع جریان کاری یک or مستتر هم بوده که در نمودار پنهان شده است. برخی گام‌ها تکرار شده‌اند. و در برخی گام‌ها بازگشت به عقب زائد بوده است. پس از ساده‌سازی شکل ۲-۱۰ حاصل می‌شود.



شکل ۲-۱۰- مراحل ساده سازی نمودار جریان‌کاری پیچیده

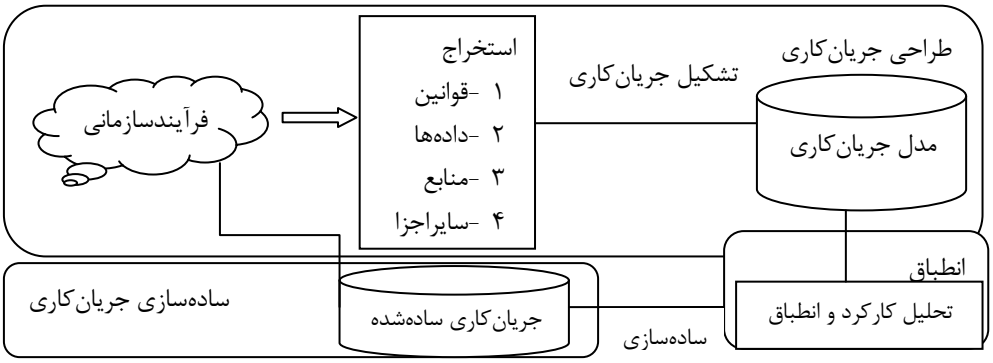
در شکل ۲-۱۰ یک سری مشابه‌ها پیدا شده است و بر اساس قوانین ارائه شده در روش ادغامی کـه در [48] مطرح شده، جریان‌کاری ساده‌تر شکل (۲-۱۰-ج) در آن حاصل شده است.

مثال کاربرد ساده‌سازی جریان‌های‌کاری را می‌توان در کاربردهای بهینـه‌سـازی چنـدوجهی (MDO[1]) یافت [85]. به عنوان مثال کار ارائه شده در [87] را ببینید که مساله ساده‌سازی جریان‌کاری را از چند جهت مورد بررسی قرار داده است.

۲-۶-۲. ساده‌سازی مفهومی

در ساده‌سازی مفهومی از مدل مفهومی استفاده می‌کنند. اسـتخراج داده‌هـای مفهـومی بـرای بدسـت آوردن چنین مدلی الزامی است. مدل فرآیند ابـزاری بسـیار مهـم و حیـاتی در آمـاده‌سـازی داده‌هـای فرآیندها می‌باشد و اهمیت آن در فرآیند تحلیل جریان‌های‌کاری قابل صرف‌نظـر نیسـت. فرآینـدهـای سازمانی در طی طراحی جریان‌کاری به وسیله‌ی مدل توصیف می‌شوند. طراح برای استخراج این مـدل لازم است قوانین کنترل، داده‌ها، منابع و سایر اجزا مانند رخدادهای آغازین و پایانی را اسـتخراج کنـد. در شکل ۲-۱۱ مدل مفهومی مرتبط با این عملیات مشاهده می‌شود.

[1] - Multi Disciplinary Optimization



شکل ۲-۱۱- مدل طراحی جریان‌کاری

در شکل ۲-۱۱ مراحل ساده‌سازی از تشکیل جریان‌کاری تا بهبود فرآیند سازمانی دیده می‌شود. همان‌طور که شکل ۲-۱۱ نشان داده است، در ساده‌سازی و کاهش پیچیدگی جریان‌کاری معمولا بخش‌های مدل‌سازی و تحلیل جریان‌کاری فرآیندها، ثبت وقایع جریان‌های‌کاری ساده شده را برای نگهداری جریان‌های‌کاری ساده‌شده داریم. این چرخه مراحل تحلیل تا ساده‌سازی جریان‌کاری را نشان می‌دهد.

۲ ۷. آنالیز آماری

۲-۷-۱. تابع وزن PMI[1]

اندازه PMI [63] امکان مدل‌کردن اطلاعات متقابل بین خصیصه‌ها و کلاس‌ها فراهم می‌کند. این اندازه از تئوری اطلاعات مشتق شده است. اطلاعات متقابل نقطه به نقطه (PMI) بین قطعه w و کلاس i برمبنای سطح هم‌رخدادی بین کلاس i و قطعه w تعریف می‌شود. هم‌رخدادی مورد انتظار کلاس i و قطعه w ، برمبنای وابستگی متقابل توسط $P_i.F(w)$، و هم‌رخدادی درست توسط $F(w).P_i(w)$ داده می‌شود. اطلاعات متقابل به صورت نسبت بین این دو مقدار تعریف می‌شود و توسط رابطه زیر مطرح می‌شود:

$$M_i(w) = \log(\frac{F(w).P_i(w)}{F(w).P_i}) = \log(\frac{P_i(w)}{P_i})$$

هنگامی که $M_i(w)$ بزرگتر از صفر باشد، عبارت w به کلاس i به طور مثبت مرتبط (هم‌رخداد) می‌شود. هنگامی که $M_i(w)$ کمتر از صفر باشد، عبارت w به کلاس i به طور منفی هم‌رخداد می‌شود. اندازه PMI شدت ارتباط را در نظر می‌گیرد.

[1] - Point-wise Mutual Information System

۲-۷-۲. متریک آماری انتساب بهینه زیر الگو (OSPA[1])

در [91] متریکی جدید برای ارزیابی کارایی چند شی‌ای معرفی شده است. این متریک مبتنی است بر ساخت واسرستین[2]؛ هنگامی که حالت‌های چندهدفی، کاردینالیتی (تخمین تعداد اهداف) مشابهی دارند، از روش واسرستین، تفسیر بهینه انتساب از دست دادن فواصل[3] را به ارث می‌برد.

این متریک به طور مشترک خطای تخمین مکان هدف و تعداد را ارزیابی می‌کند. برای هر $p \geq 1$، $c > 0$ و $X = \{x_1,...,x_m\}$، $Y = \{y_1,...y_n\}$، OSPA برای سه حالت مجزا تعریف می‌شود. اگر $m = n = 0$، آن‌گاه $\bar{d}_{c,p}(X,Y) = \bar{d}_{c,p}(Y,X) = 0$. در غیر این صورت $m \leq n$:

$$d_{c,p}(X,Y) \triangleq \frac{1}{n}(\min_{\pi \in \Pi_n} \sum_{i=1}^{m} d_c(x_i, y_{\pi(i)})^p + c^p(n-m))^{\frac{1}{p}}$$

که در آن، Π_n مجموعه جایگشت ها بر $\{1,2,...,n\}$ را درفضای حالت نشان می‌دهد.

و اگر $m > n$، $\bar{d}_{c,p}(X,Y) = \bar{d}_{c,p}(Y,X) = 0$. دو پارامتر متریک OSPA مرتبه p و برش[4] c هستند؛ مرتبه p حساسیت به نویزها را نشان می‌دهد، در حالی که برش c وزن دادن نسبی جریمه‌های منتسب به خطاهای کاردینالیتی و مکان‌یابی را مشخص می‌کند.

خطاهای کاردینالیتی و مکان‌یابی به صورت زیر تعریف می‌شوند [91]:

$$e_{c,p}^{card}(X,Y) = (\frac{c^p(n-m)}{n})^{\frac{1}{p}}$$

$$e_{c,p}^{loc}(X,Y) = (\frac{1}{n}.\min_{\pi \in \Pi_n} \sum_{i=1}^{m} d_c(x_i, y_{\pi(i)})^p)^{\frac{1}{p}}$$

اگر $m \leq n$، و $e_{c,p}^{card}(X,Y) = e_{c,p}^{card}(Y,X)$ و $e_{c,p}^{loc}(X,Y) = e_{c,p}^{loc}(Y,X)$ اگر $m > n$.

در حالت خاص که $p = 2$ باشد برای $x, y \in \chi$ داریم: $d_{c,p=2}(x,y) := \min(c, \|x-y\|)$.

۲-۸. روش‌های یادگیری

۲-۸-۱. روش یادگیری آماری LSA

تحلیل معنایی پنهان (LSA) روشی است آماری در تبدیل جریان‌های‌کاری و یا قطعات جریان‌کاری به بردار خصیصه. این روش سعی در یافتن معنی پنهان قطعات جریان‌کاری در جریان‌های‌کاری می‌کند.

---

[1] - Optimal Sub-Pattern Assignment
[2]- Wasserstein
[3] - miss-distance
[4] - cut-off

در این روش در یک جریان‌کاری، قطعات جریان‌کاری قابل مشاهده هستند، ولی عنوان آن‌ها پنهان (Latent) است.

۲-۸-۲. روش یادگیری آماری (SCL[1])

از روش یادگیری ساختارها مرتبط (SCL) [71] می‌توان برای انطباق دامنه استفاده کرد. با قطعات جریان‌کاری برچسب خورده مفروض از دامنه مبدا و قطعات برچسب نخورده از هر دو دامنه مبدا و مقصد، روش SCL ابتدا مجموعه‌ای از m خصیصه را انتخاب می‌کند که در هر دو دامنه پرتکرار رخ می‌دهند، و هم‌چنین پیش‌بین‌های خوبی از برچسب مبدا (خصیصه‌هایی که بزرگترین اطلاعات متقابل را داشتند برچسب مبدا بودند) هستند. این خصیصه‌ها، خصیصه‌های محوری نامیده می‌شوند که فضای خصیصه اشتراکی دو دامنه را نشان می‌دهند. سپس روش SCL همبستگی هر خصیصه محوری با سایر خصیصه‌های غیرمحوری در هر دو دامنه را محاسبه می‌کند. این کار ماتریس همبستگی W را تولید می‌کند که در آن سطر i بردار مقادیر همبستگی خصیصه‌های غیرمحوری با iامین خصیصه محوری می‌باشد. به طور شهودی مقادیر مثبت نشان می‌دهد که آن خصیصه‌های غیرمحوری به طور مثبت با iامین خصیصه محوری در دامنه مبدا یا در دامنه جدید همبسته شده‌اند. این کار، ارتباط خصیصه بین دو دامنه را برقرار می‌کند. پس از آن، تجزیه مقدار منفرد (SVD[2]) برای محاسبه تقریب خطی با بعد پایین θ از ماتریس W به کار می‌آید (یعنی k بردار منفرد سمت بالا چپ، پس از transpose داریم). مجموعه نهایی خصیصه‌ها برای آموزش و برای تست مجموعه اصلی خصیصه‌های x ترکیب شده با θx می‌باشد که خصیصه‌های مقدار واقعی k را تولید می‌کند. دسته‌بند و ماشین یادگیر ساخته شده با استفاده از خصیصه‌های ترکیب شده و داده برچسب زده شده در دامنه مبدا باید در هر دو دامنه مبدا و مقصد کار کند.

۲-۸-۳. روش یادگیری انتقالی (روش آماری SFA)

پن و همکاران [71] روشی مشابه SCL در بالاترین سطح ارائه کردند. این روش فقط به ترتیبی کار می‌کند که در آن فقط نمونه‌های برچسب‌خورده در دامنه مبدا و نمونه‌های برچسب نخورده در دامنه مقصد هستند. این روش فاصله بین دامنه‌ها را با استفاده از روش انتساب خصیصه طیفی (SFA) برای انتساب کلمات ویژه دامنه از دامنه‌ها مختلف به خوشه‌های یکپارچه، با کمک کلمات مستقل از دامنه به عنوان رابط، کم می‌کند. کلمات مستقل از دامنه مشابه کلمات محوری در روش SCL هستند. روش SFA ابتدا با ساخت گراف دووجهی با کلمات مستقل از دامنه به عنوان مجموعه‌ای از گره‌ها و کلمات ویژه به عنوان مجموعه سایر گره‌ها کار می‌کند. کلمه ویژه دامنه اگر هم‌برخورد باشد به کلمه مستقل از دامنه، متصل شده است. وزن اتصال، تکرار هم‌برخوردی آن‌هاست. سپس الگوریتم خوشه‌بندی طیفی براساس نظریه طیفی گراف به گراف دووجهی اعمال می‌شود تا هم‌انتسابی کلمات مستقل از

[1] - Structural Correspondence Learning
[2] - Singular Value Decomposition

دامنه و وابسته به دامنه به مجموعه‌ای از خوشه‌های خصیصه داشته باشیم. برای این کار، آن را می‌بریم در فرمت ماتریس خلوت (تُنُک[1]) و از تکنیک‌های ارزیابی شباهت بهره می‌بریم. برای اندازه‌گیری این‌که چه مقدار کلمات بیشتر از انتظار رخ می‌دهند، اگر آن‌ها مستقل باشند، از رابطه (۲-۱) استفاده می‌شود.

$$f(s,t) = \log\left(\frac{\frac{c(s,t)}{N}}{\frac{\sum_{i=1}^{n} c(i,t)}{N} \times \frac{\sum_{j=1}^{m} c(s,j)}{N}}\right) \qquad (۱-۲)$$

که در آن f(s,t) اطلاعات متقابل (PMI) بین دو کلمه s و t است و c(s,t) تعداد کلی کلمات را نشان می‌دهد که در آن دو کلمه s و t همبرخوردی دارند، و n و m به ترتیب تعداد کلی کلمات s و t را نشان می‌دهد، و

$$N = \sum_{i=1}^{n} \sum_{j=1}^{m} c(i,j)$$

سپس برای دو کلمه s و v (توسط بردارهای خصیصه s و v نشان داده می‌شوند)، وابستگی τ (v, s) از کلمه v به کلمه s به صورت رابطه (۲-۲) محاسبه می‌شود [73]:

$$\tau(v,s) = \frac{\sum_{t \in \{x | f(v,x) > 0\}} f(s,t)}{\sum_{t \in \{x | f(s,x) > 0\}} f(s,t)} \qquad (۲-۲)$$

نمره وابستگی τ (v, s) می‌تواند به عنوان خصیصه‌های شاخص وزنی PMI از کلماتی که با کلمه v به اشتراک گذاشته می‌شود، تفسیر شود.

۲-۸-۴. روش یادگیری نظریه طیفی گراف

این روش از طیف ماتریس شباهت داده استفاده می‌کند تا برای خوشه‌بندی کاهش بعد را اعمال کند. بنابراین، ایده پایه این است که گرافی وزن‌دار از داده ورودی بطریقی بسازیم که گره‌های گراف نقاط داده باشند و هر لبه وزن‌دار درجه شباهت بین هر جفت گره را نشان دهد. روش‌های خوشه‌بندی طیفی از بردارهای Eigen ماتریس affinity نرمال‌شده استفاده می‌کنند. در واقع آن‌ها ماتریس Eigen می‌سازند که مرتبط با k بزرگترین مقدار بردار Eigen هست.

۲-۸-۵. روش یادگیری عمیق (DNN[2])

در مقایسه با روش‌های یادگیری سنتی که معمولا از معماری‌های یادگیری سطحی بهره می‌برند، یادگیری عمیق اساسا از استراتژی‌ها نظارتی و یا بدون نظارت در معماری‌ها عمیق برای یادگیری خودکار بازنمایی سلسله‌مراتبی استفاده می‌کند. اخیرا برخی از روش‌ها یادگیری عمیق با شبکه‌های عصبی (DNN) ترکیب شده‌اند.

[1] - Sparse Matrix
[2] - Deep Neural Networks

۲-۸-۶. متریک ارزیابی روش یادگیری (CA[1])

براساس رابطه متریک CA می‌توان مقدار دقت تخمین دامنه جریان‌کاری را به صورت زیر محاسبه کرد:

$$CA = \frac{tn + tp}{tn + fn + tp + fp}$$

که در آن:

	truly relevant	*truly irrelevant*
retrieved wf segments	*true positive* (tp)	*false positive* (fp)
not retrieved wf segments	*false negetive* (fn)	*true negetive* (tn)

۲-۹. متریک‌های ارزیابی جریان‌کاری

۲-۹-۱ اتصالات (cp[2])

متریک cp [78] میزان اتصالات میان قطعات جریان‌کاری را نشان می‌دهد. هر قدر اتصالات از قطعات پایین‌تر به قطعات بالاتر بیشتر شود (یعنی دور داشته باشیم) با افزایش تودرتویی و پیچیده‌تر شدن جریان‌کاری مواجه می‌شویم. کم بودن میزان cp در یک جریان‌کاری نشان می‌دهد که جریان-کاری از تعداد اتصالات کمتر و ساده‌تری برخوردار است [78]. رابطه (۲-۳) را ببینید.

$$cp = \begin{cases} \dfrac{\left|\left\{(T_1,T_2) \in S \times S \mid \overline{T_1} \neq \overline{T_2} \wedge (T_1 \cap T_2) \neq 0\right\}\right|}{\left|\overline{S}\right|.(\left|\overline{S}\right|-1)}, for \left|S\right| > 1 \\ 0, for \left|S\right| \leq 1 \end{cases} \quad (۲-۳)$$

که در آن T_1 و T_2 دو قطعه از جریان‌کاری هستند. هم‌چنین S تعداد کلیه قطعات جریان‌کاری است. در صورتی که ارتباطی بین این دو قطعه برقرار نباشد $(T_1 \cap T_2) = 0$، میزان cp صفر خواهد شد. در غیر این‌صورت cp مقداری بین صفر و یک خواهد شد.

۲-۹-۲ چگالی

نسبت تعداد ارتباطات قطعات جریان‌کاری به ارتباطات کل میان قطعات را چگالی می‌نامند. هر قدر چگالی یک جریان‌کاری بیشتر باشد احتمال تودرتویی و پیچیدگی آن بیشتر است. معمولا سعی می-کنند حدود محدودیت‌هایی را برای پایین نگه داشتن چگالی در ساخت جریان‌های‌کاری در نظر بگیرند.

[1] - Classification Accuracy Rate
[2] - Coupling

فصل سوم:

کارهای انجام شده قبلی

۳ ـ۱. مقدمه

یکی از اهداف مهم کاهش پیچیدگی جریان‌کاری در اغلب مواقع دستیابی به شکل ساده‌تر جریان-کاری می‌باشد. نیاز به ساخت جریان‌کاری هست که ساختارش نزدیک به واقعیت باشد. بدین منظور رویکرد تعمیر و ساده‌سازی جریان‌های‌کاری به کمک می‌آید. خصوصا زمانی که داده‌های ناقص، گنگ و یا مبهم در گزارشات عملکرد داده‌ها دخیل شده باشد. تاکنون راه‌حل‌های مختلفی برای ساده‌سازی و کاهش جریان‌های‌کاری ارائه شده است، که شکل‌های مختلف جریان‌کاری را بر حسب اندازه و پیچیدگی مورد بررسی قرار داده‌اند. گاهی نتیجه‌ی عملیاتی شکل‌های مختلف یک جریان‌کاری با هم یکسان‌اند. به عبارت دیگر گاهی جریان‌های برابر به وجود می‌آیند؛ در این حالت ممکن است برخی شکل‌ها از سادگی بیشتری برخوردار باشد. به عنوان مثال در آن‌ها حلقه‌های کمتر، گره‌های کمتر، اتصال دهنده های کمتر یا با تعداد تلاقی کمتر، به عبارتی بهینه‌تر باشد. امروزه با توجه به بزرگ شدن جریان‌های‌کاری و نیاز به مطالعه آن‌ها به صورت فشرده‌تر مسائل کاهش پیچیدگی جریان‌کاری اهمیت ویژه‌ای یافته‌اند.

وین و همکاران در [9] برای کاهش جریان‌کاری صحت آن‌ها را مورد بررسی قرار داده اند. یک جریان‌کاری لازم است خاتمه‌پذیر و محدود باشد. چنین جریان‌کاری از صحت برخوردار است [12]. برای توضیح بیشتر در کتاب آلست و هی که در رابطه مدیریت جریان‌های‌کاری نوشته شده است، با یک مثال مجموعه قوانینی برای فرآیندهای سازمانی سرویس شخصی‌سازی شده تعریف شده است [13]. این قوانین شامل قانون مسیریابی نیز می شود که برای افزایش خوانایی شبکه جریان‌کاری قابل استفاده است. در این کتاب با بیان یک قضیه، صحت شبکه جریان‌کاری به دو خصیصه شناخته شده زنده بودن[۱] و محدود بودن[۲] شبکه مدار کوتاه شده[۳] مرتبط شده است. ولی با این وجود، بررسی صحت شبکه را با قضیه‌ای که آورده، کافی ندانسته است؛ زیرا چالش‌های خوانایی و قابلیت نگهداری را در آن حل نشده یافته است. برای بررسی صحت جریان‌کاری می‌توان شرط‌های خاتمه پذیر بودن و یا زنده بودن را مورد بررسی قرار داد. زنده بودن بدین معنا که مدل عنصر مرده ندارد. عنصر مرده یعنی با استفاده از آن به انتها برسیم و کار بدون نتیجه خاتمه پیدا کند. در این صورت با در نظر گرفتن مواردی مانند بن بست که باعث نقض شرایط صحت جریان‌کاری می‌شوند، می‌توان کاهش پیچیدگی جریان‌کاری داشت. در این صورت می‌توان عواملی را کشف کرد که منجر به بن بست شده‌اند و پیچیدگی را کاهش داد [21]. برای بهبود روند جریان‌کاری نیز به عنوان مثال می‌توان کار مقاله مندلینگ و همکاران [30] را در نظر گرفت که در آن شرط ها را در جلوترین نقطه ممکن قرار می-دهد. این کمک می‌کند روند جریان‌کاری ساده‌تر گردد.

---

[1] - Liveness
[2] - Boundedness
[3] - Short-Circuited

مسیرهای عملکردی در دنیای واقعی معمولا نه یک جهته محض هستند و نه چنـان پـیچدرپـیچ و تودرتو که انجامشان وابسته به وجود برخی شرایط شود. اگر در طراحی جریان‌کاری معیارهای محـدود کننده‌ای رعایت شود، تا حدود زیادی می‌توان از پیچیدگی‌های زائد جلوگیری کرد. در این صورت یـک جریان‌کاری خوش‌رفتار به وجود خواهد آمد. برای این کـه بتـوان یـک جریـان‌کـاری معقـول داشـت، محققین محدودیت‌هایی را برای ساخت یک جریان‌کاری تعریف می‌کنند. به عنوان مثال شرط‌های زیر را برای خوش‌رفتار بودن (معقول بودن) جریان‌کاری تعریف می‌کنند:

۱ -زنده‌بودن

۲ -کران‌دار بودن

۳ -معکوس‌پذیری

ممکن است جریان‌کاری طراحی شده با توجه به تکرار نقش ها در عملیات چنان پیچیده شـود کـه روند جریان‌کاری را دشوار سازد. در این صورت افزایش پیچیدگی فرآیند بـا توجـه بـه در نظـر گـرفتن منابع مشترک و یا نقش‌ها می‌تواند صورت بگیرد.

کارهای مرتبط در برخورد با پیچیدگی جریان‌های‌کاری و بهبود آن‌ها می‌توانند در سه گـروه اصـلی تقسیم شوند: استخراج خصیصه‌های مشترک، ارزیابی شباهت، و بهبود جریان‌های‌کاری. در ادامـه ایـن بخش پس از بیان رویکردهای برخورد با افزایش پیچیدگی جریان‌های‌کاری بـه ایـن سـه گـروه مـی-پردازیم.

۳ ۲. رویکردهای برخورد با افزایش پیچیدگی جریان‌کاری

در این جا رویکردهای برخورد با افزایش پیچیدگی را از دو جهت طراحی سـامانه و ابزارهـای مـورد استفاده برای کاهش پیچیدگی مورد بررسی اجمالی قرار می‌دهیم.

دو رویکرد در برابر افزایش پیچیدگی هنگام پیاده سازی و طراحی سامانه وجود دارد:

۱ -کاربر محدود شود: یعنی اجازه طراحی جریان‌هـای‌کـاری دلخـواه را از اول نـدهیم [10]. و عملیات پذیرفته نشود.

۲ - کاربر محدود نشـود: یعنـی اجـازه بـدهیم جریـان‌کـاری شـکل بگیـرد. سـپس آن را بـه جریان‌کاری معقول، خوش‌رفتار و ساختاریافته تبدیل کنیم [23].

در هر دو صورت، ممکن است جریان‌کاری طراحی شده چنان پیچیده شود که خوانـایی آن کـاهش یابد. به عنوان مثال شکل ۶ را در نظر بگیرید که کاربر با انتخاب یک یال موجب افزایش پیچیـدگی در نمودار جریان‌کاری شده است. با این کار یک ساختار هم‌پوشان در مدار ایجاد شده است کـه تصـمیم-گیری در مورد ساختاریافتگی جریان‌کاری را دشوار می‌سازد.

در رویکرد یک، محدودیت هایی از همان ابتدا اعمال شده است (به عنـوان مثـال خاتمـه پـذیری). برای بررسی رویکرد دو در صورتی که کاربر محدود نشود، لازم است ساده‌سـازی صـورت بگیـرد. بـرای

حل این مساله راه‌حل‌هایی در نظر گرفته اند. به عنوان مثال کایپزوفسکی و همکاران در [23] پیشنهاد داده‌اند، که ابتدا معادل‌های جریان‌کاری را طوری در نظر بگیریم که معنای جریان‌کاری بـرهم نخـورد. بدین منظور می‌توان از روش‌هایی مانند تولید گرامر برای تعریف یـک زبـان بـرای جریـان‌کـاری بهـره ببریم. سپس با یک تبدیل زبان مساله را حل کنیم. به عنوان مثال رزا و همکاران ۱۲ الگوی کـاهش را پس از در نظر گرفتن نحو انتزاعی جریان‌های‌کاری مورد بررسی قرار داده اند [38].

برای ساده‌سازی یک جریان‌کاری تعداد لایه‌های تودرتویی یک جریان‌کاری باید محدود باشد. تعداد پیوند قطعات در لایه همسان لازم است محدود و مشخص باشد. هم‌چنین بایستی تعداد پیوندهای یک قطعه از یک لایه، (با قطعات دیگر از لایه‌های پایین تر) از تعداد پیوندهای لایه‌های همسان محـدودتـر باشد. برای این که این محدودیت‌ها اعمال شود، بایستی قوانینی در سامانه اعمال شود. برطبـق تعریـف جریان‌کاری عملکردی بازگشتی دارد. ایـن عملکـرد بازگشـتی بـرای اجرایـی شـدن در دنیـای واقعـی براساس یک قانونی خاتمه می‌یابد. این قانون مـی‌توانـد از قبـل در هسـتی‌شناسـی[1] سـامانه مـدیریت جریان‌کاری تعریف شده باشد. بـه عنـوان مثـال در پـژوهش آلپـر و همکـاران بـرای خلاصـه‌سـازی از برچسب‌های معنایی مبتنی بر هستی شناسی بهره برده شده است [15]. هم‌چنین براساس این تعریـف در جریان‌کاری می‌توان به طور مناسبی هم از جریان کارهـای پیـاده سـازی شـده قبلـی (بـه صـورت زیرکلاس) استفاده کرد و هم از مسیرهای یک جهته بهره برد. چنـین جریـان‌کـاری یـک جریـان کـار منطبق بر واقعیت است. در زمینه طراحی نمودارهای جریان‌کاری عده‌ای از محققان سامانه‌های تولیـد جریان‌کاری خودشان را بگونه‌ای طراحی کرده‌اند که در یک محیط ویرایشی، جریان‌کاری توسط کـاربر طراحی شود [11]. هم‌چنین اخیرا برخی این رویکرد را پیدا کرده‌اند که خود کاربر نمـی‌توانـد جریـان-کاری را به بهترین نحو طراحی کند [7]. و لازم است سـامانه‌هـای خودکـاری بـرای ارائـه جریـان کـار متناسب با مساله مورد نظر کاربر توسط ماشین ارائه شود. تلاش بر این است که تولید جریان منحصـرا بر دوش سامانه خودکار و به دور از چشم کاربر صـورت نگیـرد. رویکـرد آن‌هـا ایـن اسـت کـه مراحـل تصمیم‌گیری برای قرار دادن هر قطعه از جریان‌کاری در سامانه خودکار، به کاربر نشـان داده شـود. تـا کاربر بنابرخواسته‌اش بتواند در تولید جریان‌کاری مطلوبش دخالت کند.

در صورتی که ساختار جریان‌کاری با تغییر پویا همراه باشد، می‌توان بـه معمـاری سـامانه خبـره بـا آنتولوژی اشاره کرد که در آن کاربر حین ساخت توسط سامانه راهنمایی شـود [7]. بـا درنظـر گـرفتن تغییرات پویا در ساختار جریان‌های‌کاری به منظور ساخت مدل‌هایی که بر چـالش‌هـای کـارایی تفـوق یابند روش‌هایی نیز برای مدل‌سازی فرآیند همکارانه [32] ارائه شده‌اند. ردگیری وضعیت فعالیت افـراد از زنجیره منبع- عمل [33] و روش‌های اخذ برچسب گره‌ها و ردگیری [34] مثـال‌هـای دیگـر در ایـن حالت هستند. استفاده از فرآیندکاوی برای یادگیری از تغییرات فرآیند در سامانه های تکـاملی [35] و کشف مدل های شبیه سازی [36]، روش‌های دیگری هستند که برای مدیریت تغییرات پویای جریـان-های‌کاری ارائه شده اند.

[1] - Ontology

به عنوان مثال در روشی که توسط شن و همکاران در [33] برای ردگیری وضعیت افراد ارائه شده است، یک روش دو مرحله‌ای برای مشخص کردن شبکه‌هایی در نظر گرفته شده است که در طول زمان تکامل می‌یابند. در گام یک این روش گره‌ها توسط خوشه بندی k-means گروه‌بندی می‌شوند و بر اساس نقش هایشان در شبکه دسته بندی شده‌اند. در گام دوم روشی برای مطالعه تکامل شبکه با استفاده از یک روش نظارتی ارائه شده است. در این روش مجموعه‌ای از رخدادها که در شبکه اتفاق می‌افتد برای نقش‌های شبکه تعریف می‌شود. سپس رخدادهای از قبل تعریف شده که در شبکه رخ می‌دهند و قوانینی پیدا می‌شوند، که آن‌ها را با استفاده از کاوش قوانین انجمنی توصیف می‌کنند.

فارغ از این نکته که ساختارهای جریان‌های کاری پویا هستند یا خیر، به طور کلی روش های کاهش جریان‌کاری با اهداف زیر مطرح می‌توانند مطرح شوند:

۱ - انتزاعی کردن رفتارهای غیرتکراری: عناصر نموداری متنوعی می‌توانند در یک نمودار وجود داشته باشند. برای فهم موثر سامانه، می‌توان به روش‌هایی برای همبسته کردن[1] عناصر نمودار در یک دید بسنده کرد.

۲ - محدود کردن سامانه به رفتارهای قابل پذیرش.

بهبود روند جریان‌کاری و کاهش پیچیدگی‌ها از جنبه های مختلفی صورت می‌گیرد [87]. بدین منظور تحلیل‌هایی اهمیت می‌یابند که به استخراج نقش‌ها و تحلیل نقاط مختلف جریان‌کاری می‌پردازند. روش معمول برای استخراج نقاط مختلف در جریان‌های‌کاری فرآیند کاوی می‌باشد. کاری که توسط آلست و فهلاند در [16] صورت گرفته است و کاری که توسط آلست و گوتچر در [28] قبل تر از آن انجام شده است، شامل ساده‌سازی برمبنای فرآیندکاوی می‌شود. با فرآیند کاوی می‌توان به بازیابی مدل فرآیندها بر مبنای میزان شباهت آن‌ها نسبت به یک مدل جستجوی مفروض، پرداخت. بدین صورت که از روی رخدادهای ثبت شده در طی اجرا، به کشف الگوهای جریان‌کاری بپردازیم. بدین ترتیب از تخصیص منابع نیز می‌توان پشتیبانی صورت داد (مثل آن‌چه که در پژوهش لیو و همکاران آمده است [17]). از سوی دیگر با استفاده از روش‌های مدرن می‌توان تحلیلی بر روند جریان‌کاری داشت. مانند روش‌های ارائه شده در [18-19]. به منظور تحلیل جریان‌های‌کاری ابتدا لازم است جریان‌کاری استخراج و کشف شود. روش های مختلفی مثل الگوریتم ژنتیک، الگوریتم آلفا [20] و الگوریتم استخراج مکاشفه‌ای[2] [22] و فازی کاو [28] برای کشف الگوهای جریان‌های‌کاری وجود دارند. در این میان الگوریتم‌های استخراج آلفا و استخراج مکاشفه‌ای از بقیه معروف‌ترند. الگوریتم آلفا که بیشتر در موارد قطعی کار می‌کند، نویز و وظایف تکراری را در نظر نمی‌گیرد. ولی الگوریتم های استخراج مکاشفه ای و ژنتیک نویز و وظایف تکراری را نیز در نظر می‌گیرند. بدین منظور می‌توان موتیف کاوی، کاوش جریان‌کاری با الگوریتم ژنتیک و غیره را نام برد. به عنوان مثال در پژوهش ماگویر و همکاران در [27] یک سری جریان‌کاری بیولوژیکی ارائه شده است. ماگویر و همکاران یک

[1] - Correlate
[2] - Heuristic Mining

سری جریان‌های کاری پایه را به عنوان موتیف در نظر گرفتند. این موتیف‌ها براساس نوع گره‌ها و یال-ها، استخراج می‌شوند. بدین منظور از اطلاعات آماری این موتیف‌ها (مثل عمق گراف و میانگین استفاده) برای انتخاب آن‌ها استفاده می‌شود [27].

در بسیاری از کارهای قبلی کاهش جریان‌کاری به منظور انتزاعی کردن رفتارهای غیرتکراری آن‌ها صورت می‌گرفته است. به عنوان مثال پژوهش آلست و گویچر [28] بدین منظور ساده‌سازی جریان‌کاری صورت گرفته است. آن‌ها در [28] از مفاهیم درشتی[1] برای تعریف سطح جزئیات و انتزاعی کردن جریان‌کاری استفاده می‌کنند. الگوریتم فازی‌کاو در یک ساده‌سازی تعاملی از متریک‌های همبستگی برای ساده‌سازی در سطح انتزاع مطلوب استفاده می‌کند [28]. پژوهش آلست و گویچر برای این کار از فیلترکردن یال‌ها برای ساده‌سازی جریان‌کاری بهره برده است؛ یعنی یال‌هایی را حذف می‌کند که همبستگی کمی دارند. ولی این روش، بیشتر به بهبود نمایش جریان‌کاری کمک می‌کند، و در تعمیر و ساده‌سازی کاربرد چندانی ندارد. هم‌چنین، در پژوهش آلست و گویچر فرآیند ساده‌سازی بدون در نظر گرفتن محتوا صورت می‌گیرد. بدین ترتیب اولویت‌بندی هم براساس نقش‌ها و منابع در ساده‌سازی نمی‌تواند وجود داشته باشد.

سایر روش‌هایی که بر مبنای انتزاعی کردن رفتارهای غیرتکراری در جریان‌کاری می‌باشند، عبارتند از:

الف- تولید گرامر برای تعریف یک زبان و استفاده از زبان‌های معادل ساده‌تر [23] و [38].

ب -ساده‌سازی با روش براساس ادغام [40].

در این روش‌ها ساده‌سازی بیشتر برمبنای ساختار می‌باشد. به عنوان مثال کایپزوفسکی در [23] پیشنهاد داده است، که ابتدا معادل‌های جریان‌کاری را طوری در نظر بگیریم که معنای جریان‌کاری برهم نخورد. بدین منظور می‌توان از روش‌هایی مانند تولید گرامر برای تعریف یک زبان برای جریان-کاری بهره ببریم. سپس با یک تبدیل زبان مساله را حل کنیم. چالشی که در این روش‌ها وجود دارد امکان وجود حلقه در جریان‌کاری است که باید در نظر گرفته شود. ضمن این که ساده‌سازی براساس ساختار در آن بر محتوا ارجحیت دارد.

در [12] توسط فلندر و همکاران برخی از موارد نقض روند یک جریان‌کاری آورده شده است. جدول ۳-۱ را ببینید. در این جدول ۵ کلاس نقض جریان‌کاری تعریف شده است.

[1] - Granularity

جدول ۳-۱. دسته‌های نقض جریان‌های‌کاری [12]

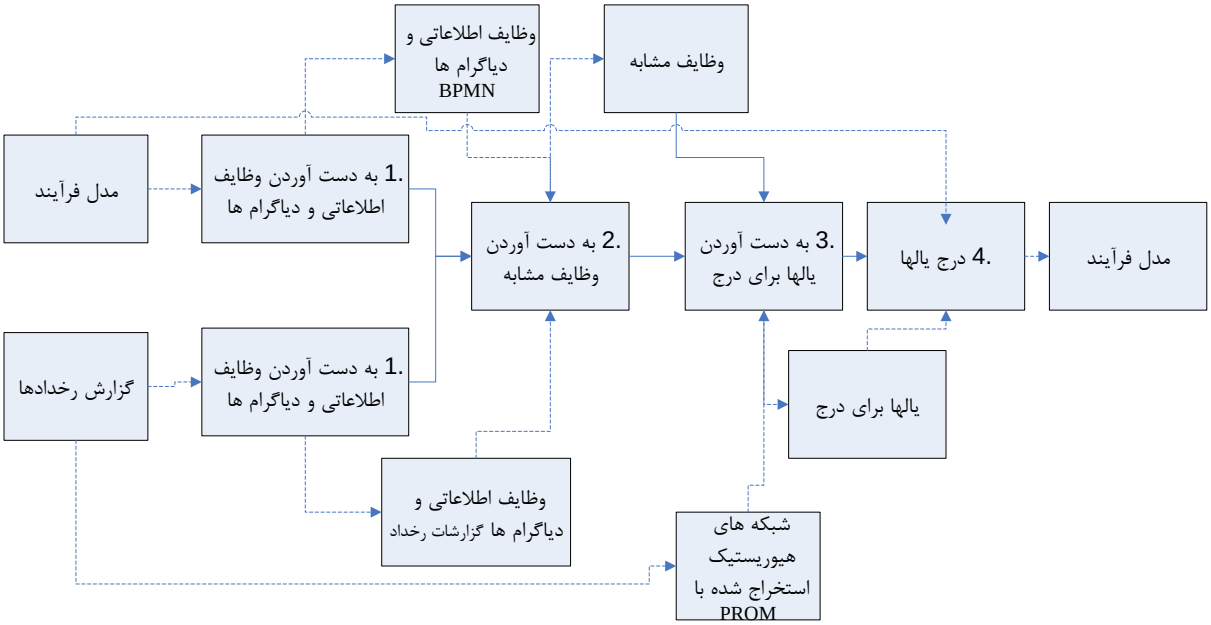
کلاس نقض	مجموعه‌های نقض ممکن
۱ (شبکه جریان‌کاری)	مکان‌های آغاز یا پایان نامناسب گره هایی که قویا همبند نیستند
۲ (صحت موقتی[1])	مکان‌هایی که توسط قطعات S پوشش داده نشده‌اند مجموعه‌های نادرست یا مجموعه‌های درهم[2]
۳ (کران‌دار بودن)	مکان‌های بدون کران دنباله‌های بدون کران
۴ و ۵ (زنده بودن)	انتقال‌های مرده دنباله‌های فاقد حیات

در واقع هنگامی که یک جریان‌کاری معقول نباشد و یا تعداد زیادی وظیفه دارد و وابستگی‌های جریان کنترلی زیادی در آن درگیر هستند، بررسی می‌تواند زمان خیلی زیادی ببرد یا حتی غیر ممکن باشد. روش‌های مختلفی برای رفتار با این پیچیدگی وجود دارد. وین و همکاران در [9] کاهش اندازه جریان‌کاری را پیشنهاد داده‌اند. به طوری که جریان‌کاری خصوصیات اولیه‌اش را با توجه به مساله تحلیل خاص نگه دارد. این کاهش اندازه به نوعی همان کاهش پیچیدگی جریان‌کاری است، که در [12] بیشتر توضیح داده شده است.

از روش‌هایی که برای محدود کردن سامانه به رفتارهای قابل پذیرش می‌توان به کار فهلاند و آلست در [16] اشاره کرد که در آن ساده‌سازی جریان‌کاری مبتنی بر فرآیندکاوی با آشکارسازی[3] معرفی شده است. در مقاله [16] جریان کاری اسپاگتی مانند که مملو از اتصالات تو در تو بود پس از فرآیندکاوی به جریان‌کاری با حداقل اتصالات و تمیز تبدیل شده‌است. روش ارائه شده فهلاند و آلست در [16] سامانه را به رفتارهای قابل پذیرش محدود می‌کند. در این روش اجراهای جزئی با ترتیب به نحوی صورت می‌گیرد که در آن بتوان ترتیب اجرای اجزای جریان‌کاری را یافت. روش ارائه شده در [16] نمونه اجرایی فرآیندهایی را می‌یابد که در گزارش‌ها ثبت شده‌اند، ولی در مدل فرآیند ظاهر نشده‌اند. سپس آن‌هایی را بازیابی می‌کند که از دست رفته‌اند. برای تعمیر مدل‌های فرآیند می‌توان هم‌چنین به روش ارائه شده توسط فرناندز-روپرو و همکاران در [37] اشاره کرد. فرناندز-روپرو و همکاران در [37] جریان‌های‌کاری تعمیر شده را به صورت فرآیندهایی در نظر گرفته‌اند که در آن‌ها تکرار اتصالات وجود ندارند. در شکل ۳-۱ روش کار ارائه شده در [37] را می‌بینید که در آن یال‌های لازم برای درج استخراج می‌شوند. این یال‌ها پس از استخراج وظایف مشابه از دیاگرام‌ها و گزارشات

[1] - Interim
[2] - Confusion
3 - Unfolding

رخداد فرآیندها به دست می‌آیند. آن‌ها برای این کار از وظایف مشابه و شبکه‌های هیوریستیک استخراج شده به کمک شبکه جریان‌کاری بهره می‌برند.



شکل ۳-۱- تکنیک ها به تعمیر BPMN با استفاده از رخدادهای ثبت شده [37]

به عنوان مثالی دیگر می‌توان به مرجع [10] اشاره کرد که الگوریتم کاهش جریان‌کاری اتفاقی (SWR[1]) را به طور تکراری با مجموعه‌ای از قوانین کاهش به یک جریان‌کاری اعمال می‌کند، تا زمانی که فقط یک تک وظیفه باقی بماند. هر زمان که یک قانون کاهش اعمال می‌شود، ساختار جریان‌کاری تغییر می‌کند. بعد از چند تکرار فقط یک وظیفه باقی خواهد ماند. هنگامی که به این حالت دست یافتیم، وظیفه باقی‌مانده شامل متریک کیفیت سرویس (QoS[2]) مربوط به جریان‌کاری در حال تحلیل می‌شود. برای همین مجموعه قوانین کاهش که می‌توانند به یک جریان‌کاری مفروض اعمال شوند، به مجموعه‌ای از عملیات معکوس مربوط می‌شوند که می‌تواند برای ساخت یک جریان‌کاری استفاده شود. در مرجع [10] فقط به ساخت جریان‌های‌کاری اجازه داده می‌شود که بر مجموعه‌ای از سامانه-های از قبل تعریف شده مبتنی هستند. این کار کاربران را از ایجاد جریان‌های‌کاری نامعتبر حفاظت می‌کند. جریان‌های‌کاری نامعتبر شامل خطاهای طراحی، مانند خاتمه ناپذیری، بن‌بست‌ها و جداشدن نمونه‌ها[3] می‌شوند [10].

[1] - Stochastic Workflow Reduction
[2] - Quality of Service
[3] - Splitting of Instances

در مرجع [10] کاردوسو و همکاران، متریک QoS را از این جهت انتخاب کرده اند که در شبکه سازی‌ها از اهمیت ویژه‌ای برخوردار است. این مرجع برای محاسبه متریک‌های QoS از مجموعه ۶ تایی قوانین کاهش مجزا استفاده می‌کند: (۱) ترتیبی، (۲) موازی، (۳) شرطی، (۴) تحمل‌پذیر در برابر خطا، (۵) حلقه، و (۶) شبکه. در این مرجع، وفاداری[1] به عنوان تابعی از طراحی موثر دیده می‌شود. کاردوسو و همکاران برای یافتن موثرترین وظایف در نتیجه نهایی طراحی پیشنهاد کرده‌اند که از مدل‌سازی خطی برای پیدا کردن وفاداری یک وظیفه نسبت به وظیفه دیگر استفاده کنیم. در مطالعه انجام شده در مرجع [10] در باب الگوهای شبکه بحثی به میان نیامده است. کاردوسو و همکاران دو دلیل برای این کار ذکر کرده اند: یک این که الگوی شبکه متمایل است تقسیمی ساختاریافته و سلسله‌مراتبی از طراحی جریان‌کاری مفروض به سطح‌ها فراهم کند. این کار به منظور ساده‌سازی فهم با گروه کردن وظایف مرتبط به واحدهای عملکردی، صورت می‌گیرد. دلیل دوم آن‌ها این است که قوانین کاهش ساده هستند؛ این کار فهم ایده فرآیند کاهش را ساده می‌کند.

۳ ۴. مباحث کارهای مرتبط

هنگام طراحی نمودار جریان‌کاری، یک رویکرد این است که از ابتدا اجازه طراحی جریان‌های‌کاری دلخواه را ندهیم [10]. رویکرد دیگر این است که اجازه دهیم جریان‌کاری شکل بگیرد. سپس آن را به جریان‌کاری معقول، خوش‌رفتار و ساختاریافته تبدیل کنیم [23]. در صورتی که رویکرد دوم را در نظر بگیریم، به طور کلی برای کاهش پیچیدگی جریان‌های‌کاری، حذف و یا تغییر چینش اتصالات جریان-های‌کاری به نحوی صورت می‌گیرد که قابلیت فهم جریان‌کاری، صحت آن و یا روند آن بهبود یابد. در تحلیل کارهای معرفی شده که به ساده‌سازی و کاهش نمودارهای جریان‌کاری پرداخته‌اند، تحقیقات مختلفی صورت گرفته است. کارهای انجام شده مرتبط را در این‌جا به سه موضوع دسته‌بندی می-کنیم: استخراج خصیصه مشترک، ارزیابی شباهت، و بهبود جریان‌کاری. در ادامه این بخش ابتدا به معرفی کارهای مرتبط با استخراج خصیصه‌ها مشترک می‌پردازیم. سپس کارهای مرتبط با ارزیابی شباهت را مورد بررسی قرار می‌دهیم، چرا که روش این رساله در بهبود جریان‌کاری از ارزیابی شباهت به عنوان بخش کلیدی کار بهره می‌برد. در نهایت کارهای مرتبط با بهبود جریان‌کاری را مورد بررسی قرار می‌دهیم.

[1] - Fidelity

الف- استخراج خصیصه مشترک

لوداسشر و همکاران [54] خصیصه‌های جریان‌های‌کاری علمی و سازمانی را مقایسه کرده‌اند. در نهایت نتیجه گرفته‌اند که ادغام روش‌های تحلیل جریان‌کاری مبتنی بر گردش داده و مبتنی بر گردش کنترل می‌تواند نتایج جدید و بینشی برای هر دو جریان‌کاری علمی و سازمانی به دست دهد. با این وجود، به خاطر اختلافات بیشتر بین جریان‌های‌کاری علمی و جریان‌های‌کاری سازمانی، مقالات ترجیح می‌دهند یا روی جریان‌های‌کاری علمی کار کنند و یا روی جریان‌های‌کاری سازمانی. به عنوان مثال، گاریجو و همکاران [55] تحلیلی عملی بر توصیفات ۲۶۰ جریان‌کاری علمی ارائه کرده اند. آن‌ها کاتالوگی از انتزاعات مفهومی مستقل از دامنه (DI[1]) و مختص دامنه (DS[2]) برای گام‌های جریان‌کاری تعریف کردند به نام موتیف جریان‌کاری علمی. موتیف جریان‌کاری نمونه که آن‌ها در کاتالوگشان به صورت دستی مشخص کردند شامل آماده‌سازی داده، حذف داده، انتقال داده، بازیابی داده و هم-پوشانی جریان‌کاری می‌شدند. آن‌ها موتیف را شبیه عبارت «الگو» تعریف کردند که به بهترین تمرین-های استقراریافته برای حل مسائل بازگشتی اشاره می‌کند. در نهایت، آن‌ها درباره توزیع انتزاعات در میان سیستم‌های جریان‌کاری مختلف بحث کردند. سپس تاکید کردند که سیستم‌های جریان‌کاری مختلف هسته‌ای مشترک از انتزاعات جریان‌کاری به اشتراک می‌گذارند. آن‌ها گام‌های ماکزیمم، مینیمم، و میانگین در جریان‌های‌کاری به ازای هر دامنه را مقایسه کردند و کاتالوگی از موتیف‌های جریان‌کاری علمی را برای انتزاعی کردن جریان‌های‌کاری ترجیح دادند. آن‌ها نشان دادند که موتیف-های آماده‌سازی مشترک‌ترین نوع موتیف هستند و سپس آن‌ها دریافتند که اغلب این موتیف‌ها در سایر حوزه‌ها و سیستم‌های جریان‌کاری یافت خواهند شد. در نهایت آن‌ها پیشنهاد دادند که تشخیصشان از موتیف‌های جریان‌کاری به عنوان مجموعه‌ای از هیوریستیک‌ها استفاده شود برای ساخت خودکار انتزاع‌های جریان‌کاری [55]. هدف کار [55] فهم گروه‌بندی گام‌های جریان‌کاری علمی است که عملیات دستکاری داده سطح بالا بامعنی شکل می‌دهند. با این حال، این کار نیز برای پوشش دادن جریان‌های‌کاری سازمانی تعمیم داده نشده بود. به علاوه، این مقاله روشی خودکار برای استخراج موتیف‌های جریان‌های‌کاری پیشنهاد نداده بود.

مقاله [15] راه‌حلی برای ساخت خودکار خلاصه‌های توصیف جریان‌کاری ارائه کرده است. در واقع، این مقاله روشی برای خلاصه‌سازی جریان‌کاری با تحلیل جریان‌های‌کاری علمی ارائه کرده است. رهیافت ارائه شده جریان‌کاری را با بدویات به خوبی تعریف شده، یعنی ترکیب، بازکردن، و حذف، مجددا می-

[1] - Domain Independent
[2] - Domain Specific

نویسد. اگرچه تاثیر خلاصه‌سازی خودکارشده آن‌ها نتایج بهتری نسبت به خلاصه‌سازی کاربران نشان داده است، اما از گرانی و هزینه‌بر بودن برچسب‌گذاری موتیف جریان‌کاری رنج می‌برد.
جالب‌تر اینکه، سایر انواع کشف موتیف جریان‌های‌کاری علمی می‌توانند بیان شوند. به عنوان مثال، با استفاده از اطلاعات انتقال حالت بین موتیف‌های جریان‌کاری، الگوهای تکراری می‌توانند از انباره استخراج شوند تا هم زیرواحدهای کاری فراهم شوند و در طراحی جریان‌کاری استفاده شوند، و هم الگوهای خوب مشخص شده برای هدایت طراحان جریان‌کاری به کار آیند [27]. در مقاله [27] جریان‌های‌کاری با جابجایی موتیف‌های تکراری با ماکروها به جریان‌های‌کاری بصری فشرده شده می‌رسند. جابجایی موتیف‌های بازگشتی با ماکروها می‌تواند انتزاع مفهومی سلسله مراتبی، فشرده‌سازی بصری، خوانایی بهبودیافته، و کارآیی وظیفه موثر از نظر هزینه فراهم کند [27]. از سایر کارهای مرتبط می‌توان به مقاله [56] اشاره کرد که هدفش تولید انتزاعات به صورت خودکار است. مقاله [56] دو متریک طول توصیف کمینه (MDL[1]) و اندازه توصیف کرده‌است تا بهترین انطباق موتیف‌ها را مبتنی بر پیداکردن گرامری آن‌ها بیابد. MDL بهترین ساختار را به عنوان آن ساختاری می‌گیرد که طول توصیف کل مجموعه‌داده را کمینه می‌کند. کار مقاله [56] الگوریتم دسته‌بندی آماده می‌کند که مبتنی بر کار برچسب‌گذاری داده متنی دستی عمل می‌کند که می‌تواند زمان‌بر و پرهزینه باشد. به‌علاوه، کاربران اغلب هنگام بیان منظورهایشان از کلمات مختلفی استفاده می‌کنند.

ب- ارزیابی شباهت

ارزیابی شباهت بخش مهم روش رساله در ساده‌سازی جریان‌کاری است. این رساله به منظور طراحی یک جریان‌کاری مرتب و ساده شده از روی مدل نامناسب یا پیچیده جریان‌کاری، شباهت جریان‌های‌کاری را ارزیابی می‌کند. شباهت در رساله حاضر اندازه‌ای است که جریان‌های‌کاری معادل را نشان می‌دهد. اهداف مرتبط با رهیافت‌ها به اندازه‌گیری‌های شباهت در جریان‌های‌کاری در جدول ۳-۲ نشان داده شده‌اند.

[1] - Minimum Description Length

جدول ۳-۲. بکارگیری ارزیابی شباهت

مرجع	مبتنی بر متن	مبتنی بر ساختار			اهداف مرتبط با ارزیابی شباهت	حوزه داده	سطح
		موتیف برچسب	موتیف ساختاری	توپولوژی			
[57]	BW	خیر	خیر	خیر	جستجو و بازیابی	S	F1
[58]	تکرار مجموعه تگ‌ها	خیر	خیر	تکرار MS	جستجو داده	S	F1
[27]	خیر	خیر	زیرگراف‌ها مشترک، ML	خیر	فشرده‌سازی بصری Wf	S	F2
[59]	خیر	خیر	زیرگراف‌ها مشترک، MM	خیر	سازمان‌دهی Wf و دسته‌بندی	S	F2
[56]	خیر	برچسب معنایی	MDL و اندازه	خیر	جستجو و بازیابی	S	F2
[60]	خیر	خیر	خیر	انطباق برچسب، GED	جستجو و بازیابی	B	F2 ،F1
[61]	BW ،BT	خیر	خیر	MS ،PS ،GED	راحت‌کردن استفاده مجدد Wf	S	F2 ،F1
[15]	خیر	برچسب معنایی	خیر	خیر	خلاصه‌سازی	S	F1
[62]	خیر	خیر	خیر	PS	بهبود طراحی Wf	S، B	F2
[83]	خیر	خیر	MM	خیر	جستجو و بازیابی	S، B	F2

اختصارات: F1= تک-قطعه، F2= چند-قطعه، S= علمی، B= سازمانی

جدول ۳-۲ روش‌های موجود مرتبط با ارزیابی شباهت جریان‌های‌کاری را نشان می‌دهد. این جدول اهداف مرتبط با ارزیابی شباهت استفاده شده در مقالات را نشان می‌دهد. کارهای مشابه که روی جریان‌های‌کاری انجام شده‌اند با هدف‌های مختلفی صورت گرفته‌اند. یعنی روش پیشنهاد شده در رساله حاضر از روش‌هایی که پژوهش‌گران قبلی برای اهدافی به جز ساده‌سازی جریان‌های‌کاری به کار برده‌اند، استفاده می‌کند. بسیاری از پژوهشگران به دنبال جستجو و بازیابی در میان داده‌های جریان-های‌کاری هستند. همین‌طور که در جدول دیده می‌شود، روش ارائه شده رساله به طور ویژه دنبال-کننده کار کوهی‌ور و زاهدی در [83] می‌باشد. دیجمان و همکاران در [60] اندازه‌های شباهت را که شامل شباهت انطباق برچسب، و GED بود مطالعه کرده‌اند. متریک انطباق برچسب مبتنی است بر مقایسه‌های دو به دو برچسب‌های اجزا، و متریک GED هر دو برچسب گره و توپولوژی مدل‌های فرآیند را به حساب می‌آورد. دو سطح درشتی ارزیابی شباهت وجود دارد. برخی تحقیقات علاقه‌مند به

ارتباط الگوهای تکی (تک قطعه) [61]، [15] هستند، در حالی که سایر تحقیقات علاقه‌مند بـه ارتبـاط زیرجریان‌های‌کاری مشترک (چندقطعه‌ای) برای تحلیل شباهت هستند [56]، [27] و [59]. کـار ارائـه شده حاضر به تحلیل شباهت در سطح چندقطعه‌ای مرتبط می‌شود. یک روش ارزیـابی شـباهت کـه از تکنیک‌های انتخاب خصیصه مبتنی بر متن بهره می‌برد کار ارائه شده توسـط کوتچنـت و همکـاران در [57] می‌باشد. کوتچنت و همکاران با جریان‌کاری به عنوان گروهی از کلمات (BWs) رفتـار مـی‌کننـد. در مقاله کوتچنت و همکاران [57] برای هر جریان‌کاری قطعه پیش‌پردازش، تعـداد رخـداد عبـارت‌هـا شمرده می‌شود. سپس مقادیر شباهت بین جفـت جریـان‌هـای‌کـاری را مـی‌شـمارد. روش کوتچنـت و همکاران با استفاده از تحلیل معنایی پنهان (LSA)، رخداد عبارت‌ها را در نظر می‌گیرد. سپس مقـادیر شباهت بین جفت جریان‌های‌کاری را تولید می‌کند. مشترک‌ترین گام انتخاب خصیصه در رهیافت‌های مبتنی بر متن گام پیش‌پردازش (گام حذف کلمات توقف و بازگردان کلمات به ریشه‌شان مثـل flies-> fly) می‌باشد [63]. مثال دیگر کار اسـتانویچ و همکـاران اسـت کـه در آن بیشـترین تکـرار مجموعـه ماژول‌ها (MS) و تکرار مجموعه تگ‌ها جستجو شده‌اند [58].

از میان کارهایی که تاکنون برای ساده‌سازی جریان‌های‌کاری صورت گرفته است، برخی بـه اصـلاح شکل جریان‌کاری می‌پردازند. به عنوان مثال این کـار بـا بهـره‌گیـری از قـوانینی ماننـد RWF[1] [9]، و بسته، ترتیبی [4] انجام می‌شود. در مقاله‌ای تحت عنوان مدیریت پیچیدگی مدل‌های فرآیند از طریـق تغییرات نحو انتزاعی نیز با در نظر گرفتن ساختار جریان‌های‌کاری به صـورت فرمـال، الگوهـایی بـرای مدیریت ارائه شـده اسـت [38]. در مقالـه رزا و همکـاران [38] سـعی شـده اسـت بـا در نظـر گـرفتن خصیصه‌های زبانی روش‌های ارائه شده به حل مساله پیچیـدگی جریـان‌هـای‌کـاری بپردازنـد. در ایـن بررسی از میان ۱۲ الگوی زبانی ارائه شده برای کاهش جریان‌های‌کـاری، الگوهـای پیمانـه‌ای کـردن و ترکیب از نظر سادگی استفاده از همه الگوهای دیگر بیشتر در جریان‌های‌کاری استفاده شده‌اند. سـری دیگر مقالات مانند پژوهش گلدمن و نگُکو در [47] با عملگرهای ریاضی، به تعبیر ریاضی جریان‌کـاری می‌پردازند و آن تعبیر ریاضی را ساده می‌کنند که ممکن است به کـاهش پیچیـدگی بیانجامـد. برخـی دیگر مانند روش مبتنی بر پردازش آماری اطلاعات استفاده موتیف [27] یا با اندازه‌گیـری پارامترهـایی

[1] - Reduced Work-Flow

مثل ارزیابی وفاداری [10] وظایف به صورت آماری به کاهش جریان‌های‌کاری می‌پردازند. مزیت روش های آماری مبتنی بر موتیف این است که به کمک توزیع موتیف‌ها می‌توان جریان‌های کاری پویا را دسته‌بندی کرد [43]. به عنوان مثال مقاله [43] با وزن‌دهی رفتار موتیف‌ها جریان‌های‌کاری را دسته-بندی می‌کند. این کار در ساده‌سازی نهایی، در جایی که همه داده‌ها را برای ساده‌سازی نداریم، می-تواند اعمال شود. آلست و همکاران در [53] روشی در تعادل بین بیش‌برازش و کم‌برازش ارائه کرده‌اند که در دو گام مدل فرآیند را از گزارشات عملکرد به نحوی استخراج می‌کند. گام اول با کشف کنترل فرآیند مدل فرآیند استخراج می‌شود. گام دوم با کمک فرد خبره بین بیش‌برازش و کم‌برازش مدل تعادل ایجاد می‌کند. بدین ترتیب یک مدل فرآیند ساده‌شده در اختیار کاربر قرار می‌دهد. روش ارائه شده در [16] نیز تعادلی بین بیش‌برازش و کم‌برازش مدل ایجاد می‌کند. ولی این روش به صورت خودکار این کار را انجام می‌دهد و دخالت فرد خبره را با تنظیم مجدد خروجی بر یک تابع کنترل کیفیت شده را حذف می‌کند. این روش به استخراج و آشکارسازی یک وضعیت از وضعیتی دیگر به صورت آماری می‌پردازد. روشی که مبتنی است بر آشکارسازی [16] روشی مناسب برای ساده‌سازی و تعمیر جریان‌های‌کاری است. هدف روش ارائه شده در [16] بستن مدل فرآیند به یک مدل ساده‌تر به نحوی است که رفتار مطلوب را به دست گیرد.

مدل‌های فرآیند که توسط روش‌های ارائه شده تاکنون ساده شده‌اند متمایل به مدلی هستند که در تعادلی در بیش‌برازش و کم‌برازش باشند. در صورتی که مدل بیش‌برازش باشد به قدری که رفتار خوب را تعمیم دهد یادگیری مناسب در آن‌ها صورت نگرفته است. هم‌چنین اگر مدل دچار کم‌برازش شده باشد هر رفتاری در آن مجاز است. ولی روش‌هایی که تاکنون ارائه شده‌اند فقط توانسته‌اند در شرایطی خاص و با ارزیابی کیفیت فرآیند به یادگیری مدل دست یابند. به عنوان مثال مقالات [16] و [44] روشی ارائه کرده‌اند که در آن‌ها در صورتی که هم یک مدل فرآیند از قبل تولید شده داشته باشیم و هم گزارشات عملکرد ثبت شده فرآیند را داشته باشیم می‌تواند به نتیجه‌گیری مناسبی دست یابد. نمونه دیگر کار آقایان گانتر و آلست در [28] می‌باشد که در آن فقط برای گزارشات عملکرد ثبت شده از قبل نظری به منظور ساده‌سازی می‌دهد. اشکالی که این روش دارد این است که در مواجه با فرآیندهایی که تا به حال ندیده است، تضمینی وجود ندارد که در هر شرایطی با کمترین مشاهدات

مدلی متعادل شده به عنوان خروجی داشته باشیم. در این رساله قصد داریم با کمترین مشاهدات به یک مدل مناسب به نحوی دست یابیم که بتواند تعادل خوبی در رفتار داشته باشد. در ادامه این بخش متریک‌های ارائه شده برای ارزیابی را معرفی می‌کنیم.

ج- بهبود جریان‌کاری و متریک‌ها

برای بهبود جریان‌کاری لازم است ابتدا متریک‌های ارزیابی روش‌های ساده‌سازی مورد بررسی قرار بگیرند. ساده‌سازی جریان‌کاری از دو دیدگاه کمی و کیفی می‌تواند مورد بررسی قرار گیرد. در این‌باره تاکنون تحقیقاتی صورت گرفته است که در ادامه ذکر می‌گردند:

بررسی از نظر کمی پیچیدگی در [29] صورت گرفته است. در [30] پیچیدگی از نظر کیفی و چالش‌های مهم آن از این جهت مورد بررسی قرار گرفته‌است. نویسنده مقاله [30] در کارهای قبلی خود روی همین مبحث از نظر قابلیت فهم کار کرده است و نتیجه گرفته است که بین قابلیت فهم نمودار جریان‌کاری و تعداد یال‌ها رابطه مستقیم منفی وجود دارد؛ بدین معنی که هر قدر تعداد یال-های به کار رفته در نمودار جریان‌کاری کمتر باشد، قابلیت فهم آن بالاتر است.

متریک‌های دیگری نیز می‌توان برای تحلیل جریان‌کاری پیشنهاد کرد. از جمله متریک کاردوز توسعه‌یافته (ECaM[1])، متریک سایکلوماتیک توسعه‌یافته (ECyM[2]) و متریک ساختاری (SM[3]) [26]. متریک ECaM متریکی است که برای شبکه‌های پتری معرفی شده است. تمرکز این متریک بر نحو مدل است و از پیچیدگی رفتار صرف‌نظر می‌کند. متریک ECyM بر رفتار در شبکه جریان‌کاری تمرکز می‌کند و از پیچیدگی خود مدل صرف‌نظر می‌کند. متریک SM الگوهای طراحی نمودار جریان‌کاری را مدنظر قرار می‌دهد. سایر متریک‌ها را می‌توان در [29]، [30] و [31] ملاحظه کرد. مقاله [31] یک سری متریک ساده به عنوان معیار ارزیابی معرفی کرده است که در جدول ۳-۳ آمده است.

بسیاری از مقالات از متریک‌های جدول ۳-۳ برای ارزیابی‌های خود بهره برده‌اند. به عنوان مثال، مقاله [37] از میان متریک‌های بالا تعداد گره‌ها و جداپذیری را استفاده کرده است. هم‌چنین متریک دیگری به اسم اتصال‌پذیری[4] برای مقایسه اولیه شبکه‌های استفاده شده به عنوان مجموعه داده‌اش تعریف کرده است. این متریک نسبت بین تعداد کلی یال‌ها در مدل فرآیند سازمانی (یعنی، گردش-

[1] - Extended Cardose Metric
[2] - Extended Cyclomatic
[3] - Structuredness Metric
[4] - Connectivity

های دنباله و پیوندها) و تعداد کلی گره‌ها می‌باشد. این اندازه‌گیری می‌گوید هر قدر اتصال‌پذیری کمتر باشد، مدل‌های فرآیند سازمانی قابلیت فهم و قابلیت تغییرشان بالاتر است.

جدول ۳-۳– متریک‌های ساده که در [31] تعریف شده اند

متریک	توصیف
تعداد گره‌ها	تعداد فعالیت‌ها و عناصر مسیریابی در یک مدل
قطر	طول طولانی‌ترین مسیر از یک گره آغاز به یک گره پایان
چگالی	نسبت تعداد کلی یال‌ها به بیشینه تعداد یال‌ها
جداپذیری[۱]	نسبت تعداد گره‌های قطع به تعداد کلی گره‌ها در مدل فرآیند
دنباله‌پذیری[۲]	درجه‌ای که مدل از دنباله‌های محض کارها درست شده
عمق	بیشینه بلوک‌های ساختاریافته تودرتو در یک مدل فرآیند
ناهمگنی انشعاب‌های اصلی[۳]	تعداد انواع مختلف انشعاب‌های اصلی که در مدل استفاده شده‌اند
دورپذیری[۴]	تعداد گره‌های یک دور برای جمع زدن همه گره‌ها
موازی پذیری[۵]	بیشینه تعداد مسیرها در یک مدل فرآیند که ممکن است به طور موازی به دلیل جداسازهای AND و OR فعال شود

مقاله [38] مطالعه کاملی بر روی سایر متریک‌ها داشته است. این متریک ها به عنوان مثال عبارتند از:

صحت، برازندگی، مناسب بودن از نظر رفتاری، مناسب بودن از نظر ساختاری، دقت رفتاری و غیره.

روش ساده‌سازی خوب است که بتواند معیارهایی را داشته باشد که یک جریان‌کاری را خوب نشان می‌دهد. در انتخاب متریک‌های ارزیابی اگر بتوان ۷ راهنمای [29] زیر را پوشش داد آن متریک خوب است. از نظر مقاله [29] جریان‌کاری خوب است که بتواند موارد زیر را رعایت کند:

۱ -استفاده کمتر از قطعات داشته باشد
۲ -به ازای هر قطعه راه‌های مسیریابی کمینه باشد
۳ -یک رخداد آغاز و پایان استفاده کرده باشد
۴ -مدل‌سازی آن تا حد امکان ساختاریافته باشد
۵ -از عناصر مسیریابی (ساختار کنترل جریانی) «یا» تا حد امکان استفاده نشود
۶ -از برچسب‌های فعالیت فعل-شی استفاده شود
۷ - مدلی را که بیش از ۵۰ عنصر دارد تجزیه کند

[1] - Separability
[2] - Sequentiality
[3] - Gateway Heterogeneity
[4] - Cyclicity
[5] - Concurency

مبتنی بر پیشنهادات بالا مدل‌های بهبودیافته بهینه می‌توانند ساخته شوند. راهنمای [29] به تلاش‌های زیادی مثل تعمیر [44]، [64] بهبود [62] و ساده‌سازی [59] جریان‌کاری مرتبط می‌شود. مبتنی بر این راهنما، تکنیک ارائه شده در [65] برای حذف خودکار رفتار غیرتکراری از گزارشات اجرای فرآیند، بهبودی بر ساخت مدل فرآیند می‌باشد. در مقاله [65] وابستگی‌ها یافت می‌شوند و با کمک ماشین خودکار[1] که از روی گزارش رخداد ساخته شده، حذف می‌شوند. سپس گزارش اصلی بر همان اساس به روز می‌شود. این کار توسط حذف رخدادهای تکی با استفاده از اجرای مجدد فرآیند مبتنی بر تنظیم صورت می‌گیرد. رهیافت دیگر برای بهبود طراحی جریان کار [62] می‌باشد. مقاله [62] وظایف قابل استفاده مجدد را استخراج می‌کند و ترتیب وظایفی (یعنی مسیرهایی) را شناسایی می‌کند که تکراری رخ می‌دهند. با استخراج این مفاهیم از حوزه کسب‌وکار و حوزه آزمایشات علمی، نویسندگان در [62] مساله داده‌کاوی را به فرآیندکاوی تطبیق داده‌اند. مقاله [62] نتیجه گرفته است که رهیافت ارائه شده از این جهت کلی است که می‌تواند با هر سامانه مدیریت جریان‌کاری علمی جفت شود. اگرچه کار مقاله [62] برای ارائه در حوزه‌های بیوانفورماتیک و تجارت الکترونیک جامعیت دارد، اما مشابه کار [61] روی جریان‌های‌کاری علمی متمرکز شده است.

رساله حاضر راه‌حلی کلی به بهبود جریان‌کاری توسط استخراج مفهوم می‌دهد. این کار هنگامی صورت می‌گیرد که برچسبی در حوزه مورد نظر (از این جا به بعد به عنوان حوزه مقصد اشاره می‌شود) نداریم اما داده برچسب‌گذاری شده در حوزه متفاوتی داریم که به عنوان حوزه مبدا (یا کمکی) مدنظر است. کار حاضر مشابه تعمیر مدل فرآیندی است که کشف مدل گردش کنترل [66] مبتنی بر داده رخداد و بررسی انطباقی دارد که مدلی از پیش‌تعریف شده به عنوان نُرم دارد. این کار مشابه حالت مقاله [44] می‌باشد: با مدل فرآیندی مفروض که با واقعیت تطبیق نمی‌خورد می‌توان مدلی را به صورتی پیشنهاد کرد که رفتار مشاهده شده در آن می‌تواند کاملا توسط مدل توصیف شود. مقاله [44] موجب اصلاح جریان کارها با به حداقل رساندن عدم تعادل می‌شود به طوری‌که بخش‌هایی از مدل هم‌چنان حفظ شود که توسط ورود به رویداد نامعتبر نبودند. بخش‌های بعدی روش پیشنهادی را ارائه می‌دهند.

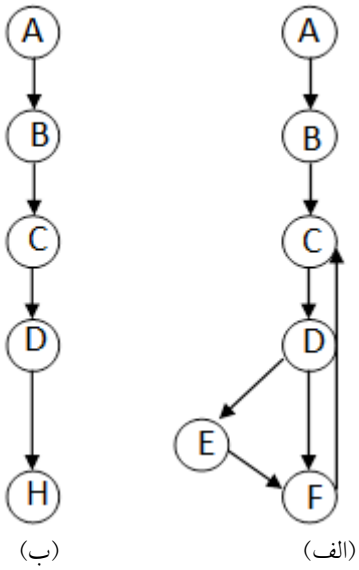
[1] - automaton

فصل چهارم:

روش پیشنهادی

۴ ـ۱ مقدمه

در این بخش روش ارزیابی و مجموعه‌های داده را معرفی می‌کنیم. در واقع جریان‌های‌کاری متمایل به کاهش زمان اجرا و بیشینه کردن کارایی عملیات درونشان هستند. از اهداف مساله ساده‌سازی نیز، تعمیر و فشرده‌سازی نمودار جریان‌کاری می‌باشد به نحوی که هزینه ارتباطات بین عناصر جریان‌کاری کاهش یابد. بنابراین، نتیجه ساده‌سازی شامل بهینه‌سازی عملکرد جریان‌های‌کاری می‌شود و خروجی شامل مدل فرآیندی ساده شده و تعمیر شده‌ای است که در آن با تکنیک‌هایی برخی گام‌ها حذف و یا فشرده شده‌اند. در این رساله، این تکنیک‌ها شامل استفاده از قوانین خطی‌سازی و ادغام می‌شود [48]. هم‌چنین با روشی بر مبنای خوشه‌بندی، ساده‌سازی ترتیب می‌دهیم که در ادامه بیشتر توضیح خواهیم داد. روش برمبنای خوشه‌بندی آن جاهایی را ساده خواهد کرد که نمی‌شود جریان‌کاری را از روی گزارش عملکرد ساده کنیم. به عنوان مثال، شکل ۴-۱ را برای خروجی این روش ساده‌سازی ببینید.



شکل ۴-۱—ساده‌سازی جریان‌کاری. الف- نمودار جریان‌کاری. ب- نمودار ساده شده شکل الف.

شکل ۴-۱ را در نظر بگیرید. در شکل ۴-۱-الف نمودار جریان‌کاری ساده نشده‌ای است که اگر گزارش عملکرد آن را نداشته باشیم نمی‌توانیم ساده‌سازی را با وجود برگشتی که دارد ترتیب دهیم. در این صورت می‌توان با تکیه بر مشاهدات قبلی که داشته‌ایم و بر مبنای خوشه‌بندی مطابق مشابهش عمل کرده و شکل ساده شده ۴-۱-ب را به دست آوریم. با تکیه بر مشاهدات قبلی و مراجعه به گزارش عملکرد معلوم شده‌است که در ۹۰٪ مواقع کاربران بعد از گام D به گام E رفته و پس از آن کار را خاتمه می‌دهند. هم‌چنین در خوشه مربوطه کاربران هیچ وقت پس از گام F بازگشت به عقبی

ندارند. بنابراین، شکل پیشنهادی ۴-۱(ب) به عنوان خروجی ساده‌تر شده در نظر گرفته می‌شود. در این شکل گام H ترکیبی از گام E و F می‌باشد.

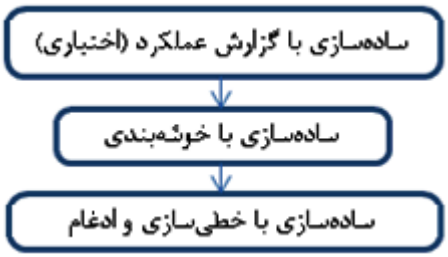
۴ ۲. محدودیت‌های بهینه سازی

محدودیت‌های بهینه سازی شامل موارد زیر است:

۱ -تنوع جریان‌کاری بالاست: با این فرض که الگوها و تنظیمات کلی برای کاربردهای خاص خیلی مفید نیستند، مطالعه جریان‌های‌کاری روز به روز بیشتر به کاربرد وابسته می‌شود. برای برخی حوزه ها مانند جریان‌های‌کاری زیستی، مطالعه گروهی از گره‌ها می‌تواند الگوهای در شبکه را بهتر مشخص کند. اما در کاربردهایی مانند مدیریت شبکه های اجتماعی، افراد بیشتر می‌توانند درباره کل سیستم خاصیت کشف کنندگی داشته باشند. در بخش قبل گفتیم که با توجه به رفتارهای موتیف‌ها می‌توان جریان‌های‌کاری پویا را خوشه‌بندی کرد.

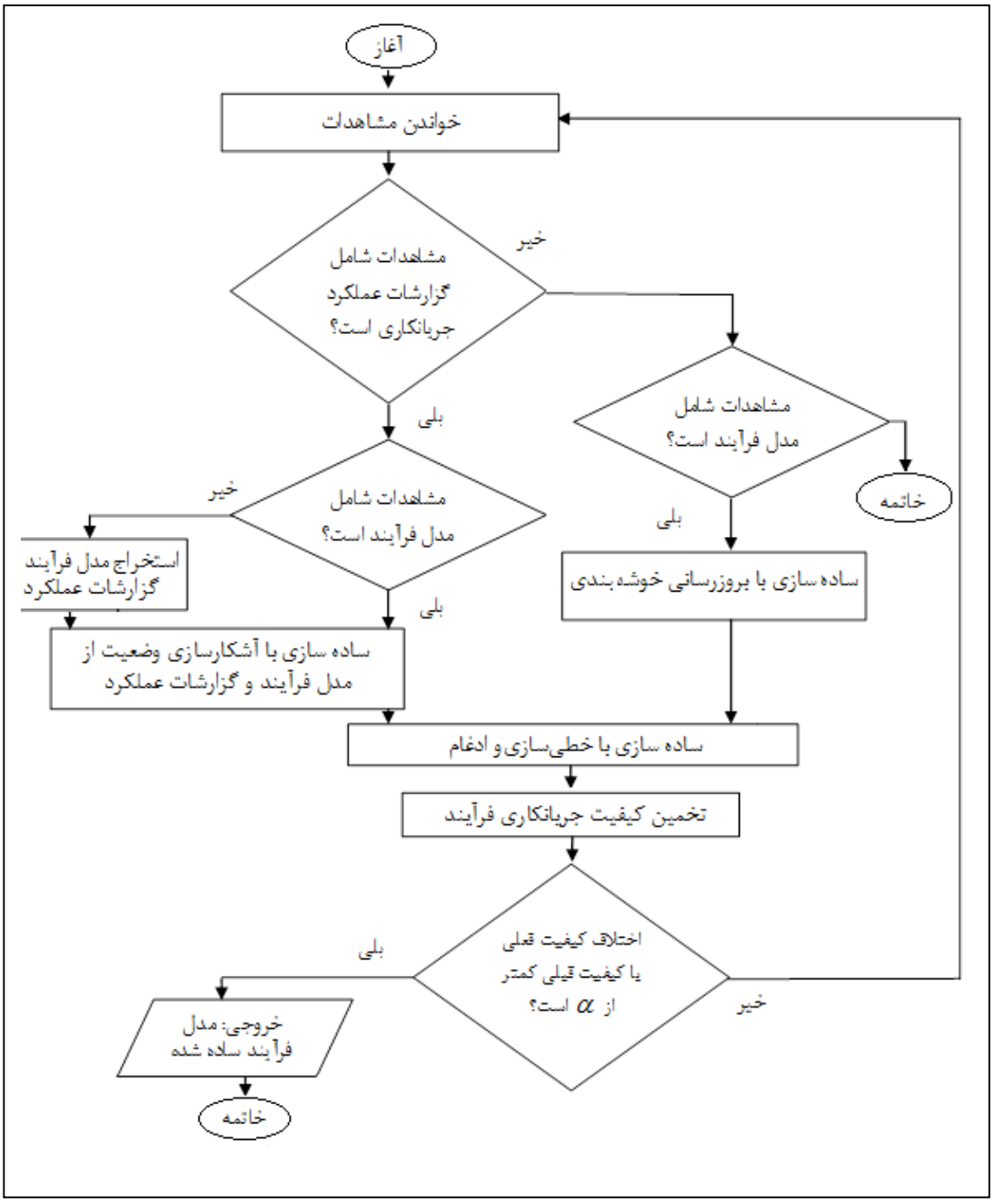
۲ -تغییر پویای ساختار جریان‌کاری: فهم فرآیندهای پویا در جریان‌کاری نیازمند در نظر گرفتن تمام انواع وابستگی‌های ساختاری جریان‌کاری (مثل وابستگی‌های کنترل، داده و موقتی) بین وظایف است [42]. با توجه به این مورد راه‌حل به ساده‌سازی مساله را ارائه می‌دهیم.

برای ساده‌سازی ابتدا از داده‌های گزارش عملکرد جریان‌کاری بهره می‌بریم. سپس خوشه‌بندی ترتیب می‌دهیم. با خوشه‌بندی اگر داده‌های گزارش عملکرد ناقص بود یا گزارش عملکرد نداشتیم می‌توانیم از دانش قبلی جریان‌های‌کاری ساده شده قبلی بهره ببریم. شکل ۴-۲ ترتیب مراحل ساده‌سازی روش ارائه شده را نشان می‌دهد.



شکل ۴-۲– مراحل ساده‌سازی جریان‌کاری

با در نظر گرفتن تکاملی بودن نمودارهای روند کاری خطی‌سازی به صورت شکل ۴-۳ می‌شود.



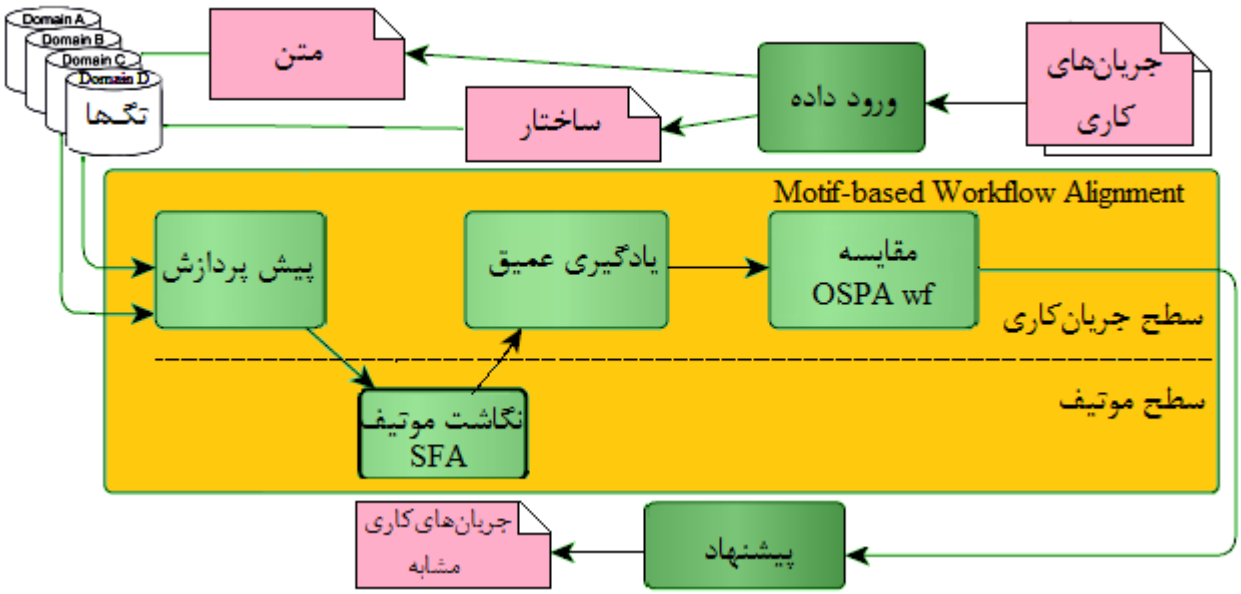
شکل ۴-۳- روند کاری ساده‌سازی جریان‌کاری

برای ساده‌سازی جریان‌کاری با خوشه‌بندی لازم است تحلیل جریان‌کاری صورت بگیرد. در ادامه به بررسی اجمالی روش می‌پردازیم.

۴ ۳. بررسی اجمالی روش

روش ما نشان می‌دهد که چگونه دانش به دست آمده از انباره‌ای از مجموعه‌داده‌هایی که از دامنه‌های مختلف می‌آیند را تنظیم کنیم. الگوریتمی که ما استفاده کرده‌ایم خصیصه‌های مشترک،

یعنی مفاهیم، را به عنوان انتزاعات مفهومی مستقل از دامنه برای گام‌های جریان‌کاری یاد می‌گیرد. سپس بعد چند گام، جریان‌کاری مفهومی‌تری را که با جریان‌کاری مفروض انطباق می‌یابد را پیش-بینی می‌کند. بنابراین، روش ما نمونه‌های گراف جریان‌کاری را در فضای برداری نهفته می‌کند. این روش نهفته‌سازی، بازنمایی موتیف‌ها را از مجموعه بزرگ داده‌های برچسب نخورده با استفاده از مدل مولد یاد می‌گیرد. این روش، مدل مولد را به عنوان مساله‌ای از استخراج مفهوم بین دامنه‌ای از جریان‌های‌کاری مختلف می‌بینند. در نتیجه، می‌توان بهترین ترتیب موتیف‌های جریان‌کاری را در ترتیب ترجیحی مبتنی بر مجموعه خصیصه‌های یادگرفته شده توسط DNN پیشنهاد داد. شکل ۴-۴ را برای بررسی اجمالی روش ارائه شده می‌توان ملاحظه کرد.



شکل ۴-۴-روش اعمال شده برای بهبود جریان‌های‌کاری

روش ما باید بتواند توصیفات مدل‌های جریان‌کاری پیچیده را به دست بگیرد. بنابراین از فرمت ISA-Tab استفاده می‌کند فرمت ISA-Tab برای مدیریت جریان‌های‌کاری علمی ارائه شده است. این فرمت یک مورد اجرایی را با چندین تجزیه و تحلیل در کنار هم قرار می‌دهد [62]. برای سایر جریان-های‌کاری در سایر دامنه‌ها، مکانیسم ثبت رخداد مبتنی بر XML در فرمت XES استفاده شده است [74]. همان‌طور که در شکل ۴-۴ می‌بینید جریان‌های‌کاری ورودی که می‌توانند به صورت نموداری یا به صورت گزارش عملکرد باشند، به عنوان ورودی دریافت می‌شوند. این جریان‌های‌کاری، دامنه‌ها و حوزه‌های کاربرد مختلفی دارند، که در شکل به صورت دامنه A-D نمایش داده شده است. در اولین مرحله ابتدایی‌ترین کاری که صورت می‌گیرد این است که تمامی تگ‌های کلیه جریان‌های‌کاری از دامنه‌های مختلف در کیسه‌ای از جریان‌های‌کاری قرار می‌گیرند. سپس در گام پیش‌پردازش، تگ‌هایی را که تعداد تکرار کمی در میان سایر دارند با درنظر گرفتن یک حد آستانه حذف می‌کنیم. این کار باعث می‌شود که حجم پردازش‌ها بعدی کم‌تر شود. پس از آن در سطح موتیف به بررسی تگ‌ها می-

پردازیم و الگوریتم SFA را روی آن‌ها اعمال می‌کنیم تا بتوانیم تمام تگ‌های مربوط به هر حوزه و دامنه را خوشه‌بندی کنیم. این کار امکان استفاده از جنبه‌های مختلف یک حوزه را فراهم می‌کند. در مرحله بعدی، همان‌طور که در شکل ۴-۴ دیده می‌شود، در سطح جریان‌کاری الگوریتم یادگیری عمیق اعمال می‌شود تا بتوانیم چگونگی ارتباط میان تگ‌هایی را بیابیم که در الگوریتم SFA حوزه-بندی شده‌اند. گام مقایسه OSPA مربوط به زمانی است که جریان کاری جدیدی برای بهبود به سیستم وارد شده‌است. در این گام با مقایسه نتیجه OSPA می‌توان از خوشه‌بندی که قبلا صورت گرفته و از نتیجه ارتباطاتی که الگوریتم یادگیری عمیق به دست داده، پیشنهادی را برای به دست آوردن جریان‌کاری ساده‌تر ارائه کرد. به طور خلاصه، گام‌های ارزیابی شباهت جریان‌کاری در دامنه-های مختلف و بهبود جریان‌کاری عبارتند از:

الگوریتم: بهبود جریان‌کاری

ورودی: گزارش جریان‌کاری به صورت L(D, t) که در آن t نوع می‌باشد و D دامنه است. **خروجی:** جریان‌کاری بهبودیافته

روش: شامل انجام مراحل زیر می‌شود:

۱ - بارگذاری مجموعه داده برای دامنه‌های مبدا و مقصد

۲ - اعمال SFA

حلقه: برای هر مبدا و مقصد

a. حذف داده‌های پرت با آستانه مفروض
b. نسبت دادن نوع برچسب‌ها به گام‌های جریان‌های‌کاری
c. جداسازی داده در دو بخش DS و DI
   i. انتخاب خصیصه‌های محوری که ارتباط پرمعنی با سایر خصیصه‌ها دارد
   ii. انتخاب خصیصه‌های خاص که ارتباطی با سایر خصیصه‌ها ندارد
d. خوشه‌بندی برای محاسبه ماتریس هم‌رخدادی که نشان‌دهنده ارتباطات بین داده‌های مختلف مجموعه‌داده است
e. محاسبه گراف دووجهی که دارای روابط بین نقاط مختلف در مجموعه‌داده است
f. پیدا کردن مقادیر k بزرگترین Eigen (ساخت ماتریس U)
g. جاسازی خصیصه‌های منحصربفرد موتیف‌های جریان‌کاری در ماتریس U

پایان حلقه

جستجو بهترین انطباق مقادیر k بزرگترین Eigen از جریان‌کاری مفروض با سایرین

۳ - اعمال مقایسه OSPA:

مقایسه ماتریس‌ها U برای یافتن نماینده خوشه؛ پیشنهاد نزدیک‌ترین موتیف جریان‌کاری بعنوان نماینده خوشه برای یک جریان‌کاری مفروض

۴ - اعمال DNN:

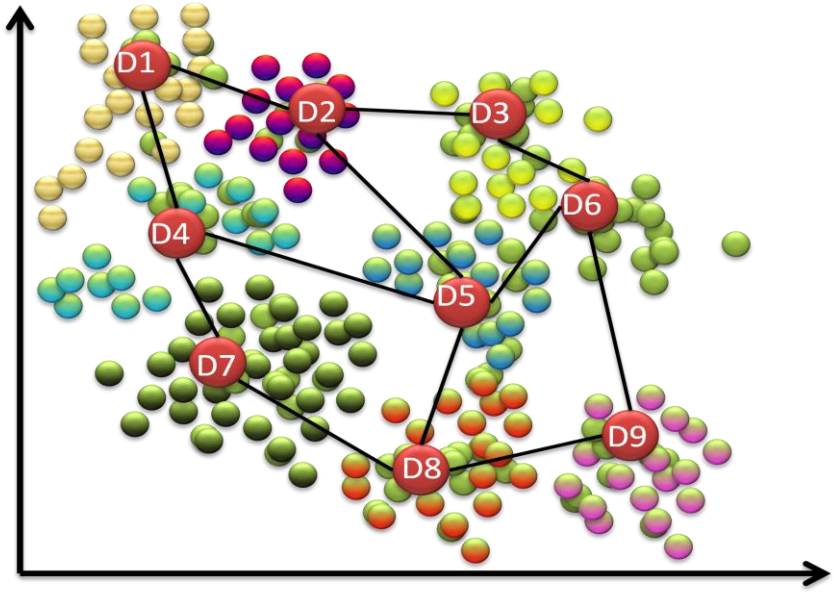
ساده‌سازی جریان‌کاری مفروض بر اساس ارتباطات نماینده خوشه

۴ ۴. اعمال الگوریتم SFA در جریان‌های‌کاری

الگوریتم SFA امکان استخراج خصیصه‌های جریان‌کاری را فراهم می‌کند. مهم‌ترین خصیصه‌های جریان کاری که در این رساله مدنظر بوده‌اند، تگ‌ها و موتیف‌ها هستند. در الگوریتم SFA تگ‌هایی را که به صورت الگویی مشترک هستند به عنوان موتیف در نظر می‌گیریم. آن‌ها را خوشه‌بندی می‌کنیم. و در نتیجه خوشه‌هایی با دامنه کاربردها مختلف به دست می‌آوریم. برای این کار تگ‌هایی که با شکلی از ارتباطات مشخص در کنار هم قرار گرفته‌اند را به وسیله میزان ارزیابی شباهت آن‌ها مورد شمارش قرار می‌دهیم تا پرتکرارترین آن‌ها در هر خوشه یافت شوند. بدین صورت هر خوشه شکل می‌گیرد.

۴-۴-۱. تحلیل جریان‌های‌کاری در سطح موتیف

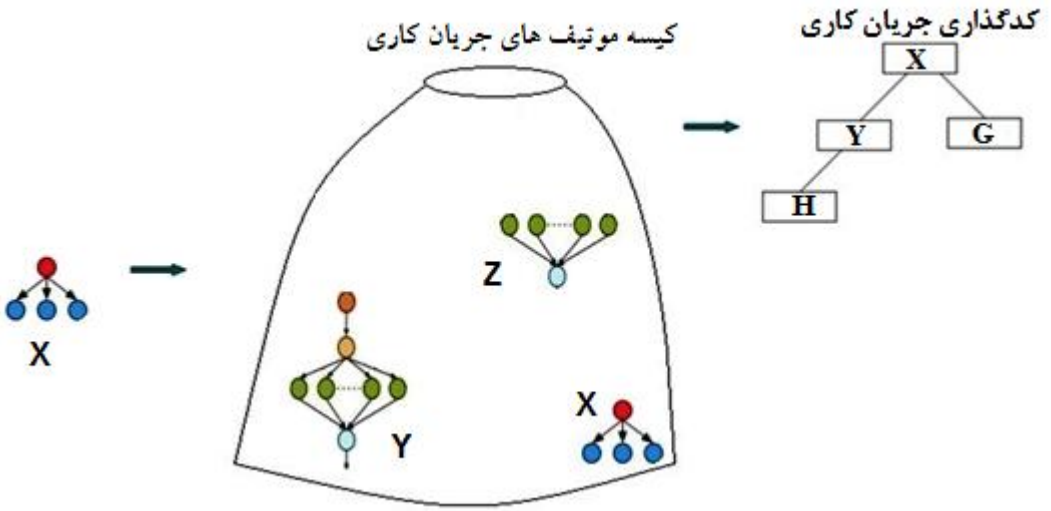
بعد از این که ترمیم و ساده‌سازی مساله با محدودیت‌های بخش ۴-۲ براساس محتوای جریان‌کاری و با داده‌های گزارش عملکرد حل شد، لازم است راه حل نمونه را برای جریان‌های‌کاری مشابه در نظر بگیریم. بدین منظور می‌توان جریان‌های‌کاری مشابه را که قبلا برای آن‌ها ساده سازی ترتیب داده‌ایم در یک دسته قرار دهیم. بدین منظور لازم است یک سری داده با توجه به فرمت جریان‌های‌کاری استخراج کنیم. خوشه‌بندی ساختاری جریان‌های‌کاری بیشترین پیچیدگی زمانی را دارد. ممکن است کدگذاری جریان‌های‌کاری برای وجود وظایف با ۱، برای عدم وجود وظایف با صفر صورت بگیرد و روش خوبی برای ورودی خوشه‌بند باشد [41]. در مقاله [41] با توصیف برداری جریان‌های‌کاری ماتریس‌های اسپارس حاصل شده است و در نتیجه کیفیت خوشه‌بندی جریان‌های‌کاری پایین آمده است. این در صورتی است که اگر از توصیف‌گرهای مبتنی بر مجموعه استفاده کنیم می‌توان اندازه جریان‌های‌کاری را متغیر در نظر گرفت. ضمن این که می‌توان از تئوری مبتنی بر مقایسه مجموعه‌ها برای ارزیابی و تحلیل آن‌ها بهره برد [46]. بدین ترتیب دیگر فرض‌های محاسبات در فضای برداری را نخواهیم داشت. البته با توجه به این‌که می‌توان در یک رده بالاتر کدگذاری جریان‌کاری را براساس موتیف‌های تشکیل دهنده آن ترتیب داد، در این صورت می‌توان نحوه خوشه‌بندی را به صورت شکل ۴-۴ در نظر گرفت.



شکل ۴-۵- خوشه‌بندی

در شکل ۴-۵، D_1-D_n دامنه‌های جریان‌های‌کاری را نشان می‌دهند . موتیف‌های ویژه و خاص باعث می‌شوند که خوشه‌ها متمرکزتر و جمع‌تر شوند. ولی پایه اصلی خوشه‌بندی، موتیف‌های محوری و مشترک می‌باشند که پرتکرار هستند.

در شکل ۴-۵ برای تست از یک نمودار جریان‌کاری نمونه استفاده شده است. از کیسه‌ای می‌توان استفاده کرد که شامل مجموعه‌ای از موتیف‌های جریان‌کاری استخراج شده از چند نمونه جریان‌کاری قبلا ساده شده باشد. سپس ساختار جریان‌کاری نمونه را با این کیسه شامل موتیف‌های جریان‌کاری برای تست تحلیل می‌کنیم. برای این کار با یک روش دسته‌بندی نمونه‌های مشابه جریان‌کاری را پیدا می‌کنیم که می‌خواهیم ساده کنیم. سپس مشابه‌ترین جریان‌کاری را از نظر ساختار انتخاب کرده و براساس ساختار بررسی می‌کنیم سادگی نمونه تست به آن نزدیک‌تر شود.



شکل ۴-۶- تست نمونه

در شکل ۴-۶ کیسه موتیف‌های مشترک از موتیف‌های X، Y و Z پر شده است. در این شکل ملاحظه می‌شود که یک جریان‌کاری X برای ساده‌سازی در نظر گرفته‌ایم که قبلا مشابه آن در کیسه در شکل ۴-۶ ذخیره شده است.

مزیت خوشه‌بندی جریان‌های‌کاری زمانی مشخص می‌شود که نمودار جریان‌کاری را داریم ولی گزارش‌های عملکرد را نداریم. در این حالت مزیت استفاده از این دسته بندی فراهم شدن امکان حدس زدن برای نمونه های مشابه است. بدین ترتیب می‌توان مشخص کرد که کدام وظایف قبل از وظایف دیگر می‌آیند. سپس مشخص کنیم منبع اجرایی هر وظیفه چه چیزی باشد.

برای تحلیل جریان‌کاری می‌توان برای هر موتیف خصیصه‌هایی مثل سطح، ورودی، خروجی، نوع، نقش، و چگالی استخراج کرد. در نظر گرفتن سطح و در نظر گرفتن نوع برای موتیف کمک می‌کند پردازش مفهومی جریان‌کاری را نیز در نظر بگیریم. منظور از بررسی مفهومی، در نظر گرفتن جایگاه و نقش موتیف در سطحش است. یعنی مشخص کردن این‌که هر بخش از جریان‌کاری چه رفتاری را نشان می‌دهد. برای این کار دو عمل الزامی است. یکی سطح‌بندی موتیف و دیگری خوشه‌بندی آن-هاست. تنها کاری که تاکنون در این زمینه انجام شده است، به بررسی دو به دوی رفتارهای گره‌ها پرداخته است [28]. سایر کارها تاکنون از نظر مفهومی صرفا به اختصاص دادن یک نوع به گره‌ها اکتفا کرده‌اند. فرض کنیم در یک جریان‌کاری یک موتیف درختی داریم که بعد از یک موتیف اتصال‌دهنده آمده است. در این صورت موتیف سطح یک درخت با یک پیوند به موتیف سطح ۲ اتصال دهنده

متصل شده است. این اتصال می‌تواند نشان‌گر این باشد که یک گره خدمات‌رسان یک گـره جداکننـده خدمات‌رسان دیگر دارد.

بنابراین روش ارائه شده برای تحلیل جریان‌کاری در سطح موتیف به صورت زیر می‌شود:

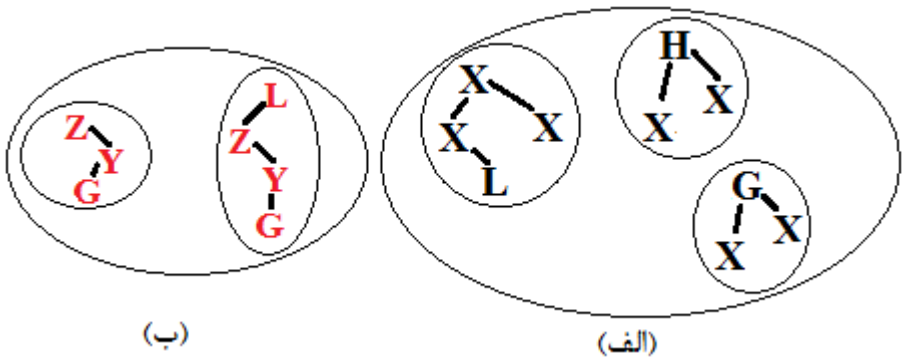
۱ -ورودی شامل تعدادی فرآیند می‌شود که مدل گرافیکی جریان‌کاری آن‌ها را همراه با گزارش عملکرد[1] برای برخی از آن‌ها داریم.

۲ -یک فرآیند از تعدادی موتیف (زیرگراف) تشکیل می‌شود.

۳ -یک موتیف یک زیرگراف پایه است که موارد مشابه آن در فرآیندهای دیگر تکرار می‌شود. این موتیف می‌تواند ساختاری باشد مثل درخت ستاره‌ای که در آن از یک گره به چند گره انشعاب داریم، یا برعکس آن از چند گره انشعاب می‌گیریم و به یک گره می‌دهیم. هم‌چنین موتیف می‌تواند مفهومی و فقط بخشی با تکرار فراوان باشد.

۴ -اگر مثلا ۱۰ فرآیند داشته باشیم، بعضی از این فرآیندها موتیف‌های مشترک دارند.

۵ -فرآیندها را براساس داشتن موتیف‌های مشترک خوشه‌بندی می‌کنیم (مثل شکل ۴-۶).



شکل ۴-۷- خوشه‌بندی فرآیندها براساس موتیف‌های مشترک

در شکل (۴-۷-الف)، نتیجه خوشه بندی ۳ فرآیندی است که در درون خـود موتیـف مشـترک X دارند. یعنی موتیف X در هر سه فرآیند تکرار می‌شود. در شکل (۴-۷-ب) موتیـف مشـترک Y در دو فرآیند تکرار می‌شوند.

۶ -اگر فقط برای یکی از فرآیندهایی که آن‌ها را خوشه‌بندی کرده‌ایم یک گزارش عملکرد داشـته باشیم، با استفاده از تحلیل آماری گزارش عملکردها، می‌توان برای سایر فرآیندهای دیگـر نیـز که گزارش عملکرد نداریم ساده‌سازی ترتیب داد. مثلا ممکن است یک موتیف که شامل چنـد گره می‌شود با یک گره جایگزین شود و یا یک و یا دو موتیف پشت سرهم با یـک موتیـف یـا یک گره جایگزین می‌شوند.

[1] - Event Log

۷ -نحوه ادغام یا تغییرات موتیف‌ها را در یک فرآیند ساده‌شده می‌توانیم به فرآیندهای دیگر تعمیم دهیم. بدون آن که در این فرآیندها از گزارش عملکردها استفاده کنیم.

۴-۴-۲. تنظیم پارامترهای مساله برای اعمال SFA

با توجه به این‌که قصد داریم استخراج خصیصه‌های جریان‌های‌کاری از چند حوزه مختلف صورت بگیرد، ابتدا برخی تعاریف را ارائه می‌کنیم.

تعریف ۱ (دامنه): دامنه D کلاسی از موجودیت‌ها در یک فضای مفهومی را نشان می‌دهد. به عنوان مثال، انواع مختلف جریان‌کاری مثل جریان‌های‌کاری در حوزه تجارت الکترونیک، بیمارستان، و بیوانفورماتیک، می‌توانند به عنوان دامنه‌های مختلف در نظر گرفته شوند. به عنوان مثالی دیگر، علوم کامپیوتری، ریاضیات و فیزیک نیز هر کدام می‌توانند به عنوان دامنه جریان‌های‌کاری در نظر گرفته شوند.

تعریف ۲ (نوع): با توجه به یک دامنه خاص D، نوع جریان‌کاری با دنباله‌ای از برچسب‌ها $t_1t_2..t_n$ تعریف شده است که در آن هر t_i نوع برچسب جریان‌کاری برای دامنه مفروض D را مشخص می‌کند. بر اساس تعاریف بالا اکنون مساله تعریف می‌شود.

تعریف ۳ (تگ DI): هر تگ DI به عنوان موتیف مستقل و مشترکی در نظر گرفته می‌شود که بیشترین تکرار را در دامنه D دارد. تگ DI به عنوان داده محوری در حین ارزیابی شباهت میان جریان‌های‌کاری مورد بررسی قرار می‌گیرد.

تعریف ۴ (تگ DS): هر تگ DS به عنوان موتیف ویژه در نظر گرفته می‌شود که در حین محاسبه به صورت تگ‌های متمم DI در نظر گرفته می‌شوند. تگ DS به عنوان داده خاص و ویژه در حین ارزیابی شباهت میان جریان‌های‌کاری مورد بررسی قرار می‌گیرد. این تگ‌ها در جریان‌کاری‌ها گرچه به تعداد کمتری مورد استفاده قرار می‌گیرند اما نقشی ویژه و مهم در جریان‌کاری دارند.

تعریف مساله (ارزیابی شباهت دامنه متقابل[1])

با داشتن مجموعه داده‌های برچسب‌دار (تگ‌ها) از دامنه مبدا مشخص مدل را آموزش می‌دهیم تا برخی از داده‌های بدون برچسب از دامنه مقصد را دسته‌بندی کنیم. در ابتدا برای هر مجموعه داده تابع نگاشتی بر هر مورد اجرایی که در گزارش رویداد ارائه شده است، اعمال می‌شود. این کار موتیف-های جریان‌کاری را کشف می‌کند. موتیف‌های جریان‌کاری که اغلب در حوزه‌های جریان‌کاری اتفاق می‌افتند، می‌توانند به عنوان بخشی از جریان‌های‌کاری به کار آیند. آن‌ها می‌توانند پس از تجزیه و

[1] - Cross-Domain

تحلیل دامنه‌های مختلف جریان‌کاری استخراج شوند. موتیف‌های مستقل جریان کاری که آن‌ها را DI نشان می‌دهیم در بین دامنه‌های مختلف وجود دارند. این موتیف‌ها نقش محوری دارند. زیرا جایگاه نسبتا ثابتی را در ساختار جریان‌کاری دارا هستند، و بسیاری از جریان‌های‌کاری آن‌ها را دارند. موتیف-های دیگری نیز وجود دارند که در بین دامنه‌های مختلف وجود ندارند. فقط گاهی در یک دامنه بخصوص دیده می‌شوند. این موتیف‌ها به علت تفاوت دامنه‌شان جایگاه‌های ثابتی ندارند، و فقط در برخی جریان‌های‌کاری مشاهده می‌شوند. که آن‌ها را DS یا الگوهای خاص می‌نامیم. با بوجود آمدن پلی میان تگ‌های DS و DI می‌توان فاصله موجود بین دامنه‌های متفاوت را پوشش داد. الگوریتم SFA با تشکیل گرافی دووجهی مدلی را بین تگ‌های DS و DI شکل می‌دهد. در این مدل مفهومی اگر دو تگ DS ارتباط هم‌رخدادی بیشتری با DIها در گراف داشته باشند، تمایل دارند بیشتر با هم، هم‌تراز شوند. هم‌چنین اگر دو تگ DI اتصالات بیشتری با DS در گراف داشته باشد، آن‌ها با احتمال بیشتری هم‌تراز می‌شوند. با یادگیری نمونه‌های مدل می‌توان پس از برچسب‌گذاری تگ‌ها براساس DNN، آن‌ها را در چینشی بهبود یافته کنار هم قرار داد و جریان‌کاری G را بهبود داد. برای دستیابی به اهداف، کار حاضر گام‌های زیر را دنبال می‌کند:

۱ -یادگیری خصیصه‌های مبدا و مقصد،
۲ -استفاده از خصیصه‌های آموخته‌شده برای پوشش نقایص کار،
۳ -آموزش مدل از بازنمایی‌های جدید جریان‌های‌کاری با برچسب‌های مربوطه براساس استخراج مفهوم از دامنه‌های مختلف،
۴ -جستجو در میان مدل‌های جریان‌کاری آموخته شده،
۵ -تطبیق جریان‌کاری پیچیده نمونه براساس مدل جریان‌کاری شناسایی شده

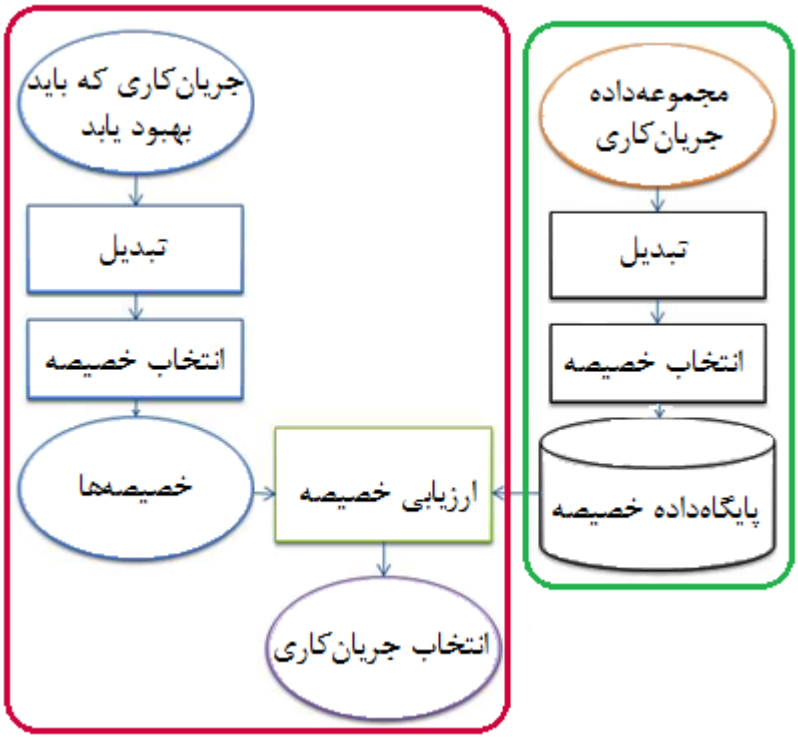
۴ ۴ ۳. ارزیابی شباهت در جریان‌های‌کاری با یادگیری انتقالی

کار حاضر تگ‌ها را به صورت خودکار با خوشه‌بندی طیفی و یادگیری انتقالی در بین دامنه‌های مختلف به یکدیگر مرتبط می‌کند. هدف یادگیری انتقالی، انتقال دانش بین موتیف‌های دامنه‌های مختلف جریان‌کاری است. روش حاضر برای انطباق دامنه‌ها از استخراج مفهوم در بین دامنه‌های متقابل استفاده می‌کند. در روش پیشنهادی برای پر کردن شکاف اطلاعات بین دامنه‌ها، الگوریتم SFA تگ‌های DS را از حوزه‌های مختلف به خوشه‌های متحدی از DI ارتباط می‌دهد. به این ترتیب، خوشه‌ها می‌توانند مورد استفاده قرار گیرند تا شکاف بین تگ‌های DS دامنه‌ها را کاهش دهند، که می‌تواند برای آموزش دقیق دسته‌بندها در حوزه هدف مورد استفاده قرار گیرد. در مقایسه با رویکردهای قبلی، SFA می‌تواند نمایشی قوی برای داده‌های متقابل دامنه را با استفاده کامل از رابطه بین تگ‌های DI و DS (موتیف‌های جریان‌کاری) از طریق هم خوشه بندی آن‌ها در فضای پنهان

مشترک کشف کند [71]. از این رو، آزمایشات با روشی از خوشه بندی برای ارزیابی شباهت جریان-های‌کاری به نام خوشه بندی طیفی صورت می‌گیرد. علاوه بر این، این رساله کسری از مشاهدات را حذف می‌کند که توسط مکانیزمی متفاوت [72] تولید شده‌اند. اگر داده‌های پرت واقعی از مجموعه-داده‌ها حذف نشده باشند، نتایج را خراب می‌کنند.

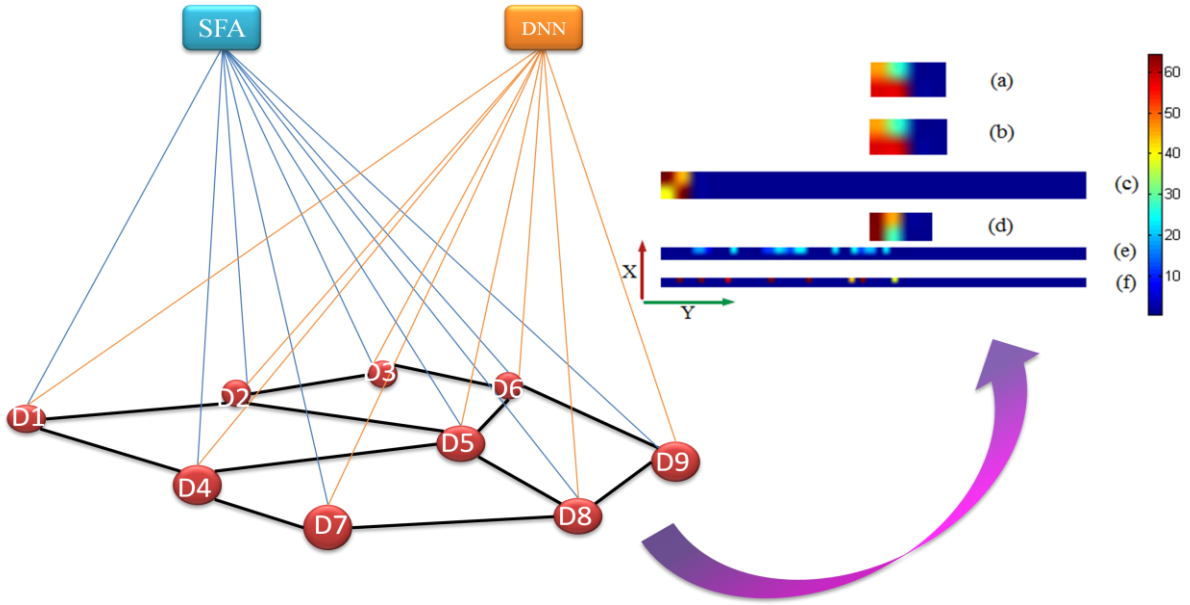
روش SFA برای استخراج مفهوم از تجزیه Eigen ماتریس هم‌رخدادی استفاده می‌کند. پس از آن-که ماتریس هم‌رخدادی که ارتباطات آماری میان DS و DI را به وجود می‌آورد. پس از آن‌که ماتریس هم‌رخدادی شکل گرفت، با یافتن K مقدار Eigen ماتریسی حاصل می‌شود که آن را U می‌نامیم. این ماتریس U امکان مقایسه جریان‌های‌کاری را فراهم می‌کند.

با بهبود جریان‌کاری براساس ارزیابی شباهت می‌توان از دانش قبلی جریان‌های‌کاری پیشین بهره برد؛ یعنی بعد از انطباق و بهبود جریان‌کاری، می‌توان از شیوه بهبود یافتن برای استفاده مجدد در بهبود جریان‌کاری مشابه استفاده کرد. خصیصه‌های مهم جریان‌های‌کاری در استفاده مجدد بالقوه‌شان موثر هستند [67]. ما از ارزیابی شباهت جریان‌کاری برای خودکار کردن بهبود جریان‌کاری استفاده می‌کنیم. برای این کار، خصیصه‌ها (تگ‌های DS و DI) مبتنی بر SFA استخراج می‌شوند [68]. روش SFA برای استخراج و نمایش آماری خصیصه‌ها به کار می‌رود. این روش که نوعی یادگیری انتقالی است برای خودکار کردن یادگیری خصیصه‌ها در کل جریان‌های‌کاری، از دامنه‌های متفاوت و از طبیعت‌های مختلف به کار می‌رود. یادگیری انتقالی زمانی که دسته‌بندی در حوزه خاصی مدنظر باشد، ولی فقط داده‌های آموزش در حوزه دیگری در اختیار باشد، بسیار قوی عمل می‌کند. حتی با وجودی که ممکن است دامنه‌ها خصیصه‌های متفاوتی داشته باشند و یا حتی توزیع داده متفاوتی را دنبال کنند [68]. جریان‌های‌کاری مشابه که قبلا بهبود یافته‌اند می‌توانند در یک پایگاه‌داده به یک گروه تقسیم شوند (شکل ۴-۸). در نهایت، ارزیابی شباهت با خصیصه‌های استخراج شده از جریان‌کاری که باید بهبود داده شود و از پایگاه داده خصیصه صورت می‌گیرد تا از بهترین جریان‌های‌کاری یافته شده مجددا استفاده شود. مبتنی بر ترتیب‌های استخراج شده توسط روشی مبتنی بر DNN جریان‌کاری بهبود می‌یابد. داده‌های مبدا مطالعه حاضر عبارتند از وقایع ثبت شده جریان‌کاری که نشان‌دهنده موردهای اجرایی جریان‌های‌کاری است.



شکل ۴-۸- ارزیابی شباهت

رساله حاضر مربوط به یادگیری موتیف‌هاست، حتی اگر داده کمی در اختیار داشته باشیم. این رساله فرض می‌کند که خصیصه‌های سطح بالایی وجود دارند که می‌توانند به آموزش مدل دسته بند کمک کنند. هدف ما در این مطالعه این است که خصیصه‌های DS غیر محوری را از دامنه‌های مختلف با تطبیق حوزه جریان‌کاری تنظیم کنیم.



شکل ۴-۹- تشکیل نماینده خوشه‌ها براساس Eigen Value (ماتریس وزن دار U)

برای ساخت خصیصه‌های سطح بالا روش‌های کدگذاری تُنُک (SC[1]) و یادگیری عمیق [69] استفاده می‌شود. این رساله از هر دو روش سود می‌برد. روش SC یک روش ساخت خصیصه بدون نظارت است و روش یادگیری عمیق به طور مستقیم از ورودی‌های خام استفاده می‌کند. روش SC توابع پایه‌ای را یاد می‌گیرد که خصیصه‌های سطح بالا را در داده مفروض یا حتی داده ورودی بدون برچسب بدست می‌گیرد [70]. ایده اصلی کدگذاری تنک این است که بردار خصیصه به عنوان ترکیبی خطی از پایه‌های کمی از مخزن داده قبلا تعریف شده است. بنابراین شامل مفهوم خلوت بودن می‌شود.

با تگ‌های نسبت داده شده به گام‌های جریان‌کاری در یک مخزن داده به صورت کیسه‌ای از تگ‌ها رفتار می‌شود و شباهت جریان‌کاری می‌تواند مشابه روش‌های شباهت میان کلمات در کیسه کلمات حساب شود [57]. تفاوت کار حاضر با مقاله [57] این است که این روش تگ‌ها را بدون نیاز به پیش گزینش خاصی توسط طراح جریان‌کاری تنظیم می‌کند، و این روش تگ‌های گام‌های جریان‌کاری را با بیشترین تکرار به عنوان موتیف‌های جریان‌کاری بین دامنه‌های مختلف انتخاب می‌کند.

در SFA ساخت فضای مفهومی (همچنین فضای جاسازی شده یا پنهان نامیده می‌شود) با استفاده از LSA انجام می‌شود. روش LSA روشی آماری ساده است که داده‌ها را بر روی یک فضای مفهومی به فضای با بعد پایین‌تر نگاشت می‌کند. تخمین همبستگی‌ها به اندازه‌گیری وابستگی‌ها بین داده‌های DS و DI کمک می‌کند تا خصیصه‌های محوری خوبی انتخاب کنیم. روش‌هایی برای تخمین همبستگی‌ها بین خصیصه‌های DS و محوری وجود دارند. نمونه‌هایی از این روش‌ها عبارتند از یادگیری ساختارهای مرتبط (SCL) و SFA [71]. روش SCL مشابه SFA می‌باشد با این تفاوت که خصیصه‌های محوری را به صورت هیوریستیک انتخاب می‌کند و تضمینی برای این‌که این انتخاب صحیح است ارائه نمی‌کند. این رساله یک الگوریتم خوشه‌بندی طیفی را، مشابه کار [73] بر روی گراف دو وجهی برای هماهنگ‌سازی خصیصه‌های DS و DI تطبیق می‌دهد. اگر خصیصه‌های DS با خصیصه‌های DI در برخی از بردارهای خصیصه هم-برخوردی داشته باشند، بین دو خصیصه در گراف دووجهی لبه ایجاد می‌شود. در نهایت، خوشه‌بندی طیفی بر گراف دووجهی انجام می‌شود که دو مجموعه از خصیصه‌ها را نشان می‌دهد. اجزای اصلی K استخراج شده توسط LSA می‌توانند به صورت خودکار خوشه‌بندی را به زیرفضایی که توسط اجزای اصلی K اندازه‌گیری شده اعمال کنند. این بدان

[1] - Sparse Coding

معنی است که تابع نگاشتی که از اجزای اصلی K ساخته شده می‌تواند داده‌های اصلی را خوشه‌بندی کند و آن‌ها را به فضای جدیدی نگاشت دهد که توسط خوشه‌ها به طور همزمان اندازه‌گیری می‌شود.

۴ ۴ ۴. توضیحات الگوریتمیک SFA

در این رساله الگوریتم SFA را به صورت ۷ گام اصلی پیاده‌سازی کرده‌ایم:

۱ - خواندن مجموعه‌داده با درنظر گرفتن حوزه مبدا و مقصد.

۲ -محاسبه نمره وابستگی و حذف داده‌ها پراکنده.

۳ -جداسازی داده‌ها مجموعه‌داده در دو بخش تگ‌های مرتبط با هم و تگ‌های بی‌ارتباط با هم بر اساس رابطه ارتباط متقابل.

- یافتن تگ‌ها خصیصه‌های محوری مرتبط که ارتباط معنایی و کاربردی با سایر خصیصه‌ها دارند.
- یافتن تگ‌ها خصیصه‌های خاص ویژه که ارتباطی با سایر خصیصه‌ها ندارند.

۴ -محاسبه ماتریس co-occurrence که نشان‌دهنده میزان ارتباط موجود میان تگ ها در جریان‌کاری‌ها مختلف در مجموعه‌داده است.

۵ -تشکیل گراف دو وجهی $DS \times DI$ که در بردارنده وجود ارتباط میان تگ‌ها مختلف مجموعه‌داده است.

۶ -یافتن مقادیر K-largest-EigenValue: از روی گراف دو وجهی، ماتریس Affinity ساخته میشود. که با ضرب شدن در ماتریس قطری مقادیر Affinity، ماتریس U به دست می‌آید که ماتریس پایه در بدست آوردن K-largest-EigenValue می‌باشد. از U در دسته‌بندی و مقایسه OSPA استفاده می‌کنیم.

۷ -پس از آن دسته‌بندی انجام می‌شود.

برای اندازه‌گیری این‌که چه مقدار تگ‌ها بیشتر از انتظار رخ می‌دهند، اگر آن‌ها مستقل باشند، رابطه (۴-۱) استفاده می‌شود.

$$f(s,t) = \log\left(\frac{\frac{c(s,t)}{N}}{\frac{\sum_{i=1}^{n} c(i,t)}{N} \times \frac{\sum_{j=1}^{m} c(s,j)}{N}}\right) \qquad (۴-۱)$$

که در آن f(s,t) اطلاعات متقابل (PMI) بین دو تگ s و t است و c(s,t) تعداد موتیف‌ها را نشان می‌دهد که در آن دو تگ s و t هم‌برخوردی دارند، و n و m به ترتیب تعداد کلی تگ‌های s و t را نشان می‌دهد، و

$$N = \sum_{i=1}^{n} \sum_{j=1}^{m} c(i, j)$$

سپس برای دو تگ s و v (توسط بردارهای خصیصه s و v نشان داده می‌شوند)، وابستگی τ (v, s) از تگ v به تگ s به صورت رابطه (۴-۲) محاسبه می‌شود [73].

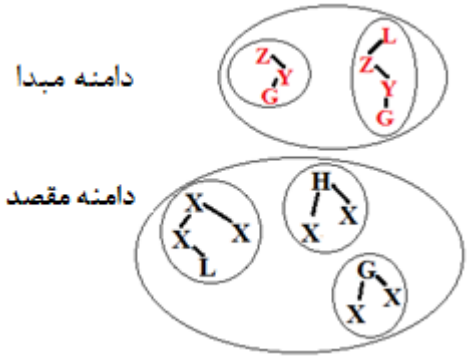
$$\tau(v, s) = \frac{\sum_{t \in \{x | f(v, x) > 0\}} f(s, t)}{\sum_{t \in \{x | f(s, x) > 0\}} f(s, t)} \qquad (۴-۲)$$

نمره وابستگی τ (v, s) می‌تواند به عنوان خصیصه‌های شاخص وزنی PMI از تگ هایی که با تگ v به اشتراک گذاشته می‌شود، تفسیر شود. توجه کنید که مقادیر PMI می‌توانند در عمل حتی بعد از تخفیف برای وقایع نادر منفی شوند. برای جلوگیری از مقادیر PMI منفی فقط وزن‌های مثبت در رابطه (۴-۲) در نظر گرفته می‌شوند [73]. استفاده از رابطه (۴-۲) برای غلبه بر تحت تاثیر قرارگرفتن از عناصر و خصیصه‌های نادر است [73]. برای محاسبه مقادیر PMI در بردار خصیصه، اطلاعات مربوط به هم‌رخدادی بین چندین تگ باید ذخیره شود. برای این کار یک ماتریس واژگانی از مجموعه‌ای بزرگ از تگ‌های DS و DI می‌توان استفاده کرد. به طور خاص، می‌توان از موتیف‌های جریان‌کاری اجتناب کرد که احتمالا دارای نمره‌های کمتری از رابطه هستند و بنابراین احتمال کمتری دارند تا همسایه یک تگ s مفروض با استفاده از تکنیک‌های محاسبه شباهت بردار تقریبی باشند.

در این رساله در پیاده‌سازی رابطه نمره وابستگی مقادیر وزنی (رابطه (۴-۲)) مقادیری که از مقدار حد آستانه کمتر باشند، را برابر صفر قرار می‌دهیم. زیرا نشان‌دهنده این است که ارتباط بین دو تگ، ارتباط ناچیزی است و از اتصال کمتری در ماتریس هم‌برخوردی دارند. این حد آستانه به صورت دستی قابل تنظیم است و در این رساله آن را $\theta=0.01$ قرار می‌دهیم. بدین ترتیب داده‌های پرت[1] حذف می‌شوند. برای این منظور، روش ما دو مرحله حذف داده‌های پرت دارد. ابتدا، آن داده‌های پرت را با اعمال آستانه‌ای به دامنه‌های داده مبدا حذف می‌کند. این آستانه‌ای استفاده می‌شود تا مقدار اتصالات را محدود کند.

[1] - Outliers

پس از اعمال آستانه، تبدیل یک ماتریس هم‌رخدادی تگ-جریان‌کاری ارتباط بین تگ‌های مختلف در داده‌ها را نشان می‌دهد. یک ماتریس U از ارتباطات تگ جریان‌کاری از دامنه‌های مبدا و مقصد به عنوان یک ماتریس $DS \times DI$ با یافتن مقادیر K بزرگترین Eigen ساخته می‌شود. اندازه ارتباط تعریف شده در رابطه (۲) برای ساخت یک مجموعه واژگان حساس به موتیف استفاده می‌شود که در آن، برای هر عنصر تگ s که با t هم‌رخداد باشد (یعنی $f(s, t) > 0$) عناصر واژگانی t لیست شده باشد به صورت نزولی اتصالات مقادیر τ (v, s). برای مثال، فرض کنید دامنه‌های مبدا و مقصد به صورت شکل ۴-۱۰ تعریف شده باشند.



شکل ۴-۱۰- مثال دامنه مبدا و دامنه مقصد

جریان‌های‌کاری دامنه مبدا مثال نشان داده شده در شکل ۴-۱۰ به صورت زیر هستند:

$G_{S1}= \{Z, Y, G\}$, $G_{S2}= \{L, Z, Y, G\}$ و جریان‌های‌کاری دامنه مقصد به صورت زیر هستند:

$G_{T1}=\{X, X, X, L\}$, $G_{T2}= \{H, X, X\}$, $G_{T2}= \{G, X, X\}$ که در آن عناصر مجموعه‌ها انتزاعات مفهومی گام‌های جریان‌کاری هستند. عناصر DI در حقیقت تگ‌های محوری هستند که تکرار آن‌ها بیشتر از سایرین است، و معمولا این تگ‌ها استفاده می‌شوند. به این‌ها موتیف‌های جریان‌کاری گویند. عناصر DS مجموعه مکمل عناصر DI. در این مجموعه‌ها مجموعه تگ DS به صورت $DS= \{H\}$ و مجموعه تگ DI به صورت $DI= \{X, G, L, Y, Z\}$ می‌باشد.

روش SFA می‌تواند به طور کامل ارتباط بین تگ‌های DI و DS را از طریق هماهنگ‌سازی آن‌ها در گراف دووجهی، برقرار کند. SFA از مجموعه خصیصه‌ها که به طور خودکار در مدل آموخته شده-اند، ماتریس U را به وجود می‌آورد. برای این کار دامنه‌های مختلفی از جریان‌کاری هماهنگ شده‌اند. در مرحله تست عملیات الگوریتم برای یک جریان‌کاری پیچیده اعمال می‌شود که باید با مدل مفهومی

ساخته شده انطباق‌هایی یافت شود. ماتریس U برای محاسبه میزان انطباق جریان‌کاری فاز تست با مدل مفهومی به وجود آمده از OSPA استفاده می‌کند. پس از مشخص شدن منطبق ترین ماتریس U بهترین چینش مبتنی بر DNN یافت می‌شود.

در گام دوم اطلاعات موتیف ترتیبی جریان‌های‌کاری با تعبیه خصیصه مبتنی بر DNN در نظر گرفته می‌شود. مبتنی بر این، روش ارائه شده درباره ترتیب موتیف‌های جریان‌کاری مرتبط با دامنه-های مختلف و بهترین ترتیب مفهومی موتیف‌های جریان‌کاری پیشنهادی تصمیم گیری می‌کند. برای همین منظور، ابتدا این پژوهش کیسه‌ای از تگ‌ها با مبدا و مقصدها می‌سازد. سپس اندازه‌های متعلق را مبتنی بر روابط (۴-۳) و (۴-۴) می‌سازد [87].

$$S_percent = \frac{A_{S_i} \times 100}{DS_size} \times \frac{A_{S_i}}{S_size}, \{i = 0,...,n\} \quad (۴-۳)$$

که در آن A_{S_i} تعداد برچسب نوع دامنه مبدا است، S_size تعداد تگ‌های مبدا، DS_size تعداد تگ‌های مبدا DS، $S_percent$ درصد تعلق موتیف‌های مشترک را به مبدا نشان می‌دهد.

$$T_percent = \frac{A_{T_i} \times 100}{DS_size} \times \frac{A_{T_i}}{T_size}, \{i = 0,...,n\} \quad (۴-۴)$$

که در آن A_{T_i} تعداد برچسب نوع دامنه مبدا است، T_size تعداد تگ‌های مقصد، DS_size تعداد تگ‌های مقصد DS، $S_percent$ درصد تعلق موتیف‌های مشترک را به مبدا نشان می‌دهد. A_{S_i} و A_{T_i} توسط $c(t_i, x_j)$ به دست می‌آیند که تعداد تگ‌های t_i را در رابطه با نوع x_j مشخص می‌کند.

چنین دسته‌بندی کمک می‌کند تا بهترین انطباق یک جریان‌کاری تست پیچیده مفروض را مشخص کرد. برای مثال اگر نوع جریان کاری مفروض بیوانفورماتیک باشد، روش بر اساس نتایج معادلات (۴-۳) و (۴-۴) نتیجه‌گیری می‌کند و گروه موتیف‌هایی را که بیشترین انطباق با بیوانفورماتیک دارند را انتخاب می‌کند.

۴.۵. اعمال الگوریتم یادگیری عمیق با DNN

رساله حاضر ابتدا از ماتریس هم‌برخوردی استفاده می‌کند و سپس برای چینش تگ‌ها از DNN بهره می‌برد. هر عملیاتی که در مراحل جریان‌کاری در SFA انجام می‌شود، توسط DNN در موتیف-های جریان‌کاری صورت می‌گیرد. به عنوان مثال، اگر تگ‌هایی در SFA به عنوان محور هستند و آن‌ها از تگ‌های DS متفاوت هستند، انواع مختلفی از عملیات روی آن‌ها انجام می‌شود. همان‌طور در DNN

عملیات انجام شده بر روی تگ‌های محوری متفاوت است، از آن‌چه در تگ‌های *DS* انجام می‌شود. با DNN، بردارهای خصیصه براساس ساختار جریان‌های‌کاری و پردازش تگ‌ها ساخته می‌شوند. مبتنی بر DNN تگ‌هایی که باهم نزدیک‌تر هستند، بردارهای مشابه دارند. به عنوان مثال، ترتیب معقول موتیف‌های جریان‌کاری می‌تواند از مفاهیم استخراج شده مشتق شود.

با یک جریان‌کاری مفروض یا جریان‌کاری برچسب نوع نخورده، جریان‌کاری به موتیف‌های منحصر به فرد با روش DNN برچسب می‌خورد. بعد از جستجوی شبیه‌ترین جریان‌کاری در مجموعه داده با جریان‌کاری پیچیده مفروض، DNN می‌تواند بینشی درباره ارتباط جریان‌کاری مفروض و جریان‌کاری یافت شده بدهد. برای پیدا کردن شباهت‌های جریان‌های‌کاری، رساله حاضر از روش مبتنی بر فاصله OSPA [59] استفاده می‌کند. روش OSPA جایگزینی سیستماتیک بر روش محاسبه خطای مربعی است [59]. در ادامه این بخش تابع هزینه‌ای معرفی شده است که به چینش تگ‌ها براساس DNN کمک می‌کند.

۴.۵.۱ تابع هزینه در یادگیری عمیق

در روش یادگیری عمیق لازم است تنظیمات نورون‌ها توسط یک تابع هزینه اعمال شود. در این رساله تابع هزینه‌ای ارائه شده است که به صورت رابطه (۴-۵) قابل بیان است. زیرا با دیدگاه رساله حاضر که رفتار آماری کاربران را مدنظر قرار می‌دهد، بهترین تنظیم را می‌توان با تنظیم تابع هزینه براساس وزن برای سطح تگ‌ها در جریان‌کاری و براساس تکرار موتیف‌های جریان‌کاری به دست آورد. ایده این است که تابع هزینه N را بیابیم که در آن تگ‌هایی که به ندرت رخ می‌دهد هزینه تقریبا زیادی دارند و تگ‌هایی که خیلی پرتکرار رخ می‌دهد هزینه کمتری دارد (و بنابراین با $(1-\alpha)$٪ ترجیح داده می‌شود:

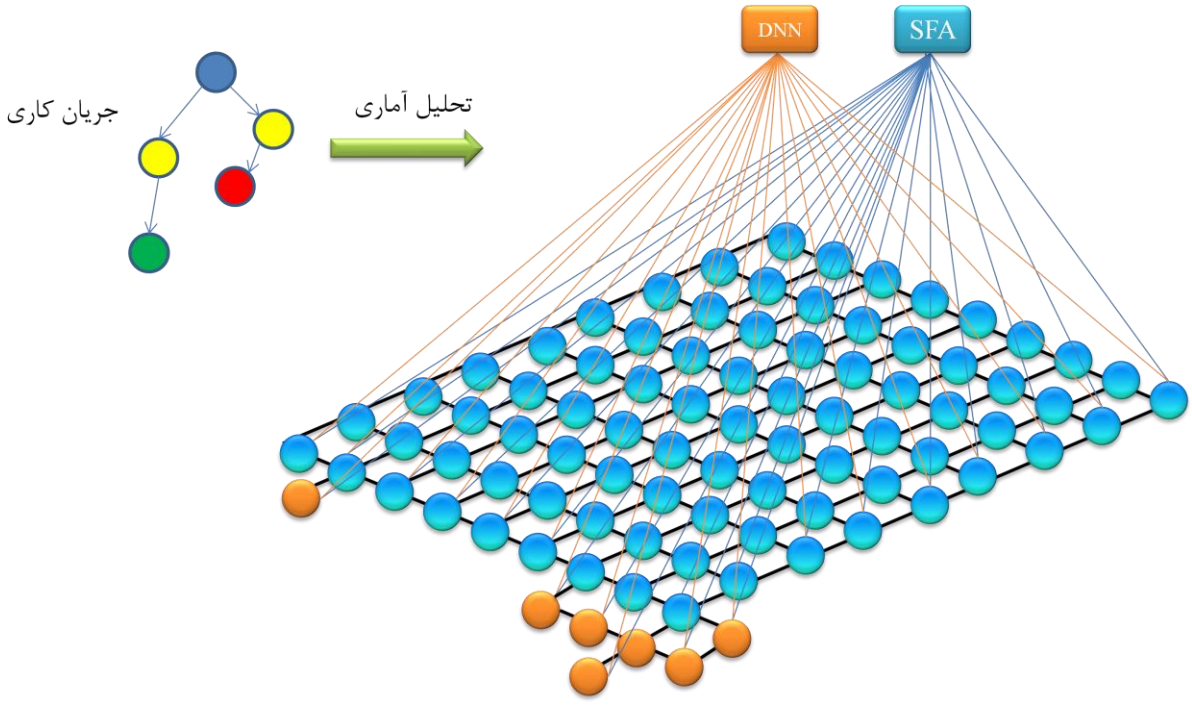
$$N = (1-\alpha)freq + weight \tag{۴-۵}$$

که در آن freq تکرار موتیف‌های جریان‌کاری است، و weight سطح‌های نسبت داده شده به تگ‌ها است. از آن‌جا که قصد داریم برمبنای رفتار آماری کاربران تصمیم بگیریم که ساختار جریان‌کاری به چه صورت باشد در رابطه تابع هزینه از weight استفاده می‌کنیم. weight سطحی است که کاربران در گزارش عملکرد جریان‌کاری برای هر تگ در نظر گرفته‌اند. اما، ممکن است کاربران در جریان‌کاری

دچار اشتباهاتی شده باشند. لذا، بررسی می‌کنیم که بیشترین تعداد کاربران برای هر تگ چه سطحی در نظر گرفته‌اند. به عبارتی دیگر، در حین بررسی در تابع هزینه فقط نظر کاربران نیست که اعمال می‌شود. بلکه رفتار آماری کاربران با مقدار freq در نظر گرفته می‌شود. عبارت $1 - \alpha$ کمک می‌کند که trade-off میان freq و weight تعدیل شود. گرچه در پیاده‌سازی این رساله، α را برابر صفر درنظر گرفته‌ایم. با محاسبه تابع هزینه N برای هر نورون در DNN ، بهبود چند جزئی بر روی جریان‌کاری اعمال می‌شود.

۴ ۶. راهکار به تغییر پویای جریان‌کاری

شکل ۴-۱۱ پویایی ایجاد پایگاه‌داده جریان‌های‌کاری را نشان می‌دهد؛ هنگامی که جریان‌کاری نمونه‌ای وارد می‌شود، دو تحلیل توسط الگوریتم‌ها SFA و DNN بر آن‌ها صورت می‌گیرد. این باعث می‌شود دانش جدیدی به صورت پویا به پایگاه داده‌ای که براساس تحلیل این دو روش ایجاد شده اضافه شود.



شکل ۴-۱۱- پویایی ساخت پایگاه‌داده

فصل پنجم:

نتایج

۵ ـ۱. مقدمه

سازمان‌ها مخصوصا هنگامی که مدل‌های فرآیند سازمانی را طراحی می‌کنند و یا با مهندسی معکوس آن‌ها را به دست می‌آورند، به ارزیابی کیفی آن‌ها بیشتر نیاز پیدا می‌کنند. زیرا به عنوان مثال اجزای نامرتبط و یا بی‌استفاده در نمودارهای تولید شده می‌تواند کیفیت کار را پایین بیاورد. امروزه روش‌های ساده‌سازی برای تخفیف این تاثیرات بسیار به کار آمده‌اند. به منظور ارزیابی تاثیر این روش-ها با استفاده از مجموعه‌ای از مدل‌های فرآیند یک مطالعه موردی ترتیب داده‌ایم. در این مطالعه از یک سری از مجموعه‌های داده بهره می‌بریم که در ادامه آورده شده اند. سپس به ارائه نتایج و تحلیل آن‌ها می‌پردازیم.

۵ ـ۲. مجموعه داده

برای ارزیابی روش‌ارائه شده با توجه به مشخصات ذکر شده می‌بایست مجموعه‌داده‌هایی استفاده کنیم که تا حد امکان جامعیت کافی را در توزیع منابع و فرآیندهای جریان‌های‌کاری داشته باشد. بدین منظور برای تشخیص این که جریان‌کاری داده شده برای ساده‌سازی متعلق به کدام دسته جریان‌کاری می‌باشد، از مجموعه داده‌هایی از دامنه‌های مختلف شامل بیمارستان، تعمیر، مالی و علمی استفاده می‌کنیم. در واقع داده‌های ما شامل ۴ مجموعه داده می‌شود که نمودارهای جریان‌کاری آن‌ها از گزارش رخدادهای[1] آن‌ها قابل استخراج است.

۱ -گزارش‌های رخداد واقعی یک بیمارستان دانشگاهی در رساله حاضر مورد استفاده قرار می‌گیرد که در ابتدا برای استفاده در اولین چالش هوش تجاری کسب و کار (BPI[2]) مورد استفاده قرار گرفته است [49]. این مجموعه داده شامل ۱۱۴۳ مورد اجرایی می‌شود.

۲ -مجموعه‌داده دیگر مجموعه داده مالی است که شامل ۴۳۶۶ مورد اجرایی می‌شود که برای چالش BPI 2012 استفاده شده است [50].

۳ -برای جریان‌های‌کاری علمی پژوهش حاضر داده‌ها را از مبدا شامل ۷۰۶۶۵ جریان‌کاری طراحی آزمایش (ArrayExpress) استخراج کرده است [75]. در این‌جا زیرمجموعه‌ای از آن استفاده شده است شامل ۲۹ جریان‌کاری علمی.

۴ -در نهایت یک مجموعه‌داده شامل ۴۱۲۳ مورد اجرایی [76] از فرآیند تعمیر تلفن برای آموزش استفاده شده است.

مجموعه داده‌های آزمون شامل یک ثبت گزارشات مربوط به فرآیند درخواست وام از موسسه مالی هلندی مورد استفاده برای چالش BPI 2017 [77] است. داده‌های این مجموعه و روند بررسی آن مشابه [50] می‌باشد. با این حال، سیستم حمایت از روند در میانه کار تغییر کرده است. به ویژه، این سیستم اکنون اجازه چند پیشنهاد برای کاربرد را می‌دهد. این داده‌ها حاوی تمام برنامه‌های ارسال

[1] - Events
[2] - Business Process Intelligence

شده از طریق سیستم برخط در سال ۲۰۱۶ و رخدادهای بعدی آن‌ها تا سال ۲۰۱۷ می‌شوند. تعداد ردیابی‌ها در این مجموعه داده ۳۱۵۰۹ می‌باشد.

در مجموعه داده‌های تعمیر، چالش BPI 2012، داده‌های بیمارستانی و نیز داده‌های چالش BPI 2017، داده‌های سری زمانی بخش مهمی از داده‌های آن‌ها را تشکیل می‌دهند. این مجموعه‌های داده شامل ثبت گزارش عملکردها می‌باشند. در آن‌ها هر مورد اجرایی با دنباله‌ای از رخدادها توصیف می‌شود. بعد از نام فعالیت آن، رخداد می‌تواند خصیصه‌های بیشتری داشته باشد. مانند زمان، منبع (منابع) درگیر، نوع تراکنش (آغاز، کامل، انصراف[1]، و غیره)، هزینه‌های نسبت داده شده و غیره.

۵ ۳. نتایج و تحلیل

الف- نتایج

ابتدا روش ما درصد وابستگی به بعضی تگ‌ها برای دامنه‌ای خاص (داده مالی استفاده شده برای چالش BPI 2017) را به عنوان گام SFA محاسبه می‌کند. جدول ۵-۱ نتایج اندازه‌گیری درصد وابستگی مجموعه تگ‌ها دامنه مالی (یعنی چالش BPI 2017) را به عنوان مبدا در کیسه تگ‌ها شامل مبدا و دامنه‌های مختلف مقصد نشان می‌دهد.

جدول ۵-۱- اندازه‌های تعلق دامنه متقابل دامنه‌های مختلف جریان‌کاری

	مقصد			
	بیوانفورماتیک	مالی	بیمارستانی	تعمیر
تعمیر٪	۰,۰	۰,۰	۰,۰	۳,۷۰
بیوانفورماتیک٪	۸,۳۰	۰,۰	۰,۰	۰,۰
مالی٪	۷۴,۵۸	۵۴,۳۲	۳۸,۶۰	۸۴,۶۲
بیمارستانی٪	۰,۰	۰,۰	۳۹,۸۹	۰,۰

نتایج جدول ۵-۱ اندازه‌های تعلق دامنه‌های جریان‌کاری مختلف را برحسب درصد با مطالعه موردی دامنه مالی چالش BPI 2017 نشان می‌دهد. جدول ۵-۱ نشان می‌دهد که تقریبا ۶۳,۰۳٪ موتیف‌ها به طور میانگین در کیسه تگ‌ها شامل دامنه مبدا چالش BPI 2017 و دامنه‌های مختلف مقصد به دامنه مالی تعلق دارند. نتایج جدول ۵-۱ نشان می‌دهد که بیشترین تگ‌های مشترک با نوع محیط اجرایی که برای آن جریان‌کاری طراحی شده است همبستگی دارند. این پژوهش این محاسبه را براساس میانگین نتایج شناسایی می‌کند.

[1] - Abort

براساس رابطه متریک نرخ دقت دسته‌بند (CA[1]) می‌توان مقدار دقت تخمین دامنه جریان‌کاری را به صورت زیر محاسبه کرد (رابطه ۵-۱):

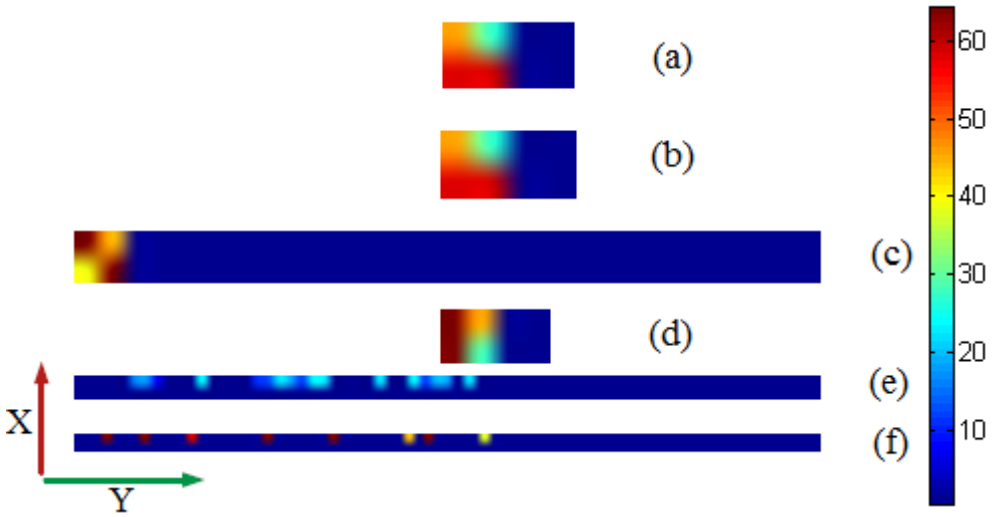
$$CA = \frac{TN + TP}{TN + FN + TP + FP} = 87\% \quad (۵-۱)$$

که در آن TP بیان‌گر دامنه واقعی صحیح جریان‌کاری و TN دامنه صحیح غیرواقعی جریان‌کاری، FP بیان‌گر دامنه واقعی غلط جریان‌کاری و FN دامنه غلط غیرواقعی جریان‌کاری می‌باشد.

تخمین داده‌ها به عنوان نتیجه انتقال دانش به طور موفقیت‌آمیز انجام شده است (تخمین داده‌ها به طور میانگین براساس رابطه (۵-۱) تا ۸۷ درصد منجر به تخمین صحیح دامنه جریان‌کاری می‌شود.). کارآیی یادگیری با حذف تلاش‌های یادگیری داده خیلی گران‌تر بهبود یافته است. در ادامه این بخش نتایج بهبود جریان‌کاری نشان داده شده است.

بعد از اندازه‌گیری شباهت جریان‌های‌کاری و یافتن بخش‌های تکراری جریان‌های‌کاری ما می‌توانیم تصمیم بگیریم که کدام دامنه‌ها ممکن است یک جریان‌کاری دیده نشده مفروض داشته باشد. سپس لازم است بدانیم که چه فرمی ممکن است فرم بهتری برای این جریان‌کاری دیده نشده مفروض باشد. به عبارتی دیگر، ما چطور می‌توانیم یک جریان‌کاری مفروض را بهبود دهیم. برای دست یافتن به فرم بهتر جریان‌کاری لازم است مدل جریان‌کاری بهتر دیگری را در دامنه مشخص شده خودش بدانیم. ما نماینده‌های هر دامنه را طوری انتخاب می‌کنیم که نماینده بهترین فرمت را در ترتیب و معنا داشته باشد. ما این نماینده را براساس تکرار و براساس نمره بالاتری که از روی گراف دووجهی بدست آمده، انتخاب می‌کنیم. سپس ماتریس U را برای این نماینده استخراج می‌کنیم (شکل ۵-۱). از سویی دیگر یک ماتریس U را برای یک جریان‌کاری دیده نشده مفروضی می‌سازیم که محتوای آن مشابه ماتریس U می‌باشد که برای نماینده انتخاب شده است.

[1] - Classification Accuracy Rate

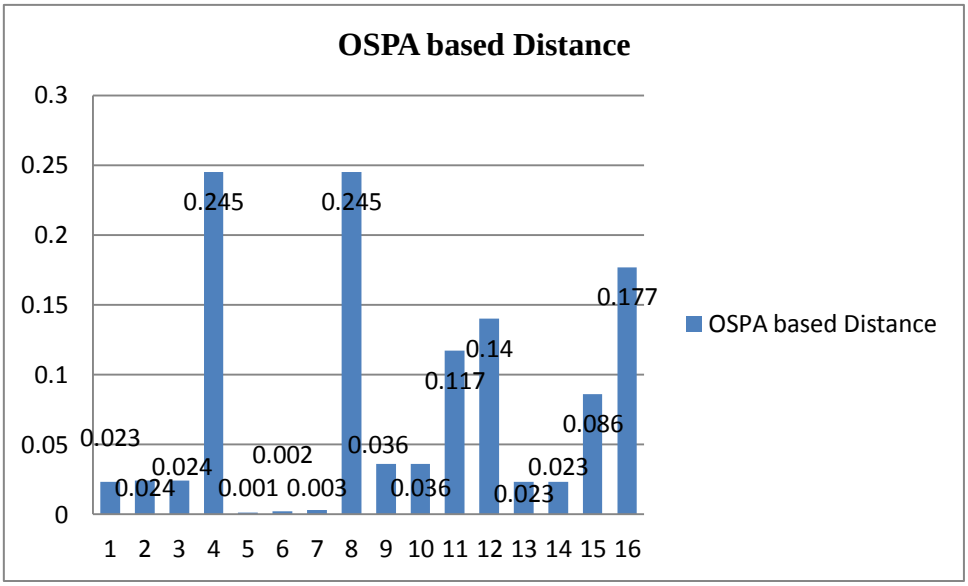


شکل ۵-۱- مقادیر ماتریس U، k=2. (a) مجموعه‌داده BPI-2012 و بیمارستانی، (b) مجموعه‌داده بیمارستانی و علمی، (c) مجموعه‌داده علمی و BPI-2017، (d) مجموعه‌داده علمی و BPI-2012، (e) مجموعه‌داده بیمارستانی و BPI-2017، (f) مجموعه‌داده BPI-2012 و BPI-2017. محور X بردارهای خصیصه را نشان می‌دهد و محور Y موتیف‌های جریان-کاری را نشان می‌دهد.

فرکانس تکراری که در آن موتیف‌ها ظاهر می‌شوند بستگی به اختلافات بین محیط‌های مختلف و دامنه‌های مختلف دارد. در رساله حاضر انواع مختلف دامنه استفاده می‌شود (مثل تعمیر، بیوانفورماتیک، مالی، و بیمارستانی). برای ساخت ماتریس U با چهار دامنه ما ۱۶ کیسه تگ بدست آوردیم. هر کیسه تگ یک ماتریس U متناظر دارد که با محاسبه مقادیر K بزرگترین Eigen ماتریس دووجهی به دست می‌آید. پارامتر K برای دست یافتن به تمایزهای دقیق‌تر بین جریان‌های‌کاری مرتبط مهم است. ما ماتریس U مربوط به ۶ کیسه تگ را در شکل ۵-۱ تصویر کرده‌ایم. برای مثال شکل ۵-۱.(a) ماتریس U مبتنی بر مجموعه‌داده‌های بیمارستانی و BPI 2012 می‌باشد. شکل ۵-۱ مقادیر ماتریس U را برای وقتی نشان می‌دهد که روش مقادیر $K=2$ بزرگترین Eigen در نمایش دوبعدی ماتریس دارد. در شکل ۵-۱، مقادیر رنگی در ماتریس U نشان‌دهنده ارتباط بین داده‌های مختلف از دامنه‌های مختلف هستند. طول بیشتر ماتریس U در شکل ۵-۱ نشان می‌دهد که داده‌های متنوع‌تری مورد بررسی قرار می‌گیرند. مقادیر غیرصفر با نقاط رنگی گرم‌تر نشان داده می‌شوند. مقادیر صفر داده پنهانی که به خاطر مکانیزم متفاوتشان توسط آستانه حذف شده‌اند را نشان می‌دهد. این مقادیر نشان می‌دهند که رابطه‌ای بین داده‌های به دست آمده از دامنه‌های مبدا و مقصد وجود دارد. بنابراین، هرقدر که تعداد نقاط رنگی و رنگ‌ها تنوع بیشتری داشته باشند داده‌های بیشتری می‌توانند تمایز یابند. همان‌طور که در شکل ۵-۱ ملاحظه می‌شود در ماتریس‌های U (a)، (b) و (d) طول ماتریس کوتاه‌تر از سایر ماتریس‌هاست. این نشان می‌دهد که این ماتریس‌ها می‌توانند محدوده کوچکی از داده‌ها را اسکن کنند. هم‌چنین، مقادیر غیرصفر ماتریس‌های (a)، (b) و (d) اطراف و در

حدود یک بخش متمرکز شده‌اند. اگرچه در ماتریس (c) داده بیشتری مطالعه شده‌است اما هنوز مقادیر غیرصفر در اطراف و در حدود یک بخش متمرکز شده‌اند. بنابراین، مبتنی بر آن، ما تقریبا رفتاری مشابه خواهیم داشت. بهترین خروجی درباره داده مقصد مربوط به ماتریس‌های (e) و (f) می‌شود. هر دو ماتریس (e) و (f) حوزه مقادیر بیشتری را نسبت به بقیه پوشش می‌دهند. این به خاطر نقاط رنگی است که به طور یکنواخت توزیع شده‌اند. همان‌طور که در شکل ۵-۱ مشاهده می‌شود هنگامی که دامنه مقادیر متفاوت یا داده زیادی دارد ماتریس U مقادیر متنوعی دارد و پیچیده‌تر و گسترده‌تر می‌شود. این پیچیدگی باعث افزایش زمان اجرا می‌شود، و ممکن است با اهمیت نباشد. چرا که زمان اجرای روش در سیستم محدود می‌شود فقط به یک بار آموزش دسته‌بند.

ما با استفاده از فاصله OSPA فاصله ماتریس U را محاسبه می‌کنیم. به عنوان مثال فاصله ماتریس U محاسبه شده برای ثبت رخداد جریان‌کاری پیچیده از سایر جریان‌های‌کاری مالی محاسبه می‌شود. نتایج این گام در شکل ۵-۲ نشان داده شده است.



شکل ۵-۲- فاصله مبتنی بر OSPA. محور افقی شماره جریان‌کاری است. محور عمودی میزان فاصله را نشان می‌دهد.

نزدیک‌ترین مدل جریان‌کاری به جریان‌کاری تست مدلی است که کمک می‌کند ترتیب مدل جریان‌کاری بهبودیافته را مشخص کنیم. از نتایج گام SFA مشخص می‌شود که جریان‌کاری تست بیشتر شبیه دامنه نوع مالی است. بنابراین مبتنی بر نتایج شکل ۵-۲ بردار DNN از ۵مین جریان‌کاری مالی ذخیره شده در مجموعه‌داده برای مشخص کردن ترتیب جریان‌کاری بهبودیافته استفاده خواهد شد.

بعد از بهبود جریان‌کاری، اتصال(cp[1]) جریان‌کاری بهبودیافته محاسبه می‌شود (جدول ۵-۲). مدل‌های ساده‌تر جریان‌های‌کاری را می‌توان به عنوان مدل‌هایی تعریف کرد که بیش از حد پیچیده نیستند. به عنوان مثال، آن‌ها بسیار بزرگ نیستند و چگالی[2] یال‌ها [3]و جفت شدن عناصر در آن‌ها کم است [78]، [79]. برای فرآیندی که شامل مجموعه‌ای از تگ‌های وظیفه (S) در ساختار جریان‌کاری می‌شود، اتصال فرآیند به صورت زیر تعریف می‌شود [78]:

$$cp = \begin{cases} \dfrac{\left|\left\{(T_1,T_2) \in S \times S \mid \overline{T_1} \neq \overline{T_2} \wedge (T_1 \bigcap T_2) \neq 0\right\}\right|}{\left|\overline{S}\right|.\left(\left|\overline{S}\right|-1\right)}, for \left|S\right| > 1 \\ 0, for \left|S\right| \leq 1 \end{cases} \qquad (۵-۲)$$

که در آن T_1 و T_2 دو تگ هستند. در صورتی که ارتباطی بین این دو تگ برقرار باشد $(T_1 \bigcap T_2) \neq 0$، صورت کسر غیر صفر خواهد شد. در غیر این صورت مقدار صفر خواهد گرفت. متریک cp میزان اتصالات میان قطعات جریان‌کاری را نشان می‌دهد. هر قدر اتصالات از قطعات پایین‌تر به قطعات بالاتر بیشتر شود (یعنی دور داشته باشیم) با افزایش تودرتویی و پیچیده‌تر شدن جریان‌کاری مواجه می‌شویم. کم بودن میزان cp در یک جریان‌کاری نشان می‌دهد که جریان‌کاری از تعداد اتصالات کمتر و ساده‌تری برخوردار است.

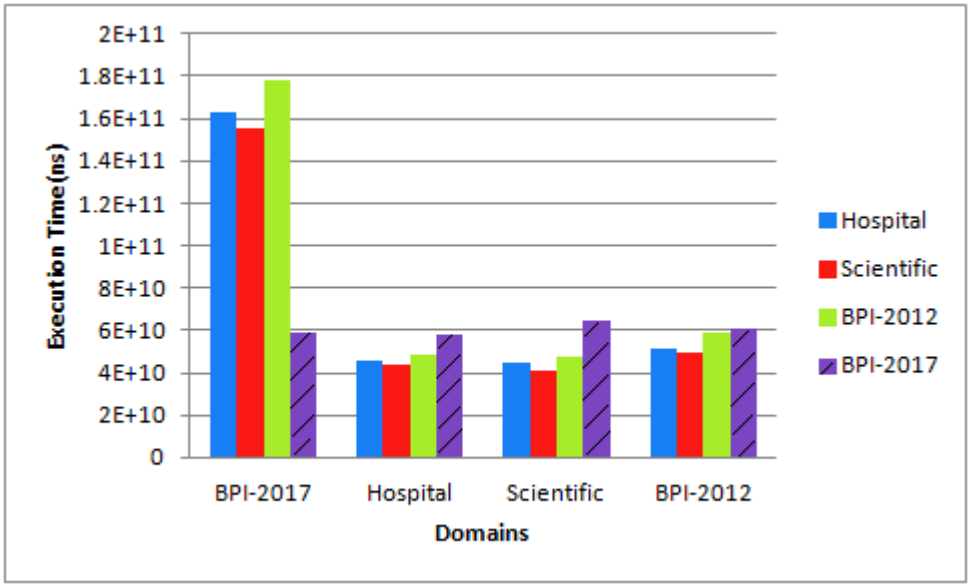
جدول ۵-۲- تحلیل بهبود جریان‌کاری

متریک	مقدار	
Cp	بهبودیافته	بهبود نیافته
	0.1860	0.5274

نتایج جدول ۵-۲ اتصال جریان‌کاری پیچیده را بعد از بهبود و ساده‌سازی نشان می‌دهد. همان‌طور که از جدول برداشت می‌شود، مقدار پارامتر اتصال جریان‌کاری به مقدار ۰,۱۸۶ کاهش خواهد یافت. در حالی که بدون استفاده از این روش این مقدار ۰,۵۲۷ می‌باشد. کاهش پارامتر اتصال به میزان حدود ۳۰ درصد، باعث کاهش برخوردهای خطازا در جریان‌کاری می‌شود. کیفیت جریان‌کاری بهبود یافته از این جهت که مقدار cp بعد از بهبود کمتر شده مطلوب است. زیرا باعث افزایش خوانایی جریان‌کاری و در نتیجه شفاف‌سازی عملکرد کاربر می‌شود [48]. در محاسبه cp منابع در دسترس

[1] - Coupling
[2] - density
[3] - arcs

می‌توانستند در موتیف‌ها دخیل شوند. ما فرض کردیم که منابع برای همه فعالیت‌ها دردسترس هستند، و استخراج موتیف را به منابع محدود نکردیم [87].



شکل ۵-۳- زمان پردازش روش ارائه شده برای اندازه مجموعه‌داده‌های مختلف – محور عمودی زمان را برحسب ns نشان می‌دهد.

از جدول ۵-۳ مشاهده می‌شود که زمان اجرا زیاد می‌شود هنگامی که مجموعه‌داده بزرگ می‌شود و تعداد تگ‌ها افزایش می‌یابد. از سویی دیگر از شکل ۵-۳ ملاحظه می‌شود لازم نیست تا داده‌ها زیاد باشند تا زمان اجرای روش افزایش یابد، و در اغلب اوقات حد آستانه با داده‌های زیاد کار می‌کند. داده شکل ۵-۳ نشان می‌دهد که در هر اجرا از دامنه‌های مبدا، روش ارائه شده می‌تواند زمان اجرا را کاهش دهد، حتی اگر دامنه مقصد شامل داده زیادی باشد. توضیح بیشتر اینکه، حجم داده BPI 2017 بیشترین است و سایر دامنه‌ها حجم داده کمتری نسبت به این دامنه دارند. واضح است که هنگامی که داده BPI 2017 به هر دامنه مبدا اضافه می‌شود، مساله با شکل سنگین‌تری از داده‌ها مواجه می‌شود.

جدول ۵-۳- مقایسه زمان اجرای روش ارائه شده

مبدا	مقصد	تعداد تگ‌ها	زمان اجرا (ns)
BPI-2012	BPI-2017	1542600	161942430734
BPI-2012	علمی	295089	47502683874
BPI-2012	بیمارستانی	454923	51955451599
BPI-2017	علمی	1247887	64293109780
BPI-2017	بیمارستانی	1407721	57628940065
علمی	بیمارستانی	160210	44093711337

ب- تحلیل و مقایسه با روش‌های مشابه

اغلب جریان‌های‌کاری واقعی از این جهت پیچیده هستند که آن‌ها خصیصه‌های بی‌اهمیت زیادی را نشان می‌دهند. با افزایش روز افزون انباره‌های جریان‌کاری، دسته‌بندی و سازمان‌دهی آن‌ها به ابزارهای متنوع کاربران نیاز پیدا می‌کند که کار دشواری است. برای سازمان‌دهی بهتر جریان‌های‌کاری واقعی، الگوریتم دسته‌بندی برای تنظیم بافت جریان‌های‌کاری از جهت ارتباط جریان‌های‌کاری باهم ارائه کردیم. در دسته‌بندی ما از برخی زیرجریان‌های‌کاری‌های پیشین به عنوان موتیف استفاده شد. ساخت ماتریس هم-رخدادی موتیف‌های جریان‌کاری در دامنه‌های مختلف اجازه می‌دهد که جریان مشابه از داده‌های یک دامنه را بازیابی کنیم که سازمان‌دهی بهتری برای نتایج جستجو فراهم می‌کند.

در این رساله روش ارائه شده را به جریان‌های‌کاری واقعی اعمال کردیم و هنگامی که تعداد تگ‌ها افزایش یافت، نتایج امکان مدیریت داده‌های بزرگ[1] ناشی از جریان‌کاری را فراهم کرده است [83]. مزیت الگوریتم ما با توجه به سایر روش‌هایی که قبلا برای دسته‌بندی ارائه شده‌اند، روش‌هایی مثل LSA و یا نظریه طیفی گراف، این است که این روش‌ها داده‌های پرت را حذف می‌کند و ارتباط بین داده‌های دامنه‌ها را در نظر می‌گیرد، و بنابراین به طور موثری به محدوده وسیعی از BPMSها اعمال می‌شود که روی دامنه‌های مختلف مثل سازمانی یا علمی کار می‌کنند. برای توضیح بیشتر مقایسه روش ما را با سایر روش‌ها در جدول ۵-۴ ببینید.

جدول ۵-۴- خلاصه خصوصیات روش‌های مختلف

روش	انتخاب خصیصه‌های محوری	مقدار Eigen	حساس به داده‌های پرت
SCL	هیوریستیکی	خیر	خیر
LSA	خیر	بله	بله
SFA	غیر هیوریستیکی	بله	خیر
نظریه طیفی گراف	خیر	بله	بله

ضمنا، اخیرا طرح‌ها و پیاده‌سازی های مختلفی از تکنیک‌هایی برای بهبود کارآیی سیستم‌هایی ارائه شده‌اند که با داده‌های بزرگ کار می‌کنند [82-80]. مثالی از مدیریت داده بزرگ در سیستم‌های مدیریت جریان‌کاری در [81] معرفی شده است که جریان‌های‌کاری را به سیستم‌های با محاسبات با کارآیی بالا (HPC[2]) مقیاس می‌کند. مثال دیگر برای مدیریت جریان‌های‌کاری با داده‌های بزرگ می-

[1] - Big Data
[2] - High Performance Computing

تواند کار ارائه شده در [84] باشد که با توجه به این‌که در ابر پیاده‌سازی شده، مقیاس‌پذیر است و از نظر هزینه به صرفه می‌باشد. راه‌حل دیگر در بازنمایی جریان‌های‌کاری است؛ می‌توان جریان‌هاکاری را کدگذاری کرد تا جریان‌های‌کاری برای کاهش زمان اجرای بیشتر و مقیاس کردن جستجوی شباهت به اندازه‌های انباره‌های فعلی شاخص‌گذاری شوند [59].

در اغلب حوزه‌های کاربردی، محاسبه اتصال فرآیند (cp) کافی نخواهد بود و روشی دیگری برای تشخیص نقض بیشتر (در اسرع وقت) که بیشتر پاسخگو باشد، مطلوب خواهد بود. این روش به عنوان مثال می‌تواند از طریق نظارت بر انطباق[1] [86] صورت گیرد. نظارت بر انطباق راهی است برای شناسایی زمانی که یک دنباله از وقایع از رفتار مورد انتظار متفاوت می‌شود. از آنجا که مدیریت فرآیندها به طور معمول نیاز به محیط‌های متصل بزرگ دارند، مفاهیمی که باید مورد نظارت قرار گیرند اغلب بسیار پیچیده می‌شوند. این کار در شبکه‌ای از گره‌های محاسباتی، برای دستیابی به مقیاس‌پذیری مورد نظر، نیاز به انتشار وظیفه نظارت بر انطباق دارد. با این حال، بسیاری از روش‌های نظارت پیشنهادی محدودیت دارند و آنها در قبال مدیریت داده‌های بزرگ و اینترنت اشیا [87] نیازمند رویکرد مسئولانه هستند. از این رو، به عنوان کاری در آینده پیشنهاد می‌کنیم رویکردهای نظارت انطباقی اتخاذ شوند تا در معماری‌های توزیع شده [87] در کنار خدمات بهبود گردش‌کار، فرآیندها استفاده شوند.

[1] - Compliance Monitoring

فصل ششم:

نتیجه گیری و پیشنهادات

۶ ـ۸ نتیجه‌گیری

ساده‌سازی یک جریان‌کاری از دو دیدگاه ساختاری و معنایی مورد بررسی قرار می‌گیرد. در این نوشتار به بررسی روش‌های موجود ساده‌سازی جریان‌های‌کاری پرداختیم.

به طور کلی اغلب روش‌های ارائه شده به صورت ساختاری به کاهش جریان‌های‌کاری پرداخته‌اند. در حالی‌که برخی از مشکلات افزایش پیچیدگی ناشی از تفکر معنایی طراح بوده است. از این رو لزوم توجه و پرداختن به این جنبه بیش از پیش احساس می‌شود. به منظور ارائه راه‌حل در این نوشتار ابتدا یک مجموعه ابزار به عنوان روش پیشنهادی ارائه شده است. وضعیت استفاده جریان‌های‌کاری به صورت آماری مورد تجزیه و تحلیل قرار گرفته تا بتوان از این تجزیه و تحلیل در ساده‌سازی آن‌ها استفاده کرد. با استفاده از تحلیل آماری صورت گرفته یک خوشه‌بندی مبتنی بر استخراج موتیف‌ها ترتیب داده شده تا بتوان عمل ساده‌سازی را خودکار کرد. عمل خوشه بندی موجب می‌شود که بتوان از نظر مفهومی و با استفاده از اطلاعات جریان‌های کاری موجود در هر خوشه همانند گزارش عملکردشان در ساده‌سازی های بعدی استفاده کرد. خوشه بندی و ذخیره‌سازی اطلاعات مشترک هر خوشه بخصوص در زمان عدم وجود گزارش از عملکرد یک جریان‌کاری می‌تواند بسیار مفید واقع شود.

روشی که ارائه داده‌ایم مبتنی است بر استخراج موتیف. در روش ارائه شده برای یادگیری موتیف‌های مشترک، یک ارزیابی شباهت متقابل و در نهایت بهبود جریان‌کاری پیشنهاد شده است. روشی که بدین منظور ارائه کرده‌ایم بر مبنای یادگیری انتقالی بوده است. روش ارائه شده پس از ملاحظه مجموعه موتیف‌های هر جریان‌کاری روشی برای خوشه‌بندی جریان‌های کاری ارائه کرد. زمانی که خوشه‌بندی صورت گرفت، می‌توان برای فرآیندهایی که گزارش عملکرد آن‌ها را نیز نداریم با کمک ادغام، ساده‌سازی ترتیب داد. این ادغام می‌تواند به این صورت باشد که به عنوان مثال ۲ موتیف پشت سرهم با یک گره جایگزین شوند. به منظور ارزیابی روش ارائه شده از چندین مجموعه‌داده استفاده کردیم تا بتوانیم عملی بودن آن را اثبات کنیم. نتایج نشان می‌دهد میزان اتصال جریان‌کاری مورد آزمایش بعد از اعمال روش بهبود یافته است که این میزان موفقیت آمیز بودن روش اعمال شده را نشان می‌دهد.

سنگینی محاسبات روش ارائه شده مربوط می‌شد به پیاده‌سازی روش SFA. پیچیدگی زمانی روش ارائه شده ضریبی است از n^3m که در آن $n<<m$ و m تعداد داده‌ها را نشان می‌دهد. هر قدر داده‌ها افزایش یابد پیچیدگی زمانی با ضریبی افزایش می‌یابد. پیشنهاد می‌شود از کارهای بعدی بهبود روش از جهت پیچیدگی زمان اجرا باشد.

۷ – منابع

[1] Guohai Zhang, Guanghui Zhou, Pingyu Jiang, Using Extended Activity-Network diagram to build a product design workflow model, in 5th International Conference on Wireless Communications, Networking and Mobile Computing, 2009.

[2] W. Sadiq, and M.E. Orlowska, Applying Graph Reduction Techniques for Identifying Structural Conflicts in Process Models, in International Conference on Advanced Information Systems Engineering. Springer Berlin Heidelberg, 1999.

[3] W.M.P Van Der Alast, Verification of Workflow Nets, In P. Azéma and G. Balbo, editors, *Application and Theory of Petri Nets 1997*, volume 1248 of *Lecture Notes in Computer Science*, pages 407-426. Springer-Verlag, Berlin, 1997.

[4] W. Sadiq and M.E. Orlowska, Analyzing Process Models Using Graph Reduction Techniques, in Journal of Information Systems, 25(2): 117-134, 2000.

[5] D. Blankenberg, and J. Taylor, A framework for collaborative analysis of encode data: making large-scale analyses biologist-friendly. Genome Res., 17(6): 960 –4, 2007.

[6] T. Oinn, and M. Greenwood, Taverna: lessons in creating a workflow environment for the life sciences. Concurrency and Computation: Practice and Experience, 18:10:1067 – 1100, 2005.

[7] A. Fiannaca, M.L. Rosa, R. Rizzo, A. Urso, S. Gaglio, An expert system hybrid architecture to support experiment management, in journal of Expert Systems with Applications, 41: 1609-1621, 2014.

[8] W-K. Chen, The Circuits and Filters Handbook: Computer Aided Design and Design Automation, Third Edition, published in University of Illinois, 2009.

[9] M.T. Wynn, H.M.W. Verbeek, W.M.P. van der Aalst, A.H.M. ter Hofstede and D. Edmond, Reduction Rules for YAWL Workflows with Cancellation Regions and OR-joins, *Information and Software Technology*, 51(6):1010-1020, 2009.

[10] J. Cardoso, J. Miller, A. Sheth, and J. Arnold, Modeling Quality of Service for Workflows and Web Service Processes, Technical Report, Computer Science, University of Georgia, 2002.

[11] V.R. Kavitha, R. Kavitha, and N.S. Kumar, Workflow Mining: Discovering Loops in Process Models, in International Journal of Computer Applications 22(6): 0975 – 8887, 2011.

[12] Ch. Flender, Th. Freytag, Visualizing the Soundness of Workflow Nets, Algorithms and Tools for Petri Nets (AWPN 2006), University of Hamburg, Germany, Department Informatics Report 267, pp. 47-52, 2006.

[13] W. Van der Alast and K. Van Hee, Workflow Management: Models, Methods and Systems, MIT press, 2014.

[14] H.V. Byelas and M.A.Swertz, Visualization of bioinformatics workflows for ease of understanding and design activities, Proceedings of the Biostec Bioinformatics-2013 conference, pp. 117–123, 2013.

[15] P. Alper, Kh. Belhajjame, C.A. Goble, Small Is Beautiful: Summarizing Scientific Workflows Using Semantic Annotations, IEEE 2nd International Congress on Big Data, 2013.

[16] D. Fahland, and W.M.P. van der Aalst, Simplifying Mined Process Models: An Approach Based on Unfoldings, *Business Process Management (BPM 2011)*, volume 6896 of *Lecture Notes in Computer Science*, pages 362-378. Springer-Verlag, Berlin, 2011.

[17] T. Liu, Y. Cheng, and Z. Ni, Mining Event Logs to Support Workflow Resource Allocation, Knowledge-Based Systems, 35:320-31, 2012.

[18] J. Nakatumba and W.M.P. van der Aalst, Analyzing Resource Behavior Using Process Mining, BPM 2009 International Workshops, 2009.

[19] R.P.J.C. Bose and W.M.P. van der Aalst, Process Mining Applied to the BPI Challenge 2012 Divide and Conquer While Discerning Resources, Technical Report: BPM Center BPM-12-16, bpmcenter.org, 2012.

[20] S. Weerapong, P. Porouhan, W. Premchaiswadi, Process Mining Using α-Algorithm as a Tool (A Case Study of Student Registration), Tenth International Conference on ICT and Knowledge Engineering, 2012.

[21] X. Ye, X. Bing, Ch. Zuo, The reduction and deadlock detection of cross-organizational business process based on Pi-calculus, in Journal of Procedia Engineering 15: 3487 – 3494, 2011.

[22] A.J.M.M. Weijters, W.M.P. van der Aalst, and A.K. Alves de Medeiros, Process Mining with the Heuristics Miner Algorithm, BETA Working Paper Series, WP 166, Eindhoven University of Technology, Eindhoven, 2006.

[23] B. Kiepuszewski, A.H.M. Hofstede, and C. Bussler, On Structured Workflow Modeling, International Conference on Advanced Information Systems Engineering. Springer Berlin Heidelberg, 2000.

[24] B. Underdahl, Business Process Management for Dummies, IBM Limited Edition, 2011.

[25] H. Li, Y. Wang, D. Zhang, M. Zhang, E. Chang, PFP: Parallel FP-Growth for Query Recommendation, In: Proceeding of the 2008 ACM conference on Recommender systems, Lausanne, Switzerland, 107-114, 2008.

[26] K.B. Lassen and W.M.P. van der Aalst, Complexity Metrics for Workflow Nets. *Information and Software Technology*, (3):610-626, 2009.

[27] E. Maguire, P. Rocca-Serra, S.-A. Sansone, J. Davies, and M. Chen, Visual Compression of Workflow Visualizations with Automated Detection of Macro Motifs, IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, 2013.

[28] C.W. Günther and W.M.P. van der Aalst. Fuzzy Mining: Adaptive Process Simplification Based on Multi-perspective Metrics. In G. Alonso, P. Dadam, and M. Rosemann, editors, *International Conference on Business Process Management (BPM 2007)*, volume 4714 of *Lecture Notes in Computer Science*, pages 328-343. Springer-Verlag, Berlin, 2007.

[29] L. Sánchez-González, F. García, J. Mendling, F. Ruiz, and M. Piattini, Prediction of business process model quality based on structural metrics, in Proceedings of the 29th international conference on Conceptual modeling, pp. 458–463, 2010.

[30] J. Mendling, H.A. Reijers, and W.M.P. van der Aalst, Seven Process Modeling Guidelines (7 pmg), Information and Software Technology, vol. 52, no. 2, pp. 127-136, 2010.

[31] K. Kluza, G.J. Nalepa, Proposal of Square Metrics for Measuring Business Process Model Complexity, Proceedings of the Federated Conference on Computer Science and Information Systems pp. 919–922, 2012.

[32] S. Forster, J. Pinggera, Collaborative Business Process Modeling, in EMISA, Vol. 206, pp. 81-94, 2012.

[33] J. Shen, E. Fitzhenry, Th.G. Dietterich, Discovering Frequent Work Procedures from Resource Connections, in Proceedings of the 14th international conference on intelligent user interfaces, pp. 277-286, 2009.

[34] S. Choobdar, F. Silva, and P. Ribeiro, Network node label acquisition and tracking, 15th Portuguese Conference on Artificial Intelligence, pp 418-430, Springer LNCS Vol. 7026, 2011.

[35] C.W. Günther, S. Rinderle, M. Reichert, W.M.P. van der Aalst, and J. Recker, Using Process Mining to Learn from Process Changes in Evolutionary Systems, *International Journal of Business Process Integration and Management*, 3(1):61-79, 2008.

[36] A. Rozinat, R.S. Mans, M. Song, W.M.P. van der Aalst, Discovering simulation models, in Jounal of Information Systems ,34: 305–327, 2009.

[37] M. Fernández-Ropero, H.A. Reijers, R. Pérez-Castillo and M. Piattini, Repairing Business Process Models as Retrieved from Source Code, BMMDS/EMMSAD, 94-108, 2013.

[38] M.L. Rosa, P. Wohed, J. Mendling, Arthur H.M. ter Hofstede, Hajo A. Reijers and Wil M.P. van der Aalst, Managing Process Model Complexity via Abstract Syntax Modifications, IEEE Transactions on Industrial Informatics, Vol. 7, No. 4, 2011.

[39] S.K.L.M.V. Broucke, J.D. Weerdt, J. Vanthienen, B. Baesens, A Comprehensive Benchmarking Framework (CoBeFra) for conformance analysis between procedural process models and event logs in ProM. In Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2013).

[40] M.L. Rosa, M. Dumas, R. Uba, R. Dijkman, Business process model merging: an approach to business process consolidation, ACM Transactions on Software Engineering and Methodology 22 (2), 2012.

[41] E. Lord, A.B. Diallo, and V. Makarenkov, Classification of bioinformatics workflows using weighted versions of partitioning and hierarchical clustering algorithms, BMC Bioinformatics, 16(68), 2015.

[42] M. Reichert, P. Dadam, Supporting Dynamic Changes of Workflows without Loosing Control, in Journal of Intell. Inf. Syst., 93-129, 1998.

[43] R. Jin, S. McCallen, E. Almaas, Trend Motif: A Graph Mining Approach for Analysis of Dynamic Complex Networks, Seventh IEEE International Conference on Data Mining (ICDM 2007). IEEE, 2007.

[44] D. Fahland, W.M.P. van der Aalst, Model Repair - Aligning Process Models to Reality, in Journal of Information Systems, 2013.

[45] M. Fernández-Ropero, Assessing the Best-Order for Business Process Model Refactoring, in Proceedings of the 28th Annual ACM Symposium on Applied Computing, 1397-140, 2013.

[46] M. Baum, B. Balasingam, P. Willett, and U.D. Hanebeck, OSPA Barycenters for Clustering Set-Valued Data, In *Proceedings of the 18th International Conference on Information Fusion (Fusion 2015)*, Washington, USA, July 2015.

[47] A. Goldman and Y. Ngoko, On Graph Reduction for QoS Prediction of Very Large Web Service Compositions, 2012 IEEE Ninth International Conference on Services Computing, 2012.

[48] T. Koohi, M. Zahedi, Linear Merging Reduction: A Workflow Diagram Simplification Method, 8th International Conference on Information and Knowledge Technology, 2016.

[49] B.F. van Dongen, Bpi challenge 2011, 2011. URL http://dx.doi.org/10.4121/uuid: d9769f3d-0ab0-4fb8-803b-0d1120ffcf54.

[50] Dongen, B.F.v.: BPI challenge 2012. Dataset. http://dx.doi.org/10.4121/uuid: 3926db30-f712-4394-aebc-75976070e91f, 2012.

[51] A.K.A. Medeiros, Genetic Process Mining, PhD. Dissertation, TU Eindhoven, 2006.

[52] S. Choobdar, On the Characterization and Comparison of Complex Networks, PhD thesis, 015.

[53] W. van der Aalst, V. Rubin, H. Verbeek, B. van Dongen, E. Kindler, C.W. Gunther, Process mining: a two-step approach to balance between underfitting and overfitting. Software and System Modeling 9(1): 87–111, 2010.

[54] B. Ludascher, M. Weske, T. McPhillips, and Sh. Bowers, Scientific Workflows: Business as Usual, BPM 2009, Springer-Verlag Berlin Heidelberg, 2009.

[55] D. Garijo, P. Alper, Kh. Belhajjame, O. Corcho, Y. Gil, C. Goble, Common motifs in scientific workflows: An empirical analysis, Future Generation Computer Systems, 2013.

[56] D. Garijo, Ó. Corcho, Y. Gil, Detecting common scientific workflow fragments using templates and execution provenance, In the proceedings of the Seventh International Conference on Knowledge Capture, 2013.

[57] A. Schoknecht, N. Fischer, A. Oberweis, Process Model Search Using Latent Semantic Analysis, In International Conference on Business Process Management, pp. 283-295, Springer, 2016.

[58] J. Stoyanovich, B. Taskar, and S. Davidson, Exploring repositories of scientific workflows, In Proceedings of the 1st International Workshop on Workflow Approaches to New Data-centric Science, ACM, 2010.

[59] T. Koohi, and M. Zahedi, Scientific Workflow Clustering based On Motif Discovery, International Journal on Computer Science, Engineering and Information Technology (IJCSEIT), 7 (4), 2017.

[60] R. Dijkman, M. Dumas, B.v. Dongen, R. Kaarik, and J. Mendling, Similarity of Business Process Models: Metrics and Evaluation, Information Systems, 36(2): 498-516, 2011.

[61] J. Starlinger, B. Brancotte, S. Cohen-Boulakia, U. Leser, Similarity Search for Scientific Workflows, in Proceedings of the VLDB Endowment (PVLDB), VLDB Endowment, 7 (12): 1143-1154, 2014.

[62] F.E. Tosta, V. Braganholo, L.Murta, and M. Mattoso, Improving workflow design by mining reusable tasks, Journal on the Brazilian Computer Society, 2015.
[63] W. Medhata, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, in Ain Shams Engineering Journal, 5: 1093-1113, 2014.
[64] A. Polyvyanyy, W.M.P. van der Aalst, A.H.M. Ter Hofstede, M.T. Wynn, Impact-Driven Process Model Repair, ACM Trans. Softw. Eng. Methodol. 25(4), 2016.
[65] R. Conforti, M.L. Rosa and A.H.M. ter Hofstede, Filtering out Infrequent Behavior from Business Process Event Logs, IEEE Transactions on Knowledge and Data Engineering, 29(2), 2017.
[66] A. Augusto, R. Conforti, M. Dumas, M.L. Rosa, F.M. Maggi, A. Marrella, M. Mecella, A. Soo, Automated Discovery of Process Models from Event Logs: Review and Benchmark, 2018.
[67] R. Bergmann and G. Müller, Similarity-Based Retrieval and Automatic Adaptation of Semantic Workflows, Springer, 31-54, 2018.
[68] S.J. Pan and Q. Yang, a Survey on Transfer Learning, IEEE Transactions on Knowledge and Data Engineering, 2009.
[69] C.C. Aggarwal, editor. Data classification: algorithms and applications, CRC Press, 2014.
[70] H. Lee, A. Battle, R. Raina, A.Y. Ng, Efficient sparse coding algorithms, In Advances in neural information processing systems, pp. 801-808, 2007.
[71] S. Jialin Pan, X. Ni, J.T. Sun, Q. Yang, and Zh. Chen, Cross-Domain Sentiment Classification via Spectral Feature Alignment, WWW 2010, April 26–30, 2010.
[72] D. Hawkins, Identification of outliers, Chapman & Hall, Reading, 1980.
[73] D. Bollegala, D. Weir, and J. Carroll, Cross-domain sentiment classification using a sentiment sensitive dictionary. IEEE Transactions on Knowledge and Data Engineering, 25 (8): 1719-1731. ISSN 1041-4347, 2013.
[74] H.M.W. Verbeek and Christian W. Gunther. XES standard definition 2.0. Technical report, BPMcenter.org, July 2014. URL http://bpmcenter.org/wp-content/uploads/reports/2014/BPM-14-09.pdf. BPM Center Report BPM-14-09.
[75] A. Tikhonov, H. Parkinson, R. Petryszak, U. Sarkans, A. Brazma, ArrayExpress update-simplifying data submissions. Nucleic Acids Res 28:43:D1113–D1116, 2015. URL https://www.ebi.ac.uk/arrayexpress/, Last Accessed: 11 January 2018
[76] http://www.processmining.org/prom/tutorials, Last Accessed: 11 January 2018
[77] B.F. van Dongen, Bpi challenge 2017, 2017. URL http://dx.doi.org/10.4121/uuid: 5f3067df-f10b-45da-b98b-86ae4c7a310b
[78] I. Vanderfeesten, H.A. Reijers, and W.M.P. van der Aalst. Evaluating Workflow Process Designs using Cohesion and Coupling Metrics, Computers in Industry, 59(5):420-437, 2008.
[79] G. Janssenswillen, N. Donders, T. Jouck, B. Depaire, A comparative study of existing quality measures for process discovery, in Journal on Information Systems, 71: 1-15, 2017.

[80] J. Qiu, Q. Wu, G. Ding, Y. Xu and Sh. Feng, A survey of machine learning for big data processing, in Journal on Advances in Signal Processing, 2016.

[81] E. Deelman, T. Peterka, I. Altintas, Ch.D. Carothers, K.K. van Dam, K. Moreland, M. Parashar, L. Ramakrishnan, M. Taufer, and J. Vetter, The future of scientific workflows, International Journal of High Performance Computing Applications, pp. 1–17, 2017.

[82] W. Tan, M.B. Blake, I. Saleh, and S. Dustdar, Social-network-sourced big data analytics. IEEE Internet Computing, 17(5), pp.62-69. 2013.

[83] T. Koohi-Var, and M. Zahedi, Cross-domain graph based similarity measurement of workflows, Journal of Big Data, 5(1):18, 2018.

[84] T. Koohi-Var, and M. Zahedi, A Cloud based Architecture for Working on Big Data with Workflow Management, International Journal of Web & Semantic Technology (IJWesT), 9 (2), 2018.

[۸۵] کوهی‌وردهکردی، طاهره؛ زاهدی، مرتضی. تحلیل SWOT نـرم افـزار «نمـآ» بـرای بهینـه سـازی طراحی چندوجهی، هفتمین کنفرانس بین‌المللی مدیریت دانش و اطلاعات، ۱۳۹۴.

[86] Loreti, Daniela, Federico Chesani, Anna Ciampolini, and Paola Mello, A distributed approach to compliance monitoring of business process event streams, Future Generation Computer Systems, https://doi.org/10.1016/j.future.2017.12.043, 2018.

[87] T. Koohi-Var, and M. Zahedi, Cross-Domain Similarity Assessment for Workflow Improvement to Handle Big Data Challenge in Workflow Management, Journal of Big Data, 5(1):26, 2018. https://doi.org/10.1186/s40537-018-0135-6.

[88] M.S. Saiyadain, Organisational behavior, Tata McGraw-Hill Education, 2003.

[89] H.L. Gantt, Organizing for work. Harcourt, Brace and Howe, 1919.

[90] M. Yaghoubi, M. Zahedi, Resource allocation using task similarity distance in business process management systems, ICSPIS 2016, Amirkabir University of Technology, 2016.

[91] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, A consistent metric for performance evaluation of multi-object filters, IEEE Trans. Signal Process., vol. 86, no. 8, pp. 3447–3457, 2008.

Abstract:

Nowadays, workflows are used extensively in scientific experiments and organizational processes. Workflows have quickly found their way into complex industrial processes. Recently, workflow improvement has become a new emerging problem. Regardless of the dynamics and workflow changes, some of the recent workflow improvement methods are based on static workflow simplification; recent methods do not have the flexibility to predict the simplified version of the workflows. Workflow changes should be taken into consideration in measurement, and the decision about workflow simplification should be based on measurement updates from workflows. In the proposed method, workflow simplification is done using statistical analysis of workflow diagrams and event logs. For this purpose, first the most frequent workflow tags are extracted. This is done using similarity evaluation. The proposed method uses these tags to identify the application domain of workflow by training a learning model. Thus, the learning model can cluster workflows. For each cluster, one workflow is selected as a cluster candidate. In this workflow candidate, the most frequent tags have a higher probability to be represented. The results of the proposed method show that the use of statistical analysis is successful in estimating simpler workflow paths. In the proposed method the value of the workflow connection parameter is reduced to 0.18. This has reduced the conflict in the workflow by approximately 30%. The advantage of the proposed method is that this method can make acceptable decisions about simplifying and improving workflows, due to the fact that it does not rely on a certain range of workflows. In the construction of a database based on the evaluation of workflow similarities, it is possible to dynamically add data from this new workflow and predict new paths to the database. Hence, the automatic workflow simplification becomes possible. Whenever we encounter a new workflow that has not already been shown to be the same, we assign it to the cluster candidate most closely resembling it. Therefore, it is possible to predict the simplifying workflow paths that we have not seen so far. Estimating the data, on average up to 87%, leads to an accurate estimation of the domain of workflow.

Key words: Statistical Process Analysis, Workflow Management Systems, Big Data.

Shahrood University of Technology

Faculty: International Campus of Kharazmi

Process Simplification by Statistical Analysis of Similar Workflows

Tahereh Koohi-Var Dehkurdi

Supervisor:

Morteza Zahedi

Date: **September 2018**