نشریه علمی مطالعات مدیریت کسب و کار هوشمند- سال نهم ـ شماره ۳۳ـ پاییز ۱۳۹۹
مقاله پژوهشی/ صفحات ۲۴۶- ۲۱۷

مدل برنامه‌ریزی ریاضی فرایندکاوی در مسئله کشف گراف وابستگی

مریم توکلی زانیانی *
محمدرضا غلامیان **

چکیده

کشف فرآیندها شاخه‌ای از فرآیندکاوی است که با استفاده از رویدادهای ثبت‌شده در یک سیستم، مدل فرآیندی را به دست می‌آورد که به‌خوبی توصیف‌کننده رفتار سیستم باشد. با توجه به این‌که یکی از بااهمیت‌ترین و پرطرفدارترین دسته روش‌های کشف فرآیند، کشف فرآیند ابتکاری است و نظر به این‌که کیفیت خروجی ارائه‌شده توسط این دسته روش‌ها شدیداً به نحوه استخراج گراف وابستگی میان فعالیت‌ها ارتباط دارد، در این مقاله برای اولین بار به ارائه رویکردی برای تبدیل مسئله کشف گراف وابستگی به یک مسئله برنامه‌ریزی باینری و همچنین معرفی تابع هدفی پرداختیم که به‌صورت هم‌زمان شاخص‌های سازگاری بازپخش و دقت مدل را در نظر می‌گیرد. میزان بها دادن به هرکدام از این شاخص‌ها از طریق آستانه‌ای که کاربر نهایی مشخص می‌کند تعیین می‌شود. شاخص‌های ذکرشده از جمله بااهمیت‌ترین معیارهای سنجش کیفیت خروجی روش‌های کشف فرآیند می‌باشند و در واقع استفاده از این رویکرد مستقیماً بر ارتقاء شاخص‌های کیفیت مدل اثر دارد. رویکرد پیشنهادی همچنین دارای این قابلیت است که با معرفی محدودیت‌های مناسب، دانش حوزه را در فرآیند استخراج مدل دخیل نماید و همچنین مدل خروجی را به سمت ارتقاء احتمال سالم بودن آن هدایت نماید. این امر در مطالعه موردی یک شرکت واقعی که در این مقاله ارائه‌شده قابل مشاهده است. در مطالعه سازمان مورد اشاره، رویکرد پیشنهادی با استفاده از محدودیت‌های تعریف‌شده بر اساس دانش حوزه و قواعد ساختاری گراف وابستگی بر روی رویدادهای ثبتی حوزه بازاریابی شرکت اعمال گردیده و نتایج آن منعکس شده است.

کلید واژگان: فرآیندکاوی، کشف فرآیندها، گراف وابستگی، برنامه‌ریزی باینری.

---

* دانشجوی دکتری، مهندسی صنایع، دانشگاه علم و صنعت ایران، تهران، ایران.

** عضو هیئت‌علمی، دانشکده مهندسی صنایع، دانشگاه علم و صنعت ایران، تهران، ایران.((نویسنده مسئول)؛ Gholamian@iust.ac.ir

تاریخ دریافت: ۱۳۹۸/۱۱/۲۱ **تاریخ پذیرش:۱۳۹۹/۰۶/۱۰**

مقدمه

کشف فرآیندها شاخه‌ای از فرآیند کاوی و یک حوزه تحقیقاتی است که با استفاده از رویدادهای ثبت‌شده از اجرای فرآیندها به تحلیل پسینی فرآیندهای کسب‌وکار می‌پردازد. یک روش *کشف فرآیند* رویدادهای ثبت‌شده را به‌عنوان ورودی می‌گیرد و مدلی را تولید می‌کند که به بهترین نحو رفتار در رویدادهای ثبت‌شده را توصیف می‌کند. به‌عنوان‌مثال، الگوریتم α که در سال ۲۰۰۴ توسط وان-در- آلست و همکاران[1]، ارائه شد یک مدل شبکه پتری[2] را با توجه به رویدادهای ثبت‌شده می‌سازد.

اقدام به کشف فرآیند به چند منظور ممکن است صورت پذیرد. در سازمان‌هایی که از پیش مدل فرآیندی وجود ندارد با استفاده از روش‌های کشف فرآیند می‌توان مدل فرآیند در حال اجرا در سازمان را به دست آورد. در سازمان‌هایی که از پیش مدل فرآیند وجود دارد، استفاده از این روش‌ها و کشف مدل فرآیند واقعی می‌تواند بینش درستی در مورد این‌که فرآیند واقعی تا چه حد به فرآیند از پیش تعیین‌شده منطبق است و چه تفاوت‌هایی وجود دارد فراهم کند. درصورتی‌که مدل فرآیند از پیش تعیین‌شده مدل قانونی باشد، کشف انحراف از مدل در واقع معنای کشف تقلب و بی‌قانونی را دارد. درصورتی‌که مدل از پیش تعیین‌شده مدل توصیفی باشد کشف عدم مطابقت‌ها و ریشه‌یابی دلایل آن‌ها می‌تواند کمک شایانی به بهبود عملکرد سیستم نماید.

برای سنجش کیفیت مدل استخراج‌شده توسط روش‌های فرآیندکاوی نیز معیارهای مختلفی وجود دارد که مهم‌ترین آن‌ها سازگاری، بازپخش[3] و دقت[4] است. سازگاری اندازه می‌گیرد که تا چه اندازه مدل کشف‌شده می‌تواند به‌صورت دقیق موارد ثبت‌شده در وقایع ثبتی را بازتولید کند. برخلاف سازگاری، دقت نسبت رفتار مجاز توسط مدل که در رویدادهای ثبتی مشاهده نشده‌اند را نشان می‌دهد در واقع مدل با دقت پایین یک مدل با برازش کم[5] است (سارنو و سانگکونو[6]، ۲۰۱۹)؛ بنابراین اتخاذ روشی که هنگام استخراج مدل، بهینه کردن شاخص‌های

1. Van der Aalst et al.
2. Petri net
3. Replay fitness
4. Precision
5. Underfitted
6. Sarno & Sungkono

سنجش کیفیت مدل را در نظر داشته باشد رویکردی کارآمد در زمینه استخراج مدل فرآیند به نظر می‌رسد.

از سوی دیگر یکی از عوامل رایج در دنیای واقعی نویز است. نویز می‌تواند تحت دو شرایط ظاهر شود: دنباله رویدادهایی که به‌نوعی به‌صورت نادرست ثبت شده‌اند (به‌عنوان‌مثال به دلیل درست پیکره‌بندی نشدن موقتی سیستم) و یا دنباله رویدادهایی که موقعیت‌های استثنایی را نشان می‌دهند. به‌طور خلاصه می‌توانیم بگوییم نویز هر رفتار با دفعات تکرار پایین در رویدادهای ثبت شده است. بدیهی است که برای این‌که یک روش برای داده‌های واقعی بتواند به‌درستی کاربرد داشته باشد باید در برابر نویز مقاوم باشد.

یکی از پرطرفدارترین الگوریتم‌های کشف فرآیند که قادر به سروکار داشتن با داده‌های دارای نویز است و از سرعت بالا برخوردار است، الگوریتم کاوشگر ابتکاری است که توسط ویجترز و همکاران[1] در سال ۲۰۰۶ ارائه شد. الگوریتم کاوشگر ابتکاری دو مرحله اصلی دارد. در مرحله اول گراف وابستگی ساخته می‌شود. این گراف وابستگی شامل روابط وابستگی می‌شود که زمان ساخت مدل شبکه پتری در نظر گرفته خواهند شد. در مرحله دوم، مفهوم نقاط انشعاب/ اتصال در گراف وابستگی تعیین می‌شوند؛ یعنی با استفاده از الگوهایی در خصوص دفعات تکرار وابستگی‌ها و در مورد نوع انشعاب/ اتصال و وابستگی‌های با فاصله زیاد تصمیم گرفته می‌شود. تاکنون توسعه‌های فراوانی برای الگوریتم مذکور انجام شده است اما در تمامی توسعه‌های مورد اشاره در مرحله ساخت گراف وابستگی، بهینه کردن شاخص‌های سنجش کیفیت مدل در نظر گرفته نشده است و همچنین ماهیت این روش‌ها به‌گونه‌ای است که امکان استفاده از دانش حوزه در هنگام استخراج گراف وابستگی وجود ندارد. از سوی دیگر احتمال زیادی برای غیرسالم[2] بودن مدل‌های کشف‌شده توسط روش‌های مذکور وجود دارد بنابراین استفاده از رویکردی که بتواند گراف وابستگی را به سمت سالم بودن هدایت نماید و در کنار ایجاد امکان بهینه کردن شاخص‌های سنجش کیفیت مدل، بتواند دانش حوزه را در فرآیند استخراج مدل دخیل نماید، امکان ایجاد بهبود چشمگیری در این روش را فراهم می‌نماید.

با توجه به موارد مذکور، استفاده از روش‌های مدل‌سازی ریاضی و برنامه‌ریزی باینری یک رویکرد مناسب جهت کشف گراف وابستگی است. استفاده از این رویکرد نه‌تنها بهینه

1. Weijters et al.
2. Sound

کردن معیارهای سنجش کیفیت مدل را در حین کشف گراف وابستگی مقدور می‌سازد، بلکه این امکان را فراهم می‌نماید تا مقاومت در برابر نویز، الزام سالم بودن مدل و استفاده از دانش حوزه از طریق نحوه تعریف تابع هدف و محدودیت‌های مدل تعیین شود.

استفاده از برنامه‌ریزی ریاضی در کشف فرآیندها تازگی ندارد و پژوهش‌های فراوانی بر روی کشف فرآیندها با استفاده از این رویکرد انجام شده است. دسته اول این پژوهش‌ها از نظریه مناطق[1] استفاده می‌کنند و از مدل برنامه‌ریزی ریاضی برای کشف مکان‌ها[2] در شبکه پتری بهره می‌گیرند به‌عنوان‌مثال می‌توان به پژوهش وان در ورف و همکاران[3] در سال ۲۰۰۸ و وان زلست و همکاران[4] در سال‌های ۲۰۱۵ و ۲۰۱۸ اشاره کرد. کلیه این پژوهش‌ها از مدل برنامه‌ریزی ریاضی برای کشف مکان‌ها در شبکه پتری بهره می‌گیرند و برای کشف هر مکان، یک بار مدل برنامه‌ریزی عدد صحیح خطی اجرا می‌شود؛ بنابراین در هر بار اجرای مدل، بهینه‌سازی به‌صورت محلی اتفاق می‌افتد. دسته دوم روش‌ها مانند پژوهش‌های پرودل و همکاران[5] در سال‌های ۲۰۱۵ و ۲۰۱۸ از برنامه‌ریزی ریاضی برای کشف فرآیندها در هنگام مواجهه با داده‌های با نویز بالا و فرآیندهای متغیر استفاده کرده‌اند. مدل فرآیند خروجی این پژوهش‌ها گرافی ساده است و این پیش‌فرض وجود دارد که کلیه اتصالات ورودی[6] و انفصالات خروجی[7] گره‌ها[8] از نوع انفصال انحصاری[9] (XOR) هستند؛ بنابراین استخراج ساختارهای پیچیده‌تر توسط این تحقیقات پشتیبانی نمی‌شود. قابل‌ذکر است که علی‌رغم این‌که سنجش کیفیت مدل‌های استخراج‌شده نهایتاً توسط معیارهای سنجش کیفیت مدل مانند سازگاری و دقت انجام می‌شود، تابع هدف هیچ‌کدام از دو دسته اشاره‌شده در راستای بهینه‌سازی این شاخص‌ها نیست.

بنابراین بر اساس جست‌وجوهای انجام شده، علی‌رغم قابلیت‌های بالقوه و مزایای ذکرشده موجود در این حوزه، تاکنون پژوهشی در راستای استفاده از برنامه‌ریزی ریاضی در مسئله کشف گراف وابستگی انجام نشده است و همچنین هدف هیچ‌یک از پژوهش‌های موجود

1. Theory of Regions
2. Place
3. Van der Werf et al.
4. Van Zelst et al.
5. Prodel et al.
6. Incoming joins
7. Outgoing splits
8. Nodes
9. Exclusive Disjunction

در زمینه کاربرد برنامه‌ریزی ریاضی در کشف فرآیندها در راستای بهینه‌سازی شاخص‌های سنجش کیفیت مدل نبوده است. لذا در این راستا این مقاله برای اولین بار به ارائه رویکردی جهت تبدیل مسئله کشف گراف وابستگی به مسئله برنامه‌ریزی باینری و همچنین معرفی تابع هدف مربوط به این مسئله در راستای بهینه‌سازی شاخص‌های سازگاری و دقت مدل پرداخته است. قابل‌ذکر است که با توجه به قابلیت‌های ذاتی برنامه‌ریزی ریاضی، با در دست داشتن رویکرد و تابع هدف مربوطه، امکان تعریف محدودیت‌های مدل بر اساس دانش حوزه و نظر کارشناسان خبره و همچنین ماهیت گراف وابستگی وجود دارد. ضمن آن‌که این امر در پژوهش موردی استخراج مدل فرآیند حوزه بازاریابی شرکت بهسازان توسعه آرمان (بتا) که در زمینه تولید مواد پوشش‌های ضد حریق و عایق سرد PIR مشغول به کار است به کار گرفته شده و نتایج آن در این مقاله ارائه گردیده است.

مبانی نظری و پیشینه پژوهش

در زمینه کشف فرآیند، الگوریتم α را می‌توان یکی از اساسی‌ترین روش‌ها در این حوزه در نظر گرفت. این روش بر اساس چکیده‌ای از رویدادهای ثبت شده، یک مدل، مثلاً یک شبکه پتری می‌سازد. این روال کاویدن شامل سه مرحله است: مرحله چکیده کردن، مرحله استنتاج و مرحله ساخت. در مرحله چکیده کردن، هر رویداد در هر یک از دنباله رویدادهایی که در رویدادهای ثبت شده وجود دارد جست‌وجو می‌شود و روابط ترتیبی پایه‌ای بین کارها ثبت می‌شود. در مرحله استنتاج، روابط ترتیب پیشرفته از روابط پایه نتیجه‌گیری می‌شوند و کارهای جدید (به‌عنوان‌مثال کارهای نامرئی و کارهای دوگانه) ممکن است از طریق چرک‌نویس‌ها تولید شوند. در مرحله ساختن، مدل نهایی با توجه به چند قانون ابتکاری از طریق روابط ترتیب پیشرفته ساخته می‌شود. الگوریتم α فرض می‌کند که مدل بالقوه یک شبکه پتری مربوط به جریان کار ساخت‌یافته (شبکه SWF[1]) سالم است که فاقد حلقه‌های کوتاه است. این الگوریتم، اولین الگوریتم از سری α بود و سایر الگوریتم‌های این سری از جمله کارهای آلوس دی مدیروس و همکاران[2] در ۲۰۰۵، هی و همکاران[3] در ۲۰۱۸، لی و همکاران[4] در ۲۰۰۷، روسی

1. Structured Workflow Net
2. Alves de Medeiros et al.
3. He et al.
4. Li et al.

و داویدراجو[1] در ۲۰۱۸، سان و همکاران[2] در ۲۰۱۹، ون و همکاران[3] در ۲۰۰۷ و ۲۰۰۹ توسعه‌ای از آن هستند. هدف این توسعه‌ها بسط دادن طبقه شبکه‌هایی است که مدل بالقوه می‌تواند به آن تعلق داشته باشد. برخی دیگر از الگوریتم‌های سری آلفا مانند پژوهش املیا افندی و سارنو[4] در ۲۰۲۰ بر قابلیت استفاده از رویدادهای ثبتی با زمان‌های آغاز و پایان ثبت شده برای هر رویداد تمرکز دارند.

با وجود این‌که الگوریتم‌های متفاوتی که در قسمت الگوریتم‌های سری α عنوان کردیم قادر به سر و کار داشتن با خیلی از ساختارهای نمودار کنترل هستند، اما همه آن‌ها این مشکل را دارند که قادر به سروکار داشتن با نویز که یک عامل رایج در دنیای واقعی است، نیستند. الگوریتم‌های سری α ارائه‌شده در زمان استنتاج روابط ترتیب پیشرفته برای ساخت مدل فرآیند، تعداد روابط ترتیب را در نظر نمی‌گیرند؛ بنابراین، برای حل این مشکل در سال ۲۰۰۶ الگوریتم کاوشگر ابتکاری[5] توسط ویجترز و همکاران ارائه شد. این الگوریتم از طریق ارائه معیارهای بر پایه تکرار فعالیت‌ها با نویز مقابله می‌کند (ویدگوف و همکاران[6]، ۲۰۲۰). کاوشگر ابتکاری می‌تواند در حضور نویز کار کند و برای نشان دادن رفتار اصلی در وقایع ثبت شده (مثلاً بدون نشان دادن جزئیات و استثناها) می‌تواند به کار رود. این روش می‌تواند مجموعه وسیعی از ساختارهای مدل فرآیند را استخراج کند. فعالیت‌های نامرئی (فعالیت‌هایی که در مدل وجود دارند اما در رویدادهای ثبتی وجود ندارند)، مستقیماً به همین شکل کاویده نمی‌شوند اما شبکه ابتکاری کاویده شده نیازی به حضور فعالیت‌های نامرئی برای مدل کردن پرش از روی فعالیت‌ها و یا ساختارهای پیچیده مسیریابی ندارد (بعد از تبدیل به شبکه‌های پتری مدل شامل فعالیت‌های نامرئی لازم برای نشان دادن چنین ساختارهایی خواهد بود).

در سال ۲۰۱۲ توسط بوراتین و همکاران[7] یک توسعه از کاوشگر ابتکاری ارائه شد که قادر به سر و کار داشتن با داده‌های جریان رویدادها[8] بود. در سال ۲۰۱۵ نیز بوراتین توسعه دیگری از کاوشگر ابتکاری به نام کاوشگر ابتکاری ++ ارائه نمود که فعالیت‌ها را به همراه بازه

1. Roci & Davidrajuh
2. Sun et al.
3. Wen et al.
4. Amelia Effendi & Sarno
5. Heuristics Miner
6. Vidgof et al.
7. Burattin et al.
8. Streaming Event Data

زمانی‌شان در نظر می‌گرفت یعنی به‌جای این‌که برای هر رویداد تنها یک نقطه زمانی در نظر بگیرد، در نظر می‌گرفت که رویداد در یک بازه زمانی اتفاق افتاده است و از اطلاعات زمان شروع و پایان آن رویداد برای رسیدن به مدل فرآیند دقیق‌تر استفاده می‌کرد. در سال ۲۰۱۱ نیز ویجترز و روبیرو[1] یک نسخه بهبودیافته از کاوشگر ابتکاری به نام کاوشگر ابتکاری منعطف[2] را ارائه داد که خروجی آن به‌صورت یک شبکه علی بود. با وجود این‌که این نمادسازی بسیار شبیه شبکه‌های ابتکاری است یک تفاوت مهم در نحوه بیان قیدهای ورودی و خروجی هر فعالیت وجود دارد. در سال ۲۰۱۷ توسط واندن بروک و دی ویردت[3] یک توسعه از الگوریتم کاوشگر ابتکاری به نام Fodina ارائه گردید که بر روی پایداری در برابر نویز و انعطاف‌پذیری تمرکز دارد. این الگوریتم رویکرد پایدارتری در خصوص داده‌های دارای نویز ارائه داده و قادر به کشف فعالیت‌های تکراری[4] است و همچنین دارای امکان گزینه‌های پیکربندی منعطف برای هدایت روال کشف فرآیند با توجه به ورودی‌های ارائه‌شده توسط کاربر نهایی است.

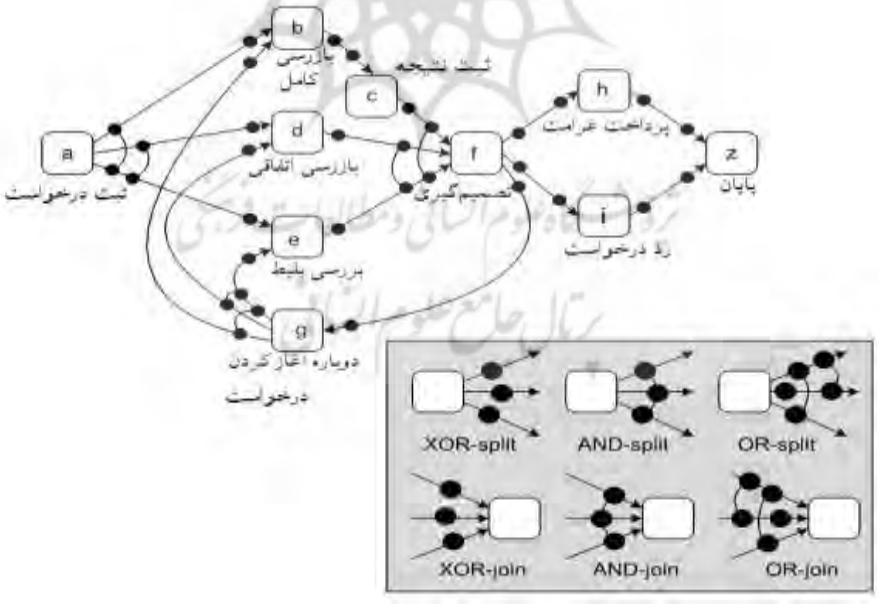
با وجود این‌که از نظر تئوریک پایداری در برابر نویز بایستی به‌صورت ضمنی منجر به بیشتر شدن دقت مدل می‌شود اما کماکان Fodina بر بهینه کردن دقت و یا سازگاری مدل تمرکز ندارد و تضمینی بر این امر که تغییر پارامترهای روش منجر به بهبود شاخص‌های کیفیت مدل شود وجود ندارد. از سوی دیگر وجود گزینه‌های پیکره‌بندی منعطف منجر به دستیابی به برخی از ساختارهای مدنظر کاربر می‌شود ولی هنوز محدودیت‌های فراوانی در رابطه با نوع ساختارهایی که کاربر می‌تواند تعیین کند وجود دارد و عملاً امکان به‌کارگیری انواع محدودی از دانش حوزه در استخراج مدل فرآیند وجود دارد. به‌عنوان‌مثال، کاربر نمی‌تواند تعیین کند که در مدل خروجی دو فعالیت از پیش تعیین‌شده با یکدیگر توالی نداشته باشند؛ بنابراین توسعه روش‌هایی که قادر به حل مشکلات فوق‌الذکر باشند ضروری به نظر می‌رسد.

معرفی شبکه‌های علی

شبکه‌های علی یک نحوه نمایش مدل فرآیند هستند که مناسب فرآیندکاوی هستند (وان در آلست، ۲۰۱۴). یک شبکه علی یک گراف است که در آن گره‌ها نمایش‌دهنده فعالیت‌ها و

1. Weijters & Ribeiro
2. Flexible Heuristics Miner
3. vanden Broucke & De Weerdt
4. Duplicate Activities

کمان‌ها نمایش‌دهنده روابط علی هستند. هر فعالیت یک مجموعه از قیدهای[1] ورودی ممکن و مجموعه‌ای از قیدهای خروجی‌های ممکن دارد. به‌عنوان‌مثال، شبکه علی نمایش داده‌شده در شکل ۱ که مربوط به مدل فرآیند رسیدگی به درخواست‌های پرداخت غرامت در یک شرکت بیمه است را در نظر بگیرید. مجموعه قیدهای ورودی فعالیت a مجموعه تهی است، زیرا فعالیت آغازین است. برای خروجی‌ها فعالیت a دو مجموعه قید ممکن داریم: {b,e} و {d,e}. این بدان معنی است که پس از فعالیت a یا b و e به‌صورت موازی با هم (هم‌زمان) انجام می‌شوند و یا d و e به‌صورت موازی با هم انجام می‌شوند. فعالیت f دو قید ورودی ممکن ({c,e} و {d,e}) و سه قید خروجی ممکن ({g}, {h}و {i}) دارد؛ بنابراین، قبل از f یا c و e به‌صورت موازی انجام می‌شوند و یا d و e به‌صورت موازی انجام می‌شوند و پس‌ازآن یکی از g، h و یا i انجام می‌شود. فعالیت z فعالیت پایانی است که دو قید ورودی و قید خروجی تهی دارد. این فعالیت برای تولید یک نقطه پایان یکتا اضافه شده است. فرآیند رسیدگی به هر درخواست با فعالیت شروع a (ثبت درخواست) آغاز می‌شود و با فعالیت پایانی z (پایان) خاتمه می‌یابد. قابل‌توجه است که برخلاف شبکه‌های پتری، هیچ مکانی در شبکه‌های علی وجود ندارد و منطق مسیریابی صرفاً بر اساس قیدهای ورودی و خروجی است.



شکل ۱: شبکه علی فرآیند رسیدگی به درخواست‌های پرداخت غرامت در یک شرکت بیمه

1. Binding

تعریف ۱(شبکه‌های علی): یک شبکه علی چندتایی $C=(A, a_i, a_o, D, I, O)$ است به‌نحوی که:

- A یک مجموعه محدود از فعالیت‌ها، $\mathbf{a_i \epsilon A}$ فعالیت آغازین و $\mathbf{a_o \epsilon A}$ فعالیت پایانی است.
- $\mathbf{D \subseteq A \times A}$ رابطه وابستگی است.
- $\mathbf{\rho(A) = \{A' | A' \subseteq A\}}$ به‌نحوی‌که $\mathbf{AS = \{X \subseteq \rho(A) | X = \{\emptyset\} \vee \emptyset \notin X\}}$ مجموعه توانی **A** است. بنابراین اعضا **AS** مجموعه‌ای از مجموعه‌های فعالیت‌ها هستند.
- $I \in A \rightarrow AS$ مجموعه قیدهای ورودی ممکن برای هر فعالیت را مشخص می‌کند.
- $\boldsymbol{O \in A \rightarrow AS}$ مجموعه قیدهای خروجی ممکن برای هر فعالیت را مشخص می‌کند.

به‌نحوی که:

- $\boldsymbol{D = \{(a_1, a_2) \in A \times A | a_1 \in \cup_{as \in I(a_2)} as\}}$
- $\boldsymbol{D = \{(a_1, a_2) \in A \times A | a_2 \in \cup_{as \in O(a_1)} as\}}$
- $\boldsymbol{\{a_o\} = \{a \in A | O(a) = \{\emptyset\}\}}$ و $\boldsymbol{\{a_i\} = \{a \in A | I(a) = \{\emptyset\}\}}$
- تمامی فعالیت‌ها در گراف **(A,D)** در مسیری از $\boldsymbol{a_i}$ به $\boldsymbol{a_o}$ قرار داشته باشند

روش‌های کشف فرآیند ابتکاری

در این قسمت به هسته اصلی روش‌های کشف فرآیند ابتکاری مانند کاوشگر ابتکاری و کاوشگر ابتکاری منعطف می‌پردازیم. در حالت کلی این روش‌ها در چهار گام اجرا می‌شوند.

گام اول به دست آوردن روابط پایه‌ای میان فعالیت‌ها در رویدادهای ثبتی است. با فرض L به‌عنوان رویدادهای ثبتی، A_L به‌عنوان مجموعه محدود فعالیت‌های اتفاق افتاده در L و همچنین a و b به‌عنوان فعالیت‌های عضو A_L اطلاعات زیر از رویدادهای ثبتی قابل‌استخراج است:

$|a|$ نشان‌دهنده تعداد دفعات اتفاق افتادن فعالیت a در رویدادهای ثبتی است (فراوانی a).

$|a > b|$ نشان‌دهنده تعداد دنبال کردن‌های مستقیم a توسط b است یعنی دفعاتی در لاگ رویدادها است که فعالیت a مستقیماً (یعنی بلافاصله) توسط فعالیت b دنبال شده است.

$|a >> b|$ نشان‌دهنده تعداد دفعاتی در رویدادهای ثبتی است که تکرار میان a و b اتفاق افتاده است. (یعنی پس از a بلافاصله b و سپس دوباره a اتفاق افتاده است).

$|a >>> b|$ نشان‌دهنده تعداد دنبال کردن‌های غیرمستقیم a توسط b است، یعنی تعداد دفعاتی در رویدادهای ثبتی است که فعالیت a نهایتاً توسط فعالیت b دنبال شده است (قبل از ظاهر شدن بعدی a و یا b). توجه کنید که تعداد هر دنبال کردن مستقیم در دنبال کردن‌های غیرمستقیم نیز شمرده می‌شود.

گام دوم شامل ساخت گراف وابستگی است با استفاده از معیارهای وابستگی و یا شاخص‌های علی که روابط علی پایه‌ای بین فعالیت‌ها (دنبال کردن‌ها و مقدم بودن‌ها) را توصیف می‌کند. برای سنجش میزان وابستگی میان دو فعالیت معیارهای مناسب متنوعی در ادبیات ارائه‌شده که بسیاری از آن‌ها در قالب الگوریتم ابتکاری بر پایه وابستگی می‌باشند. به‌عنوان‌مثال، می‌توان به مقالات واندن بروک و دی ویردت در سال ۲۰۱۷، ویجترز و همکاران در ۲۰۰۶ و ویجترز و ریبیرو در ۲۰۱۱ اشاره کرد. حال‌آنکه در برخی از تحقیقات مرتبط نیز از این مفهوم استفاده شده است، به‌عنوان‌مثال، پژوهش ماروستر و همکاران[1] در سال ۲۰۰۶ چنین معیاری را برای ساخت یک مجموعه داده به کار برده است که از آن برای یادگیری قوانین استفاده می‌شود. لی و همکاران[2] نیز در سال ۲۰۱۹ برای ساخت یک ماتریس وابستگی بر اساس ماتریس تکرار بهبود یافته، معیار وابستگی الگوریتم کاوشگر ابتکاری را بهبود دادند و از این امر جهت از بین بردن روابط غیرقطعی در رویدادهای ثبتی استفاده کردند.

پژوهش‌های ویجترز و همکاران در سال ۲۰۰۶ و ویجترز و روبیرو در سال ۲۰۱۱ از فرمول زیر برای محاسبه $|a \rightarrow b|$ که نشانگر میزان رابطه وابستگی میان a و b است، استفاده نموده‌اند:

$$|a \rightarrow b| = \begin{cases} \dfrac{|a > b| - |b > a|}{|a > b| + |b > a| + 1} & if\ a \neq b \\ \dfrac{|a > a|}{|a > a| + 1} & if\ a = b \end{cases} \quad (۱)$$

1. Măruşter et al.
2. Li et al.

واندن بروک و دی ویردت در سال ۲۰۱۷ نیز با این ایده که دنبال کردن مستقیم فعالیت a توسط فعالیت b همیشه شاهد مناسبی بر ضد دنبال کردن مستقیم فعالیت b توسط فعالیت a نیست، معیار مذکور را به‌صورت زیر تغییر داده است.

$$|a \to b| = \begin{cases} \dfrac{|a > b|}{|a > b| + |b > a| + 1} & if\ a \neq b \\ \dfrac{|a > a|}{|a > a| + 1} & if\ a = b \end{cases} \qquad (۲)$$

در **گام سوم** مجموعه قیدهای ورودی و خروجی برای هر فعالیت (که نمایانگر انشعاب‌ها و اتصالات XOR و AND هستند) کاویده می‌شود. در تعریف اصلی الگوریتم کاوشگر ابتکاری یک معیار جداگانه برای به دست آوردن اطمینان از اتفاق افتادن دو فعالیت به‌صورت موازی مورد استفاده قرار گرفته است. در کاوشگر ابتکاری منعطف روابط XOR و AND به طرق متفاوت کاویده می‌شوند. ایده کلی روش به این صورت است که برای یک فعالیت a با فعالیت‌های وابسته b و c تعداد دفعاتی که a فقط توسط b، فقط توسط c و توسط هر دو b و c دنبال شده است شمرده می‌شود. بر اساس فراوانی هر یک از الگوهای ممکن (یعنی $\{\{b\},\{c\}\}$ و $\{\{b,c\}\}$) قیدهای خروجی تعیین می‌شوند.

در **گام چهارم** که اختیاری است وابستگی‌های با فاصله زیاد کاویده می‌شوند. برای این منظور، یک شاخص و آستانه وابستگی دیگر با استفاده از مقدار $|a \gg > b|$ استفاده می‌شود.

در الگوریتم‌های کاوش ابتکاری، زمانی که معیار وابستگی برای دو فعالیت بالاتر از آستانه تعیین‌شده توسط کاربر باشد، یک وابستگی (یک کمان میان دو فعالیت) به گراف وابستگی بین دو فعالیت اضافه می‌شود. استفاده از معیار وابستگی جهت ساخت گراف وابستگی، توان مقابله با نویز این دسته روش‌ها را به نحو چشمگیری افزایش می‌دهد. علی‌رغم این‌که این امر مزیت سرعت محاسباتی بالا و پیچیدگی کم را به همراه دارد ولی دارای معایبی به شرح آنچه در بخش ۲,۱ اشاره شد، است. از جمله آن‌که با وجود اینکه خروجی نهایی الگوریتم با شاخص‌هایی در مورد سازگاری بازپخش و دقت مدل سنجیده می‌شوند، در خصوص قرار دادن و یا حذف هر کمان به‌صورت محلی و فارغ از تأثیری که بر روی شاخص‌های مذکور و کل گراف خواهد گذاشت (به‌عنوان‌مثال، ایجاد

ناهنجاری‌هایی از قبیل بن‌بست) بررسی و تصمیم‌گیری می‌شود. همچنین امکان استفاده از دانش ارزشمند کارشناسان و خبرگان سازمان در روال کشف گراف وابستگی وجود ندارد.

از سوی دیگر استفاده از رویکرد برنامه‌ریزی ریاضی می‌تواند در راستای اصلاح نقایص مذکور مزایای فراوانی داشته باشد از جمله این‌که می‌تواند در استخراج مدل فرآیند پیشنهادی مستقیماً بهینه کردن معیارهای سنجش کیفیت مدل را هدف قرار دهد و با تعریف محدودیت‌های مناسب، امکان دخیل کردن دانش حوزه در روال کشف گراف به‌سادگی وجود خواهد داشت. همچنین می‌توان محدودیت‌هایی در راستای برآورده کردن الزامات ساختاری و تضمین سالم بودن گراف حاصله، به مدل ریاضی اضافه نمود. قابل‌ذکر است که استفاده از این رویکرد می‌تواند در مواجهه با مدل‌های فرآیند بسیار بزرگ از مشکل زمان حل بالا و پیچیدگی محاسباتی رنج ببرد؛ ولی در عمل در بسیاری از موارد حصول به نتیجه با کیفیت بهتر، بر زمان حل پایین برتری دارد. در جمع‌بندی اثر علمی این تحقیق، تبدیل مسئله کشف گراف وابستگی به یک مسئله برنامه‌ریزی باینری با هدف بهینه‌سازی دو معیار مهم سنجش کیفیت مدل (سازگاری و دقت) است که در قالب توسعه یک تابع هدف ریاضی همراه با محدودیت‌های ساختاری ارائه شده است. لذا با توجه به موارد مطروحه در قسمت بعدی به ارائه رویکردی برای تبدیل مسئله کشف گراف وابستگی به یک مسئله برنامه‌ریزی باینری و ارائه تابع هدف پیشنهادی می‌پردازیم.

مدل ریاضی پژوهش

متغیرهای مدل و نحوه تبدیل مسئله کشف گراف وابستگی به برنامه‌ریزی باینری

با توجه به نقایص ذکرشده روش‌های موجود کشف گراف وابستگی و قابلیت‌های بالقوه برنامه‌ریزی ریاضی جهت رفع آن‌ها، در این بخش برای اولین بار رویکردی برای تبدیل مسئله کشف گراف وابستگی، به مسئله برنامه‌ریزی ریاضی ارائه می‌گردد و همچنین تابع هدف مرتبط با این رویکرد به نحوی پیشنهاد می‌گردد که در راستای بهینه کردن مهم‌ترین معیارهای سنجش کیفیت مدل، یعنی سازگاری و دقت مدل، باشد.

در همین راستا، هدف مدل ریاضی را به دست آوردن روابط پیش‌نیازی و موازی بودن میان فعالیت‌های $a \in A_L$ در نظر گرفته‌ایم که میزان تابع هدف ذکرشده در بخش بعدی را بهینه کنند. روابط پیش‌نیازی مستقیم میان هر دو فعالیت i و j (یعنی وجود کمان از سمت فعالیت i به فعالیت j در گراف وابستگی) توسط متغیر باینری $P_{i,j}$ نشان داده می‌شود.

درصورتی‌که این متغیر برابر با یک باشد به این معنی است که فعالیت i مستقیماً پیش‌نیاز فعالیت j است. متغیر باینری $C_{i,j}$ نیز وجود رابطه هم‌زمانی (موازی بودن) میان فعالیت‌ها را نشان می‌دهد. درصورتی‌که این متغیر برابر با یک باشد به این معنی است که فعالیت i موازی با فعالیت j است و بین آن‌ها رابطه هم‌زمانی وجود دارد. به ازای متغیر $C_{i,j}$ در گراف وابستگی هیچ عنصر فیزیکی مرئی وجود ندارد. ولی این متغیر در تعیین میزان شاخص‌های سنجش کیفیت مدل و همچنین در توسعه‌های آتی این پژوهش که بر روی کاوش قیدهای خروجی و ورودی و ارتقاء احتمال سالم بودن مدل خروجی تأکید دارند، نقشی اساسی خواهد داشت.

تابع هدف

در ادبیات کشف فرآیندها چهار بعد کیفیت وجود دارند که معمولاً برای بحث کردن در مورد نتایج تکنیک‌های کشف فرآیند به کار می‌روند. این چهار بعد شامل سازگاری بازپخش، دقت، عمومیت و سادگی هستند (بویجس و همکاران[1]، ۲۰۱۲ و ۲۰۱۴). ازآنجاکه در این میان دو شاخص سازگاری و دقت از اهمیت بیشتری برخوردار هستند، تابع هدف پیشنهادی با تمرکز بر این دو معیار ساخته می‌شود. از سوی دیگر این دو معیار در تضاد با یکدیگر هستند و بالا بردن سازگاری یک مدل منجر به پایین آمدن دقت آن و بالعکس بالا بردن دقت یک مدل منجر به پایین آمدن سازگاری آن می‌شود. لذا راهکار پیش‌بینی‌شده برای در نظر گرفتن هم‌زمان این دو شاخص، بیشینه کردن دقت مدل است درحالی‌که معیارهایی برای دستیابی به حد معقولی از سازگاری رعایت شود. بدیهی است که استفاده از این رویکرد منجر به پایداری بالای روش نسبت به نویز می‌شود. در این راستا ابتدا لازم است که در خصوص معیارهای فوق‌الذکر بحث بیشتری انجام شود؛ بنابراین در ابتدا معیارهای مذکور معرفی می‌شوند و سپس به معرفی تابع هدف پیشنهادی می‌پردازیم.

معیارهای سازگاری و دقت

شاخص سازگاری اندازه می‌گیرد که تا چه اندازه مدل کشف‌شده می‌تواند به‌صورت دقیق موارد ثبت شده در لاگ رویدادها را بازتولید کند. به این منظور از مفهوم دنبال کردن مستقیم استفاده می‌شود. به این صورت که هنگامی می‌توانیم بگوییم که فعالیت i مستقیماً توسط j دنبال

1. Buijs et al.

می‌شود که پس از i، بلافاصله j اتفاق بیافتد. می‌توان ادعا کرد که هر چه مدل اجازه وقوع تعداد بیشتری از دنبال کردن‌های مستقیم موجود در L را بدهد میزان سازگاری مدل بیشتر می‌شود و درصورتی‌که همه دنبال کردن‌های مستقیم موجود در L اجازه وقوع در مدل فرآیند را داشته باشند مدل به‌صورت کامل قابلیت بازپخش L را دارد.

از سوی دیگر شاخص دقت نشان می‌دهد که تا چه حد رفتارهای مجاز در مدل در لاگ رویدادها موجود است. در واقع هر چه میزان رفتارهایی که توسط مدل مجاز شناخته می‌شوند اما در لاگ رویدادها وجود ندارند بیشتر باشد، دقت مدل به‌دست‌آمده پایین‌تر است. جهت اندازه‌گیری معیار دقت نیز می‌توان از دنبال کردن‌های مستقیم مجاز توسط مدل و دنبال کردن‌های مستقیم موجود در لاگ رویدادها استفاده کرد. به این صورت که هر چه تعداد دنبال کردن‌های مستقیمی که توسط مدل مجاز هستند ولی در لاگ رویدادها موجود نیستند بیشتر باشد، دقت مدل کمتر می‌شود.

با استفاده از قواعد زیر می‌توان دنبال کردن‌های مستقیمی که مدل به آن‌ها اجازه وقوع می‌دهد را شناسایی کرد:

- درصورتی‌که در مدل ایجادشده $\boldsymbol{P_{i,j} = 1}$ یعنی مدل این اجازه را می‌دهد که فعالیت **i** مستقیماً (یعنی بلافاصله) توسط فعالیت **j** دنبال شود.
- درصورتی‌که در مدل ایجاد شده $\boldsymbol{C_{i,j} = 1}$ یعنی مدل این اجازه را می‌دهد که فعالیت **i** مستقیماً توسط فعالیت **j** دنبال شود و بالعکس، یعنی فعالیت **j** مستقیماً توسط فعالیت **i** دنبال شود.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

تابع هدف پیشنهادی

همان‌گونه که ذکر شد تابع هدف پیشنهادی به‌گونه‌ای است که بهینه کردن آن تضمین‌کننده این امر است که در حین رعایت معیارهایی برای دستیابی به حد معقولی از سازگاری، دقت مدل بیشینه گردد؛ بنابراین تابع هدف زیر پیشنهاد می‌گردد:

$$\min \sum_i \sum_j P_{i,j} \times d_{i,j} + \sum_i \sum_j C_{i,j} \times d_{i,j} + \sum_i \sum_j (1 - P_{i,j} - C_{i,j}) \times f_{i,j} \quad (۳)$$

$d_{i,j}$ و $f_{i,j}$ جریمه‌هایی هستند که برای روابط پیش‌نیازی مستقیم و موازی بودن در مدل اعمال می‌شوند. برای آگاهی از نحوه محاسبه آن‌ها ابتدا لازم است با مفهوم اطمینان[1] آشنا شویم.

Conf(a>b): میزان وابستگی فعالیت‌های a و b را نشان می‌دهد و مطابق فرمول زیر محاسبه می‌شود:

$$\text{conf}(a > b) = \frac{|a > b|}{|a|} \quad (۴)$$

هر چه Conf(a>b) به یک نزدیک‌تر باشد یعنی رابطه دنبال کردن مستقیم a توسط b قوی‌تر و هر چه Conf(a>b) به صفر نزدیک‌تر باشد یعنی رابطه دنبال کردن مستقیم a توسط b ضعیف‌تر است.

با توجه به مفاهیم معرفی‌شده $d_{i,j}$ و $f_{i,j}$ به‌صورت زیر محاسبه می‌شوند:

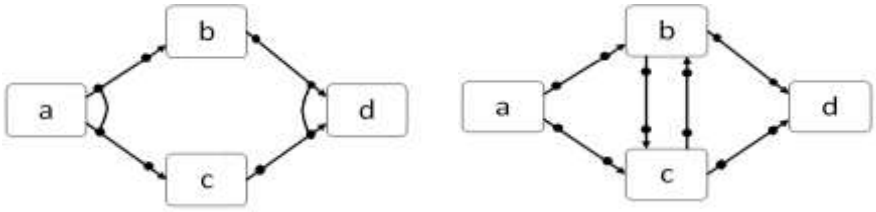
$$d_{i,j} = \begin{cases} M & conf(i > j) = 0 \\ 1 - \text{conf}(i > j) & 0 < \text{conf}(i > j) < th \\ 0 & \text{conf}(i > j) \geq th \end{cases} \quad (۵)$$

$$f_{i,j} = \begin{cases} 0 & \text{conf}(i,j) < th \\ M & \text{conf}(i,j) \geq th \end{cases} \quad (۶)$$

بنابراین با در نظر گرفتن M به‌عنوان یک عدد خیلی بزرگ، می‌توان ادعا نمود که با کمتر شدن دو قسمت اول تابع هدف، دقت مدل بیشتر می‌شود؛ درحالی‌که قسمت سوم تابع هدف تضمین‌کننده این است که درصورتی‌که اطمینان(i,j) بزرگ‌تر از آستانه th باشد حتماً یکی از دو رابطه توالی مستقیم و یا موازی بودن میان آن‌ها برقرار است.

البته باید در نظر داشت که تابع هدف مذکور دارای مقدار یکسانی برای دو مدل موجود در شکل ۲ است؛ بنابراین تابع هدف باید به نحوی اصلاح گردد که میزان متفاوتی برای موازی بودن دو فعالیت و یا وجود تکرار میان دو فعالیت اختیار کند. (در صورتی می‌توانیم بگوییم رابطه میان a وb تکرار است که پس از a بلافاصله b و سپس دوباره a اتفاق بیافتد و بالعکس)

1. Confidence



(ب) **(الف)**

شکل ۲: یکسان بودن میزان تابع هدف فرمول ۳ برای مدل‌های الف و ب

تابع هدف اصلاح‌شده به‌صورت زیر است:

$$\min \sum_i \sum_j P_{i,j} \times d_{i,j} + \sum_i \sum_j C_{i,j} \times d_{i,j} + \sum_i \sum_j (1 - P_{i,j} - C_{i,j}) \times f_{i,j} + \sum_i \sum_j L_{i,j} \times dL_{i,j} + \sum_i \sum_j (1 - L_{i,j}) \times fL_{i,j} \quad (۷)$$

به‌گونه‌ای که $L_{i,j}$ متغیری باینری است که درصورتی‌که مقدار یک را اختیار کند نشان‌دهنده وجود تکرار میان دو فعالیت i و j است و بالعکس اگر مقدار صفر را اختیار کند نشان‌دهنده عدم وجود تکرار میان دو فعالیت i و j است و همچنین:

$$dL_{i,j} = \begin{cases} M & \text{Conf}(i \gg j) = 0 \\ 1 - \text{Conf}(i \gg j) & 0 < \text{Conf}(i \gg, j) < th \\ 0 & \text{Conf}(i \gg j) \geq th \end{cases} \quad (۸)$$

$$fL_{i,j} = \begin{cases} 0 & \text{Conf}(i \gg j) < thL \\ M & \text{Conf}(i \gg j) \geq thL \end{cases} \quad (۹)$$

در اینجا نیز پشتیبان(a>>b) نشان‌دهنده تعداد دفعاتی در لاگ رویدادها است که تکرار میان a و b اتفاق افتاده است. (یعنی پس از a بلافاصله b و سپس دوباره a اتفاق افتاده است) و Conf(a>>b) به‌صورت زیر محاسبه می‌شود:

$$\mathrm{Conf}(a \gg b) = \frac{|a >> b| + |b >> a|}{|a| + |b|} \quad (۱۰)$$

به طریق مشابه می‌توان ادعا کرد که با کمتر شدن قسمت چهارم تابع هدف، دقت مدل بیشتر می‌شود؛ درحالی‌که قسمت پنجم تابع هدف تضمین‌کننده این است که درصورتی‌که Conf(i>>j) بزرگ‌تر از آستانه thL باشد، حتماً رابطه تکرار میان آن‌ها برقرار است.

نتایج تجربی

در این قسمت به ارزیابی و اعتبار سنجی تابع هدف ارائه شده می‌پردازیم. به این منظور ابتدا توسط داده‌های ساختگی تابع هدف پیشنهادی ارزیابی شده و پس از تحلیل حساسیت پارامترهای مدل، نتایج اعمال آن در یک مورد مطالعاتی سازمانی ارائه می‌شود. قابل‌ذکر است جهت انجام تحلیل‌های مربوطه، تهیه خروجی سایر روش‌های کشف فرآیند ابتکاری توسط نرم‌افزار ProM 6.4 انجام شده است؛ بدین ترتیب که جهت تهیه خروجی الگوریتم کاوشگر ابتکاری از افزونه "Mine for a Heuristics Net using Heuristics Miner"، کاوشگر ابتکاری منعطف از افزونه "Mine for a Causal Net using Heuristics Miner" و خروجی Fodina از افزونه " Mine causal net with fodina " استفاده شده است.

نتایج اعمال روش بر روی داده‌های ساختگی

به‌منظور ایجاد رویدادهای ثبتی ساختگی از مدل فرآیندهای ساختگی ارائه شده در پژوهش مدرویس در سال ۲۰۰۶ که تاکنون توسط بسیاری از مقالات در این حوزه جهت ارزیابی نتایج به کار برده شده و نرم‌افزار ProM 6.4 استفاده گردیده است. برای هر مدل‌های انتخابی توسط افزونه[1] "Perform a simple simulation of a (stochastic) petri net" یک رویدادهای ثبتی ساختگی ایجاد شده و سپس توسط افزونه "Add noise to an event log" به میزان ۱۰ درصد نویز به هر یک از رویدادهای ثبتی مذکور اضافه شده است. فهرست رویدادهای ثبتی ساخته‌شده به همراه اطلاعات مرتبط در جدول ۱ قابل مشاهده است.

1. Plugin

جدول ۱: فهرست رویدادهای ثبتی ساخته‌شده به همراه اطلاعات مرتبط

نام رویداد ثبتی ساخته‌شده (مجموعه داده)	نام مدل فرآیند استفاده شده	تعداد فعالیت‌ها	تعداد نمونه‌ها	درصد نویز
A12n0.1	A12	۱۲	۵۰۰۰	۱۰
herbstFig6p18n0.1	herbstFig6p18	۵	۵۰۰۰	۱۰
Al1.n0.1	Al1	۷	۵۰۰۰	۱۰
|2|Skipn0.1	|2|Skip	۶	۵۰۰۰	۱۰
bn3n.01	bn3	۴۰	۵۰۰۰	۱۰

جهت ایجاد چند مدل فرآیند مختلف برای هر رویداد ثبتی فوق‌الذکر از روش Fodina استفاده شد. به این صورت که از ۵ تنظیمات مختلف پارامترهای روش مذکور استفاده شده است که منجر به حصول ۵ مدل فرآیند مختلف به ازای هر رویداد ثبتی شده است. فهرست تنظیمات استفاده شده Fodina در جدول ۲ قابل مشاهده است.

جدول ۲: فهرست تنظیمات استفاده شده Fodina به ازای هر مدل

تنظیمات / پارامترهای Fodina	تنظیمات مربوط به مدل اول	تنظیمات مربوط به مدل دوم	تنظیمات مربوط به مدل سوم	تنظیمات مربوط به مدل چهارم	تنظیمات مربوط به مدل پنجم
Dependency threshold	۸۰	۸۵	۹۰	۹۵	۱۰۰
Length –one-loops threshold	۹۰	۹۰	۹۰	۹۰	۹۰
Length –two-loops threshold	۹۰	۹۰	۹۰	۹۰	۹۰
Long distance dependency threshold	۹۰	۹۰	۹۰	۹۰	۹۰
Precision/generalization threshold	۰	۰	۰	۰	۰
Dependency divisor	۱	۱	۱	۱	۱
Duplicate task threshold	۱۰	۱۰	۱۰	۱۰	۱۰

مدل‌های فرآیند به‌دست‌آمده توسط افزونه " Convert causal net to reduced petri net" به شبکه پتری تبدیل شده و سپس به‌منظور بررسی عملکرد تابع هدف ارائه شده در

سنجش سازگاری و دقت مدل‌های کشف‌شده، برای هر مدل به‌دست‌آمده، با استفاده از رویکرد ارائه شده در پژوهش وان در آلست و همکاران در سال ۲۰۱۲ و به‌واسطه افزونه " Measure Precision/Generalization" دقت و با استفاده از روش ارائه شده در پژوهش آدریانسیاه و همکاران[1] در سال ۲۰۱۱ و به واسطه افزونه " Replay a log on petri net for conformance analysis" سازگاری بازپخش مدل محاسبه گردیده است. نتایج به‌دست‌آمده در برابر مقدار تابع هدف پیشنهادی برای هر مدل مقایسه شده‌اند. مقادیر تابع هدف پیشنهادی نیز تحت ۴ تنظیمات مختلف پارامترها به‌دست‌آمده است. جدول ۳ تنظیمات استفاده شده را نشان می‌دهد.

جدول ۳: مقادیر تنظیمات مختلف پارامترها جهت محاسبه تابع هدف

تنظیمات / پارامترها	تنظیمات اول	تنظیمات دوم	تنظیمات سوم	تنظیمات چهارم
Th	۰,۶	۰,۷	۰,۸	۰,۹
ThL	۰,۱۵	۰,۱۷۵	۰,۲	۰,۲۲۵

خلاصه نتایج به‌دست‌آمده در جدول ۴ قابل مشاهده است. به ازای هر مجموعه داده و هر تنظیمات تابع هدف بهترین مقدار تابع هدف برای هرکدام از مدل‌های ساخته‌شده به‌صورت پررنگ نشان داده شده است. همان‌گونه که در بخش قبل ذکر شد M یک عدد به‌اندازه کافی بزرگ است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

1. Adriansyah et al.

جدول ٤: خلاصه نتایج به‌دست‌آمده برای ارزیابی تابع هدف پیشنهادی

مجموعه داده استفاده شده	مدل‌های کشف‌شده توسط Fodina / معیارها	مدل اول	مدل دوم	مدل سوم	مدل چهارم	مدل پنجم
A12n0.1	سازگاری بازپخش	۰,۹۹۰۶	۰,۹۸۹۶	۰,۹۸۸۶	۰,۹۸۷۹	۰,۹۸۷۹
	دقت	۰,۴۵۷۶	۰,۸۴۵۱	۰,۸۵۷۶	۱,۰۰۰۰	۱,۰۰۰۰
	تابع هدف به‌دست‌آمده از تنظیمات اول[1]	۱۷,۴۹۰۱	۱۰,۵۲۸۴	۸,۵۴۶۴	۵,۴۳۷۱	۵,۴۳۷۱
	تابع هدف به‌دست‌آمده از تنظیمات دوم[2]	۱۷,۴۹۰۱	۱۰,۵۲۸۴	۸,۵۴۶۴	۵,۴۳۷۱	۵,۴۳۷۱
	تابع هدف به‌دست‌آمده از تنظیمات سوم[3]	۱۸,۰۲۶۷	۱۱,۰۶۵۰	۹,۰۸۳۱	M	M
	تابع هدف به‌دست‌آمده از تنظیمات چهارم[4]	۱۹,۶۸۴۵	۱۱,۹۷۹۳	M	۲M	۲M
herbstFig6p18n0.1	سازگاری بازپخش	۰,۹۱۴۰	۰,۹۱۳۷	۰,۹۱۳۷	۰,۹۱۲۷	۰,۸۹۵۸
	دقت	۰,۴۰۴۰	۰,۴۲۸۹	۰,۴۲۸۹	۰,۴۵۳۶	۰,۴۷۴۷
	تابع هدف به‌دست‌آمده از تنظیمات اول	۲۸,۸۴۵۲	۲۶,۱۰۳۱	۲۶,۱۰۳۱	۲۳,۷۹۷۰	۲۲,۷۸۷۶
	تابع هدف به‌دست‌آمده از تنظیمات دوم	۲۹,۳۱۸۶	۲۶,۹۷۸۸	۲۶,۹۷۸۸	۲۴,۴۶۶۶	M
	تابع هدف به‌دست‌آمده از تنظیمات سوم	۲۹,۹۸۷۱	۲۷,۵۶۹۶	۲۷,۵۶۹۶	۲۴,۸۸۷۲	M
	تابع هدف به‌دست‌آمده از تنظیمات چهارم	۳۰,۵۷۴۶	۲۸,۵۵۷۵	۲۸,۵۵۷۵	M	۲M
Al1.n0.1	سازگاری بازپخش	۰,۹۹۵۳	۰,۹۹۴۰	۰,۹۹۳۸	۰,۹۶۵۷	۰,۹۳۴۲
	دقت	۰,۳۸۵۴	۰,۴۰۹۵	۰,۵۴۸۹	۰,۵۵۴۷	۰,۵۶۷۳
	تابع هدف به‌دست‌آمده از تنظیمات اول	۱۶,۴۶۳۳	۱۵,۴۷۶۷	۱۰,۹۹۴۵	۷,۲۳۴۳	۴,۵۳۲۷

1. Th=۰,٦, ThL=۰,۱۵
2. Th=۰,۷, ThL=۰,۱۷۵
3. Th=۰,۸, ThL=۰,۲
4. Th=۰,۹, ThL=۰,۲۲۵

مجموعه داده استفاده شده	مدل‌های کشف‌شده توسط Fodina / معیارها	مدل اول	مدل دوم	مدل سوم	مدل چهارم	مدل پنجم
	تابع هدف به‌دست‌آمده از تنظیمات دوم	۱۶,۴۶۳۳	۱۵,۴۷۶۷	۱۰,۹۹۴۵	۷,۲۳۴۳	۴,۵۳۲۷
	تابع هدف به‌دست‌آمده از تنظیمات سوم	۱۷,۶۷۵۸	۱۶,۴۳۴۵	۱۱,۸۶۶۵	M	۲M
	تابع هدف به‌دست‌آمده از تنظیمات چهارم	۱۹,۴۵۶۷	۱۷,۷۹۴۵	۱۳,۵۶۴۲	M	۲M
|2|Skipn0.1	سازگاری بازپخش	۰,۹۵۷۸	۰,۹۵۱۲	۰,۹۳۹۰	۰,۹۲۰۱	۰,۹۰۳۸
	دقت	۰,۵۷۴۶	۰,۵۹۰۶	۰,۶۷۶۴	۰,۸۳۸۵	۰,۸۹۹۲
	تابع هدف به‌دست‌آمده از تنظیمات اول	۱۰,۸۷۳۲	۷,۰۵۷۵۳	۵,۹۷۳۲	M	M
	تابع هدف به‌دست‌آمده از تنظیمات دوم	۱۲,۸۵۳۰	۹,۸۷۹۲	۸,۰۳۴۷	M	M
	تابع هدف به‌دست‌آمده از تنظیمات سوم	۱۳,۹۸۷۹	۱۰,۹۳۷۱	M	۲M	۲M
	تابع هدف به‌دست‌آمده از تنظیمات چهارم	۱۵,۰۸۳۷	۱۲,۵۸۰۱	M	۲M	۲M
bn3n.01	سازگاری بازپخش	۰,۸۴۷۶	۰,۸۳۹۸	۰,۸۲۳۴	۰,۸۱۹۸	۰,۸۰۴۷
	دقت	۰,۳۵۴۶	۰,۵۳۶۱	۰,۶۵۶۷	۰,۷۷۶۹,	۰,۸۸۹۸
	تابع هدف به‌دست‌آمده از تنظیمات اول	۳۲,۸۳۱۲	۲۸,۳۵۲۱	۲۴,۸۹۲۰	۱۹,۳۸۲۱	۱۵,۰۲۸۱
	تابع هدف به‌دست‌آمده از تنظیمات دوم	۳۳,۲۱۸۴	۲۸,۸۳۱۱	۲۶,۹۸۲۱	۲۰,۸۷۳۲	۱۵,۷۹۳۲
	تابع هدف به‌دست‌آمده از تنظیمات سوم	۳۶,۹۸۳۲	۳۲,۸۷۳۲	۳۰,۹۷۲۳	۲۳,۸۴۲۱	۱۹,۱۱۳۴
	تابع هدف به‌دست‌آمده از تنظیمات چهارم	۴۰,۹۳۷۲	۳۵,۸۶۲۱	۳۳,۹۷۳۲	**۲۷,۷۶۵۲**	M

با توجه به نتایج به‌دست‌آمده مشاهده می‌شود که معیارهای سازگاری بازپخش و دقت با هم رابطه عکس دارند. به این صورت که هرچه سازگاری بازپخش مدل بالاتر باشد دقت آن پایین‌تر و بالعکس هرچه دقت مدل بالاتر باشد سازگاری بازپخش آن پایین‌تر است. از سوی دیگر هر چه مقدار آستانه‌های Th و ThL پایین‌تر باشد، تابع هدف پیشنهادی مدل با دقت بالاتر (و بنابراین سازگاری بازپخش پایین‌تر) را به‌عنوان بهترین مدل شناسایی می‌کند. با بالا رفتن

آستانه‌های مذکور برخی مدل‌های با میزان سازگاری بازپخش پایین (و بنابراین دقت بالا) حذف می‌گردند و مدل‌های با دقت پایین‌تر اما سازگاری بازپخش بالاتر به‌عنوان بهترین مدل شناسایی می‌شوند؛ بنابراین تابع هدف پیشنهادی در انتخاب مدل با دقت بیشینه که دارای حد معقول و قابل تنظیمی از سازگاری باشد، موفق عمل نموده است.

تحلیل حساسیت و راهنمای شاخص‌های سازگاری و دقت

جهت ارائه راهنمایی در خصوص شاخص‌های سازگاری و دقت، تحلیل حساسیت پارامترهای Th و ThL انجام شد که نتایج آن به‌صورت دیداری در شکل ۱-الف در پیوست الف ارائه شده است. جهت انجام این امر از مجموعه داده‌ها و مدل‌های ایجاد شده در قسمت قبل استفاده شده است. برای هر یک از مجموعه داده‌های ذکرشده دو نمودار ارائه گردیده است که در یک نمودار مقادیر دقت و سازگاری هر مدل نمایش داده شده و در نمودار دیگر مقادیر تابع هدف پیشنهادی هر مدل به ازای مقادیر مختلف پارامترهای Th و ThL قابل مشاهده است. با توجه به این موضوع که کمینه‌سازی تابع هدف پیشنهادی مورد نظر است، مشاهده می‌شود که به ازای داده‌های استفاده شده با بالا رفتن مقادیر آستانه Th و ThL و در نتیجه بی‌نهایت شدن تابع هدف مدل‌های با بالاترین دقت (و پایین‌ترین سازگاری بازپخش)، آن مدل‌ها از گزینه‌های انتخاب حذف شده و مدل‌های با دقت پایین‌تر (و سازگاری بازپخش بالاتر) به‌عنوان مدل با بهترین تابع هدف شناخته می‌شوند.

کاربرد مدل پیشنهادی در مطالعه سازمانی

به‌منظور ارزیابی عملکرد تابع هدف پیشنهادی در مواجهه با داده‌های واقعی، روش پیشنهادی بر روی رویدادهای ثبتی حوزه بازاریابی شرکت بهسازان توسعه آرمان (بتا) اعمال گردید. شرکت مذکور در زمینه تولید مواد پوشش‌های ضد حریق و عایق سرد PIR فعال است. در حال حاضر برای حوزه بازاریابی شرکت، فرآیند از پیش تعریف‌شده‌ای وجود دارد تا کارمندان شرکت در حوزه بازاریابی از آن استفاده کنند. ولی شرکت در این زمینه با چالش‌هایی مواجه شده است از جمله این‌که پیش‌بینی می‌شود که فعالیت‌های انجام شده در بخش بازاریابی به‌صورت دقیق از فرآیند مذکور تبعیت نکنند و در عمل مدیریت تصویر شفافی از فرآیندی که واقعاً در حال اجرا است ندارد. این در حالی است که شناسایی فرآیند در حال اجرا برای مدیریت شرکت دارای

اهمیت فراوانی است زیرا به شرکت این امکان را می‌دهد تا موارد انحراف از فرآیند از پیش تعیین‌شده را شناسایی کند و تحلیل دلایل انحراف از فرآیند کمک فراوانی به بهبود فرآیند مورد نظر خواهد نمود. انتظار می‌رود که انحرافات به دلایلی از قبیل سهل‌انگاری کارمندان و یا کاربردی و واقع‌بینانه نبودن مدل از پیش تعیین‌شده اتفاق افتاده باشد. شناسایی هرکدام از این موارد می‌تواند منجر به بهبود شناخت مدیریت از کارمندان شرکت و افزایش دانش نسبت به محیط و عملکرد مشتریان گردد. با توجه به این‌که انجام این امر بدون داشتن تصویری از فرآیند واقعاً در حال اجرا در شرکت غیرممکن است، لذا تصمیم گرفته شد با استفاده از رویدادهای ثبتی موجود فرآیند واقعی کشف گردد.

رویدادهای ثبتی مذکور شامل رویدادهای ثبت شده شرکت در حوزه بازاریابی از شناسایی مشتریان بالقوه تا حصول نتیجه اعم از انعقاد قرارداد و یا عدم تمایل مشتری به همکاری، است. رویداد ثبتی مورد اشاره شامل ۲۶۴۷ مورد، ۲۲۷۵۸ رویداد و ۲۱ فعالیت می‌باشند. به‌منظور حفظ امنیت داده‌های شرکت مورد نظر در این مقاله از نام اصلی فعالیت‌ها استفاده نشده است و فعالیت‌ها تغییر نام داده شده‌اند.

در مرحله اول پاکسازی رویدادهای ثبتی انجام گردید. جهت انجام این امر برخی از دنباله[1] رویدادهای مربوط به هر مورد[2] که رویدادهای مرتبط با آن‌ها ناقص به نظر می‌رسید، حذف گردیدند. جهت انجام این امر از نظر کارشناسان شرکت استفاده گردید. به این نحو که دنباله‌هایی که با رویدادهایی آغاز یا پایان یافته بودند که در عمل آغاز یا پایان یک دنباله توسط آن‌ها غیرممکن بود به‌عنوان دنباله‌های ناقص در نظر گرفته شدند و حذف گردیدند. در نتیجه این امر تعداد ۱۰۶ دنباله از رویدادهای ثبتی حذف گردیدند. سپس تعداد تکرار هرکدام از فعالیت‌ها محاسبه شد و ۴ عنوان فعالیتی که در کمتر از یک درصد موارد حضور داشتند به‌عنوان نویز در نظر گرفته شده و حذف گردیدند. در انتها نیز جهت یکسان‌سازی فعالیت‌های آغازین و پایانی دنباله‌ها به کلیه دنباله رویدادهای مربوط به هر مورد، فعالیت‌های شروع و پایان ساختگی اضافه گردید.

در مرحله بعدی جهت مقایسه با نتایج سایر الگوریتم‌های کاوش ابتکاری، داده‌های شرکت با استفاده از الگوریتم‌های کاوشگر ابتکاری، کاوشگر ابتکاری منعطف و Fodina تحلیل گردید. تحلیل توسط پارامترهای مختلف هر الگوریتم انجام گردید و بهترین نتایج

1. Trace
2. Case

حاصل‌شده از هر روش بر اساس نظر کارشناس خبره مربوطه در شکل‌های ۲-الف تا ۴-الف در پیوست الف قابل مشاهده است. نتایج به‌دست‌آمده با توجه به دانش حوزه توسط کارشناسان مربوطه راضی‌کننده نبودند. دلایل عدم رضایت ایشان به شرح زیر است:

- فراوان بودن تعداد کمان‌ها در گراف به‌دست‌آمده که باعث ناخوانا شدن و بنابراین عدم کارایی لازم مدل شده است.
- عدم امکان دنبال شدن یک فعالیت توسط خودش به‌جز فعالیت‌های **t25** و **t6**
- عدم امکان وجود بیش از ۳ فعالیت ورودی و خروجی برای هر فعالیت

دلیل این امر آن است که با وجود این‌که روش‌های به کار شده از بهترین روش‌های مدل‌سازی فرآیندها در مواجهه با داده‌های دارای نویز می‌باشند، به نظر می‌رسد که در خصوص رویدادهای ثبتی استفاده شده روش‌های مذکور به‌صورت کامل موفق به فائق آمدن بر نویز رویدادهای ثبتی نشده‌اند و دقت مدل‌های به‌دست‌آمده قابل‌قبول نیست.

در مرحله بعدی برای تابع هدف معرفی‌شده محدودیت‌هایی تعریف گردید. دسته اول محدودیت‌ها مربوط به الزامات ساختاری گراف وابستگی هستند که مطابق زیر می‌باشند.

- فعالیت آغازین نباید هیچ کمان ورودی و فعالیت پایانی نباید هیچ کمان خروجی داشته باشد.
- به‌غیراز فعالیت آغازین هر فعالیت بایستی حداقل یک کمان ورودی داشته باشد.
- به‌غیراز فعالیت پایانی هر فعالیت بایستی حداقل یک کمان خروجی داشته باشد.

پس از مشورت با کارشناسان خبره حوزه مربوطه نتیجه گرفته شد که تنها محدودیت‌های فوق کافی نیستند و با توجه به دانش حوزه، محدودیت‌های زیر نیز که توسط کارشناسان مربوطه مطرح شدند به مدل برنامه‌ریزی ریاضی اضافه گردیدند.

- تعداد کل کمان‌های گراف نمی‌تواند بیشتر از ۲۵ کمان باشد.
- هیچ فعالیتی نباید بیشتر از سه کمان ورودی داشته باشد.
- هیچ فعالیتی نباید بیشتر از سه کمان خروجی داشته باشد.
- به‌جز برای فعالیت‌های **t25** و **t6** نباید از هیچ فعالیتی کمانی به سمت خود آن فعالیت وجود داشته باشد.

به‌منظور اولویت دادن به عامل دقت در تابع هدف، مقدار آستانه Th را برابر با ۰/۹ و مقدار آستانه ThL را برابر با ۰/۲۲۵ (که آستانه‌های بالایی محسوب می‌شوند) قرار داده شد. نتیجه

به‌دست‌آمده از اعمال مدل فوق‌الذکر در شکل ۵-الف قابل مشاهده است. نتیجه به‌دست‌آمده از نظر کارشناس خبره مربوطه راضی‌کننده‌تر از نتایج الگوریتم‌های قبلی است؛ زیرا خواناتر و قابل‌استفاده است و مطابق با دانش حوزه است.

در مرحله بعد با استفاده از روش به کار برده شده، توسط کاوشگر ابتکاری منعطف سمانتیک‌های اتصال و انفصال[1] کاویده شدند و شبکه علی مربوط به گراف وابستگی به‌دست‌آمده استخراج گردید. سپس سازگاری بازپخش و دقت نتیجه شبکه علی حاصل‌شده در برابر نتایج حاصل از هر یک از الگوریتم‌های محک فوق‌الذکر مورد مقایسه قرار گرفت. سازگاری بازپخش و دقت مطابق بخش ۴,۱ محاسبه گردیده و در جدول ۵ ارائه شده است. مشاهده می‌شود که دقت به‌دست‌آمده از روش پیشنهادی به طرز محسوسی از دقت نتایج سایر الگوریتم‌ها بالاتر است حال‌آنکه سازگاری بازپخش آن به میزان کمی پایین‌تر از سازگاری بازپخش الگوریتم‌های کاوشگر ابتکاری و Fodina است که قابل‌قبول محسوب می‌گردد. در این قسمت به‌منظور استفاده از تنها یک معیار برای مقایسه سازگاری و دقت نتایج از F_score که میانگین هارمونیک سازگاری و دقت است استفاده شده است. مشاهده می‌شود که نتیجه به‌دست‌آمده از روش پیشنهادی دارای F_score بالاتری نسبت به سایر روش‌ها است و بنابراین علاوه بر نظر کارشناس مربوطه از نظر معیارهای سنجش کیفیت مدل نیز بر نتایج حاصل از سایر روش‌ها برتری دارد.

جدول ۵. معیارهای سنجش کیفیت هرکدام از روش‌های اعمال‌شده

معیار کیفیت / روش	سازگاری بازپخش	دقت	F-Score
الگوریتم کاوشگر ابتکاری	۰,۸۳۴۲	۰,۷۹۳۹	۰,۸۱۳۶
الگوریتم کاوشگر ابتکاری منعطف	۰,۷۹۲۴	۰,۷۵۳۶	۰,۷۷۲۵
الگوریتم Fodina	۰,۸۵۰۹	۰,۸۱۹۹	۰,۸۳۵۱
روش پیشنهادی	۰,۸۲۶۱	۰,۹۳۲۰	۰,۸۷۵۸

نتیجه‌گیری و تحقیقات آتی

با توجه به این موضوع که خروجی‌های روش‌های کشف فرآیندها نهایتاً با برخی معیارهای سنجش کیفیت مدل ارزیابی می‌شوند و نظر به اهمیت فراوان دو شاخص سازگاری بازپخش و دقت در این حوزه، در این مقاله با رویکرد بهینه‌سازی و برنامه‌ریزی ریاضی، مسئله کشف

1. Split and Join Semantics

فرآیندها به یک مدل برنامه‌ریزی ریاضی با تابع هدف در راستای بهبود سازگاری بازپخش و دقت مدل تبدیل گردید. به این منظور، با در نظر گرفتن اهمیت و پرطرفدار بودن روش‌های کشف فرآیند ابتکاری و وابسته بودن کیفیت جواب ارائه شده توسط آن‌ها به نحوه استخراج گراف وابستگی در این مقاله برای اولین بار پس از ارائه رویکردی برای تبدیل مسئله کشف گراف وابستگی به یک مسئله برنامه‌ریزی باینری، به معرفی تابع هدفی پرداختیم که به‌صورت هم‌زمان بهینه کردن سازگاری و دقت مدل را در نظر می‌گیرد. میزان بها دادن به هرکدام از این شاخص‌ها از طریق آستانه‌ای که کاربر نهایی مشخص می‌کند، تعیین می‌شود. رویکرد ارائه شده همچنین این قابلیت را دارا است که با معرفی محدودیت‌های مناسب، دانش حوزه کارشناسان خبره مربوطه را در فرآیند استخراج گراف وابستگی دخیل نمایند و همچنین مدل خروجی را به سمت ارتقاء احتمال سالم بودن آن هدایت نماید. در انتها روش پیشنهادی بر روی رویدادهای ثبتی حوزه بازاریابی یک شرکت واقعی فعال در زمینه تولید مواد پوشش‌های ضد حریق و عایق سرد PIR اعمال گردید و مشخص گردید روش پیشنهادی در دو زمینه قابلیت به‌کارگیری دانش حوزه و معیارهای سنجش کیفیت مدل بر سایر روش‌های مقایسه شده برتری دارد.

در توسعه آتی این پژوهش می‌توان محدودیت‌هایی را در راستای سالم بودن گراف وابستگی و مدل فرآیند خروجی ارائه نمود. همچنین می‌توان تابع هدف و محدودیت‌ها را به‌گونه‌ای تکمیل و تعریف نمود که به‌جای استخراج گراف وابستگی، استخراج مدل فرآیند نهایی از طریق آن ممکن باشد. نهایتاً می‌توان تابع هدف را به‌گونه‌ای اصلاح نمود که معیارهای دیگری از سنجش کیفیت مدل مانند سادگی[1] و عمومیت[2] را نیز در بر بگیرد.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

1. Simplicity
2. Generality

منابع

Adriansyah, A., van Dongen, B. F., & van der Aalst, W. M. P. (2011, August). *Conformance* Checking *Using Cost-Based Fitness Analysis*. In Proceedings of the 2011 IEEE 15th International Enterprise Distributed Object Computing Conference. (pp. 55-64). IEEE Computer Society.

Alves de Medeiros, A. (2006). *Genetic process mining.* Ph.D. thesis, TU Eindhoven.

Alves de Medeiros, A., van Dongen, B. F., van der Aalst, W. M. P., & Weijters, A. J. M. M. (2005). Process Mining for Ubiquitous Mobile Systems: An Overview and a Concrete Algorithm. In: L. Baresi, S . Dustdar, H. Gall, & M. Matera (Eds.) *Ubiquitous Mobile Information and Collaboration Systems* (Springer, Berlin Heidelberg 2005).

Amelia Effendi, Y. & Sarno, R. (2020). Time-based α+ miner for modelling business processes using temporal pattern. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 18(1), 114-123.

Buijs, J. C. A. M., van Dongen, B. F. & ,.van der Aalst, W. M. P. (2012). On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery. In: R. Meersman, H. Panetto, T. Dillon, S. Rinderle-Ma, P. Dadam, X. Zhou, S. Pearson, A. Ferscha, S. Bergamaschi, & I. Cruz (Eds.) *On the Move to Meaningful Internet Systems: OTM 2012* (Springer, Berlin Heidelberg 2012).

Buijs, J. C. A. M., van Dongen, B. F., & van der Aalst, W. M. P. (2014). Quality Dimensions in Process Discovery: The Importance of Fitness, Precision ,Generalization and Simplicity. *International Journal of Cooperative Information Systems, 23*(01), 1-39.

Burattin, A. (2015). Heuristics Miner for Time Interval. In: A. Burattin (Ed.) *Process Mining Techniques in Business Environments: Theoretical Aspects, Algorithms, Techniques and Open Challenges in Process Mining* (Springer International Publishing, Cham 2015).

Burattin, A., Sperduti, A., & van der Aalst, W. M. P. (2012). Heuristics Miners for Streaming Event Data. *Computing Research Repository* abs/1212.6383.

He, Z., Du, Y., Wang, L., Qi, L., & Sun, H. (2018). An Alpha-FL Algorithm for Discovering Free Loop Structures From Incomplete Event Logs. *IEEE Access, 6*, 27885-27901.

Li, J., Liu, D., & Yang, B. (2007). Process Mining: Extending α-Algorithm to Mine Duplicate Tasks in Process Logs. In: K.-C. Chang, W. Wang, L. Chen, C. Ellis, C–.H. Hsu, A. Tsoi, & H. Wang (Eds.)

Advances in Web and Network Technologies, and Information Management (pringer, Berlin Heidelberg 2007).

Li, W., Fan, Y., Liu, W., Xin, M., Wang, H. & Jin, Q. (2019) A Self-Adaptive Process Mining Algorithm Based on Information Entropy to Deal With Uncertain Data. *IEEE Access*, 7, 131681-131691.

Măruşter, L., Weijters, A. J. M. M., van der Aalst, W. M. P., & Van Den Bosch, A. (2006). A Rule-Based Approach for Process Discovery: Dealing with Noise and Imbalance in Process Logs. *Data Mining and Knowledge Discovery, 13*(1), 67-87.

Prodel, M., Augusto, V., Jouaneton, B., Lamarsalle, L. & Xie, X. (2018). Optimal Process Mining for Large and Complex Event Logs. *IEEE Transactions on Automation Science and Engineering, 15(3)*, 1309-1325.

Prodel, M., Augusto, V., Xie, X.,Jouaneton, B., & Lamarsalle, L. (2015). Discovery of patient pathways from a national hospital database using process mining and integer linear programming. *in 2015 IEEE International Conference on Automation Science and Engineering (CASE)*. Gothenburg, 1409-1414

Roci, A., & Davidrajuh, R. (2018). A Polynomial-Time Alpha-Algorithm for Process Mining. *International Journal of Simulation-Systems, Science & Technology, 19(5),12.1-12.7* .

Sarno, R. & Sungkono, K. (2019). A survey of graph-based algorithms for discovering business processes. I*nternational Journal of Advances in Intelligent Informatics,* 5(2), 137-149.

Sun, H., Du, Y., Qi, L., & He, Z. (2019). A Method for Mining Process Models With Indirect Dependencies via Petri Nets. *IEEE Access, 7*, 81211-81226.

van der Aalst, W. M. P. (2014). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Berlin Heidelberg: Springer.

van der Aalst, W. M. P., Adriansyah, A., & van Dongen, B. F. (2012). Replaying history on process models for conformance checking and performance analysis. *WIREs Data Mining and Knowledge Discovery*, *2*(2), 182-192.

van der Aalst, W. M. P., Weijters, A. J. M. M., & Maruster, L. (2004). Worklfow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, *16*(9), 1128-1142 .

van der Werf, J. M. E. M., van Dongen, B. F., Hurkens, C. A. J., & Serebrenik, A. (2008). *Process Discovery Using Integer Linear Programming*. In: van Hee K.M., Valk R. (eds) Applications and Theory of Petri Nets (Springer, Berlin Heidelberg 2008).

van Zelst, S. J., van Dongen, B. F., & van der Aalst, W. M. P. (2015). *ILP-Based Process Discovery Using Hybrid Regions.* In Proceedings of the ATAED 2015 workshop. (pp. 47-61). CEUR-WS.org.

van Zelst, S. J ,.van Dongen, B. F., van der Aalst, W. M. P., & Verbeek, H. M. W. (2018). Discovering workflow nets using integer linear programming. *Computing*, *100*(5), 529-556.

vanden Broucke, S. K. L. M., & De Weerdt, J. (2017). Fodina: A robust and flexible heuristic process discovery technique. *Decision Support Systems*, *100*, 109-118.

Vidgof, M., Djurica, D., Bala, S. & Mendling, J. (2020). Cherry-Picking from Spaghetti: Multi-range Filtering of Event Logs. In: S. Nurcan, I. Reinhartz-Berger, P. Soffer, J. Zdravkovic (Ed.) *Enterprise, Business-Process and Information Systems Modeling* (Springer International Publishing, Cham 2020).

Weijters, A. J. M. M., van der Aalst, W. M. P., & alves de Medeiros, A. (2006). Process Mining with the Heuristics Miner-algorithm. *BETA Working Paper Series WP, 166*, Eindhoven University of Technology.

Weijters, A. J. M. M., & Ribeiro, J. T. S. (2011, April). Flexible Heuristics Miner (FHM). *In 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)* (pp. 310-317). IEEE.

Wen, L., van der Aalst, W. M. P., Wang, J., & Sun ,J. (2007). Mining process models with non-free-choice constructs. *Data Mining and Knowledge Discovery, 15*(2), 145-180.

Wen, L., Wang, J., van der Aalst, W. M. P., Huang, B., & Sun, J. (2009). A novel approach for process mining based on event types. *Journal of Intelligent Information Systems, 32*(2), 163-190.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی