شناسایی رانش مفهوم در نگارهٔ رویداد با استفاده از اطلاعات آماری گونه‌ها

فرشته جوادزاده، مهدی یعقوبی* و سهیلا کرباسی

گروه مهندسی کامپیوتر، دانشکده فنی مهندسی گرگان، دانشگاه گلستان، گرگان، ایران

چکیده

در سال‌های اخیر مدیریت فرآیندهای سازمانی (BPM)، به‌دلیل افزایش کارایی سازمان‌ها بسیار مورد توجه قرار گرفته است. استخراج و تحلیل اطلاعات فرآیندهای سازمانی بخش مهمی از این ساختار است؛ اما این فرآیندها در طول زمان پایدار نیستند و به مرور دچار تغییر می‌شوند که به این تغییرات، رانش مفهوم در فرآیند گفته می‌شود. کشف رانش‌های مفهوم یکی از چالش‌های موجود در حوزهٔ مدیریت فرآیندهای سازمانی است. در این مقاله الگوریتمی برای شناسایی رانش‌های مفهوم در نگارهٔ رویداد ارائه شده که براساس تحلیل توزیع گونه‌های دنباله در اجرای فرآیند است. در این روش با حرکت دو پنجره روی نگارهٔ رویداد، دو بردار ویژگی از گونه‌های دنباله‌های دو پنجره حاصل و سپس با استفاده از آزمون‌های آماری گونه‌های دو پنجره با یکدیگر مقایسه و درنهایت رانش‌ها شناسایی می‌شوند. آزمایش‌های صورت‌گرفته روی پایگاه‌های داده مصنوعی، درستی روش و برتری آن را نسبت به روش‌های پیشین نشان می‌دهند.

واژگان کلیدی: رانش مفهوم، نگارهٔ رویداد، فرآیند کاوی، فرآیندهای سازمانی، گونه

Concept drift detection in event logs using statistical information of variants

Fereshteh javadzadeh, Mehdi Yaghoubi* & Soheila Karbasi

Department of Computer Engineering, Faculty of engineering,

Golestan university, Gorgan, Iran

Abstract

In recent years, business process management (BPM) has been highly regarded as an improvement in the efficiency and effectiveness of organizations. Extracting and analyzing information on business processes is an important part of this structure. But these processes are not sustainable over time and may change for a variety of reasons, such as the environment, human resources, capital market changes, seasonal, and climate changes. These changes in business processes are referred to as concept drift in event logs. The discovery of concept drifts is one of the challenges in business process management. These drifts may occur suddenly, gradually, periodically, or incrementally. This paper proposes an algorithm for identifying sudden concept drifts in event logs that are created by BPM. Each execution of the process instance follows a specific path in the process model called a trace, all traces that follow the same path in process model are called a variant. The proposed algorithm is based on the distribution of trace variants in the execution of processes. In this method, by moving two sliding windows on the event log, two feature vectors are derived from the two windows trace variants, these windows are named reference and detection windows. Then variants of the two windows are compared by applying statistical G-test and finally the drifts are identified. In statistics, G-test is likelihood-ratio or maximum likelihood statistical significance test. Experiments on artificial databases show the correctness of the method and its superiority to the previous methods. In the proposed method, the detection accuracy is 0.06% better than state-of-the-art methods on average

Keywords: Concept drift, event log, process mining, business processes, variant

* **Corresponding author** * **نویسندهٔ عهده‌دار مکاتبات**

• **تاریخ ارسال مقاله:** ۱۳۹۸/۵/۴ • **تاریخ پذیرش:** ۱۴۰۰/۳/۲ • **تاریخ انتشار:** ۱۴۰۱/۰۳/۳۱ • **نوع مطالعه:** پژوهشی

سال ۱۴۰۱ شمارهٔ ۱ پیاپی ۵۱

[Downloaded from jsdp.rcisp.ac.ir on 2026-09-12]

[DOI: 10.52547/jsdp.19.1.75]

۱- مقدمه

سامانه‌های اطلاعاتی که قادر به پشتیبانی فرآیندهای سازمانی هستند، روزبه‌روز درحال افزایش هستند. در این سامانه‌ها، از اجرای فرآیندهای سازمانی، اطلاعاتی تحت عنوان نگارهٔ رویداد[1] ذخیره می‌شود که در فرآیند کاوی[2] مورد تحلیل و آنالیز قرار می‌گیرد. *فرآیند کاوی*، مجموعه‌روش‌هایی است که با استفاده از اطلاعات استخراج شده از نگارهٔ رویداد، فرآیندهای سازمانی را کشف و بررسی می‌کند و درنهایت باعث بهبود اجرای فرآیندها در سازمان‌ها می‌شود [1].

نقطهٔ شروع فرآیندکاوی، نگارهٔ رویداد است که شامل مجموعه‌ای از رویدادها است. در این نگاره‌ها، هر رویداد به یک کار[3] (یک گام تعریف‌شده در مدل فرآیند) اشاره می‌کند و به یک *پرونده*[4] (نمونه‌ای از اجرای فرآیند) خاص مرتبط است. رویدادهای یک پرونده با یک برچسب زمانی در نگارهٔ رویداد ثبت و به‌صورت اثری از «اجرای» فرآیند دیده می‌شوند. در نگاره‌های رویداد اطلاعات دیگری نیز درباره رویدادها ذخیره می‌شود که شامل منابع اجراکننده فعالیت (مانند شخص یا دستگاه) و غیره است.

روش‌های فرآیند کاوی معاصر، به‌طورعمومی فرآیندها را پایدار در نظر می‌گیرند؛ درصورتی‌که امروزه در واقعیت فرآیندهای سازمانی در طول زمان و در فواصل مختلف ممکن است دچار تغییر و تحول شوند. این تغییرات به دلایل مختلفی هم‌چون تغییر در قوانین سازمان‌ها، شرایط فصلی، تغییر در عرضه و تقاضا، تعادل بارِکاری، تعدیل نیرو و یا بروز بلایا و فجایع طبیعی صورت می‌گیرد. برای مثال، در بیمارستان‌ها با توجه به پیشرفت و تغییر مواد دارویی، روند درمان بیمار تغییر می‌کند. این تغییرات ممکن است در طول زمان تأثیرات عمیقی بر روی کارایی فرآیندهای سازمانی بگذارند؛ به همین دلیل شناسایی محل وقوع و دلیل این تغییرات، برای مدیران سازمان‌ها از اهمیت ویژه‌ای برخوردار است. به این حالت که فرآیند در حال اجرا دچار تغییر می‌شود، رانش مفهوم[5] در فرآیند می‌گویند [2]. در بیانیه فرآیندکاوی، رانش مفهوم به‌عنوان چهارمین چالش فرآیندکاوی از یازده چالش معرفی شده است [3].

رانش مفهوم به‌طورعمومی می‌تواند به چهار حالت عمدهٔ ناگهانی[6]، تدریجی[7]، دوره‌ای[8] و افزایشی[9] اتفاق بیافتد که در شکل (۱) نمایش داده شده است. در تغییرات ناگهانی، فرآیند جدید جایگزین فرآیند موجود می‌شود و به‌طورمعمول در مواقع اضطراری یا با تغییر قوانین صورت می‌گیرد. در تغییرات تدریجی نیز فرآیند موجود به‌وسیلهٔ فرآیند جدید جایگزین می‌شود اما برخلاف تغییر ناگهانی، در اینجا هر دو فرآیند برای مدت‌زمانی با هم در حال اجرا هستند و سپس فرآیند پیشین به‌تدریج کنار گذاشته می‌شود. تشخیص تغییرات تدریجی نسبت به تغییرات ناگهانی دشوارتر است، زیرا نقطه زمانی مشخصی برای شروع تغییر وجود ندارد. این تغییرات در یک بازه زمانی مشخص می‌شوند و رفتارهای قبلی به‌تدریج به رفتارهای جدید تغییر پیدا می‌کنند. در رانش دوره‌ای، مجموعه‌ای از فرآیندها بعد از مدت‌زمانی، دوباره ظاهر می‌شوند که می‌تواند به‌دلیل تغییرات فصلی در سازمان باشد. همان‌طور که در شکل (۱) نشان داده شده است، در رانش افزایشی، فرآیند M_N با تغییرات کوچک و افزایشی جایگزین فرآیند M_1 شده است. این دسته از رانش‌ها اغلب در سازمان‌هایی مشاهده می‌شوند که در حال ارتقای کیفیت فرآیندهای خود هستند.

روش‌های شناسایی رانش مفهوم را می‌توان در دو دستهٔ بزرگ برون‌خط[10] و برخط[11] طبقه‌بندی کرد [4]. روش‌های برون‌خط برای مواردی مناسب هستند که هدف از نتایج به‌دست‌آمده از تحلیل روی تغییرات، بهبود و طراحی مجدد فرآیندها است و تغییرات ایجادشده در فرآیند به‌صورت بلادرنگ کشف و بررسی نمی‌شوند. برخلاف آن، در روش‌های برخط، تغییرات باید به‌صورت بلادرنگ شناسایی شوند و این امکان را به سازمان می‌دهند که به‌سرعت واکنش مناسب نشان دهد و اقدامات لازم برای تطبیق با شرایط جدید را به عمل آورد.

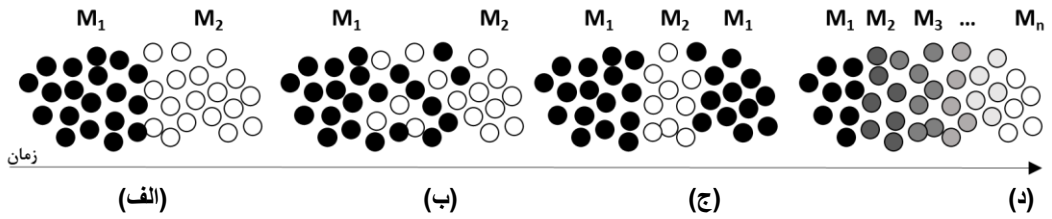
رانش مفهوم در فرآیند درواقع تغییر در مسیرهای مدل فرآیند است. این تغییرها می‌تواند حذف، ایجاد، تکرار و غیره برای یک مسیر یا چند مسیر در گراف مدل فرآیند باشد، درنتیجه در توزیع آماری اطلاعات ذخیره‌شده در اجرای فرآیند تأثیر می‌گذارد. در این مقاله الگوریتمی مبتنی بر توزیع آماری گونه‌ی دنباله‌ها در اجرای فرآیند

[1] Event log
[2] Process Mining
[3] task
[4] case
[5] Concept Drift

[6] sudden
[7] gradual
[8] recurring
[9] incremental
[10] Offline
[11] Online

[Downloaded from jsdp.rcisp.ac.ir on 2026-09-12]
[DOI: 10.52547/jsdp.19.1.75]

برای شناسایی رانش‌های مفهوم ارائه می‌شود. در این روش از دو پنجرهٔ هم‌طول با عنوان‌های *پنجرهٔ مرجع* و *پنجرهٔ اکتشافی* استفاده شده است که نگارهٔ رویداد را به بخش‌های کوچک‌تر تقسیم می‌کند؛ سپس هرکدام از این دو پنجره‌ها به بردارهایی از توزیع گونهٔ دنباله‌های خود تبدیل می‌شوند. در مرحلهٔ بعد با استفاده از آزمون آماری G، توزیع گونه‌های دو پنجره با یکدیگر مقایسه می‌شوند و درنهایت نقاطی که مقدار آزمون G در آن‌ها از یک مقدار آستانه بیشتر باشد، به‌عنوان رانش مفهوم شناسایی می‌شوند.



(شکل-۱): انواع رانش‌ها. (الف) رانش ناگهانی (ب) رانش تدریجی (ج) رانش دوره‌ای (د) رانش افزایشی.

محور افقی: زمان، تغییر تیرگی: مدل فرآیندها، دایره‌ها: نمونه فرآیندها (پرونده) [4]

(Figure-1): Different types of drifts. (a) Sudden drift. (b) Gradual drift. (c) Recurring drift. (d) Incremental drift. x-axis: time, color: process models. circles: process instances [4]

در ادامهٔ این مقاله در بخش ۲، پژوهش‌های پیشین در این زمینه معرفی شده است. در بخش ۳ روش پیشنهادی و ایدهٔ اصلی بیان می‌شود. ارزیابی روش پیشنهادی در بخش ۴ انجام و در بخش ۵ نیز به نتیجه‌گیری پرداخته می‌شود.

۲- پژوهش‌های پیشین

پژوهش‌های زیادی در زمینهٔ رانش مفهوم در داده‌کاوی و یادگیری ماشین صورت گرفته است. برای مثال در مقالهٔ کلینکنبرگ و همکاران [5] با استفاده از کمینه‌سازی خطا در یادگیرنده ماشین بردار پشتیبان (SVM[1]) رانش شناسایی می‌شود؛ اما به‌طور کلی روش‌های توسعه‌یافته در این حوزه با ساختارهای دادهٔ ساده سروکار دارد (اعم از متغیرهای عددی یا دسته‌بندی و بردارها)، درحالی‌که در فرآیندهای سازمانی، تغییرات در ساختارهای پیچیده‌تری صورت می‌گیرد (ساختارهای شامل حلقه، هم‌زمانی، تداخل، لغو و غیره). به همین دلیل نمی‌توان روش‌های تشخیص رانش مفهوم در داده‌کاوی را به‌راحتی به فرآیندکاوی انتقال داد.

در حال حاضر روش‌های متفاوتی در زمینهٔ رانش مفهوم در فرآیندکاوی ارائه شده است که هرکدام از آن‌ها قادر به شناسایی انواع خاصی از رانش‌ها هستند. در روش ارائه‌شده توسط بوز و همکاران [4][6]، روشی برای شناسایی رانش مفهوم با استفاده از استخراج الگوها (یا ویژگی‌ها) از دنباله‌ها ارائه شده است. در این روش از پنجرهٔ غیرهم‌پوشان و پیوسته با طول ثابت استفاده شده است و ویژگی‌های استخراج‌شدهٔ دنباله‌ها با استفاده از آزمون آماری با یکدیگر مقایسه می‌شوند. به‌دلیل استفاده از پنجره با طول ثابت، نتایج به‌دست‌آمده در این مقاله وابستگی زیادی به طول انتخاب‌شده دارد. برای حل این مشکل مارتجوشو و همکاران، ایدهٔ استفاده از پنجرهٔ تطبیقی را مطرح می‌کنند [7]. در پنجرهٔ تطبیقی، اندازهٔ پنجره با توجه به ویژگی‌های جریان داده به‌صورت خودکار تغییر می‌کند. همچنین روشی برای شناسایی تغییرات دوره‌ای ارائه شده است که در آن به جای درنظر گرفتن جمعیت‌هایی با تعداد دنباله‌های ثابت، جمعیت‌ها را در دنباله‌هایی با اندازهٔ متغیر در نظر می‌گیرد تا مقیاس‌های زمانی متفاوت را در برگیرد. برای آن‌که صحت و دقت کار افزایش یابد، باید فضای زیادی از ویژگی‌ها را به‌صورت خودکار جستجو کرد یا از کاربر خواسته شود که ویژگی های مورد نظر خود را تعیین کند، به این معنی که کاربر باید از ساختار و خصوصیات تغییر دانش قبلی داشته باشد. در همهٔ این موارد ممکن است نتوان همهٔ تغییرات را شناسایی کرد؛ زیرا این احتمال وجود دارد که ویژگی‌های انتخاب‌شده همهٔ انواع رانش‌ها را پوشش ندهند. همچنین این روش‌ها با توجه به نیاز به استخراج و تحلیل حجم زیادی از ویژگی‌ها با ابعاد بزرگ، قابلیت مقیاس‌پذیری ندارند.

در مقالهٔ ماراجی و همکاران روشی خودکار و مقیاس‌پذیر برای شناسایی تغییرات ارائه شده است [8]. ایدهٔ اصلی این مقاله، اعمال آزمون فرضیه آماری، روی

[1] Support Vector Machine

[Downloaded from jsdp.rcisp.ac.ir on 2026-09-12]

[DOI: 10.52547/jsdp.19.1.75]

توزیع اجراهای[1] مشاهده شده در دو پنجرهٔ زمانی متوالی است. «اجرا»، مجموعه‌ای از دنباله‌هایی است که به‌واسطه رابطهٔ هم‌زمانی بین فعالیت[2]های خود، با یکدیگر معادل هستند. از آن‌جا که در روش‌های مبتنی بر دنباله، چالش هم‌زمانی در شناسایی رانش وجود دارد، آنها سعی کردند با استفاده از روش مبتنی بر *اجرا*، این مشکل را حل کنند. مسألهٔ شناسایی رانش مفهوم در فرآیند به پیداکردن نقطه‌ای از زمان است که در آن مجموعهٔ اجراهای قبل از این نقطه با مجموعهٔ اجراهای بعد از این نقطه، از نظر آماری اختلاف معناداری داشته باشند [8].

استوار و همکاران در ادامهٔ پژوهش ماراجی [8]، علاوه‌بر اطلاعات آماری «اجرا» در سابقهٔ رویدادها از روابط دوتایی بین رویدادها که در الگوریتم آلفا+ معرفی شده بود (الگوریتم آلفا و آلفا+ دو تا از الگوریتم‌های رایج و پایه‌ای در بحث فرآیندکاوی است که ساختار گراف مدل فرآیند را از نگارهٔ رویداد استخراج می‌کند) برای شناسایی رانش مفهوم استفاده کردند. این روابط دوتایی شامل هم‌روندی[3]، تقدم-تأخر[4]، حلقهٔ به‌خود[5] و حلقهٔ به‌طول دو[6] (دوتایی) می‌باشد. اما متأسفانه در نتایج بهبود چندانی نسبت به روش ماراجی و همکاران حاصل نشد [16].

لی و همکاران یک ویژگی توسعه‌پذیر ارائه کرده‌اند که با استفاده از روش‌های پنجرهٔ زمانی و *کاوش ابتکاری*[7]، رانش مفهوم را در نگارهٔ رویدادهای ناقص، شناسایی و مکان‌یابی می‌کند. این مقاله از الگوریتم فرآیند کاوی ژنتیک (GPM[8]) استفاده کرده و همچنین آن را با استفاده از *کاوش ابتکاری وزنی* (WHM[9]) و دیفرانسیل تکاملی (DE[10]) بهبود بخشیده است. سپس با استفاده از این الگوریتم بهبودیافته یک مدل فرآیند جدید برای فرآیندهای تغییر یافته کشف می‌کند [9].

در روشی دیگر سیلیگر و همکاران از ساخت مدل‌های فرآیند برای شناسایی رانش‌ها استفاده کرده‌اند [10]. ایدهٔ اصلی روش آن‌ها مقایسهٔ ساختاری مدل‌های فرآیند و استفاده از معیارهای گراف برای شناسایی تغییرات فرآیند است. در این روش نگارهٔ رویداد با استفاده از دو پنجره تطبیقی متوالی به دو زیرنگاره کوچک‌تر تقسیم (پنجره مرجع و پنجره اکتشافی) و سپس مدل فرآیند زیرنگاره‌ها استخراج شده و از روی آن‌ها چند معیار گراف (مانند درجه گراف، تعداد یال‌ها، درجه رأس و غیره) محاسبه می‌شود؛ سپس با اعمال آزمون آماری G روی معیارهای گراف دو زیرنگارهٔ متوالی، رانش‌های مفهوم شناسایی می‌شوند. سیلیگر برای استخراج مدل فرآیند از الگوریتم *کاوش ابتکاری* استفاده کرده است؛ زیرا این الگوریتم انتزاع خوبی از نگارهٔ رویداد می‌دهد، سرعت خوبی دارد و هم‌زمانی رویدادها را به‌صورت خودکار تشخیص می‌دهد. از آنجایی‌که *کاوش ابتکاری* قادر نیست حلقه‌های طولانی را به‌درستی تشخیص دهد، در نتیجه این روش نمی‌تواند رانش‌های به‌وجودآمده در فرآیندهای دارای حلقه را به‌درستی شناسایی کند [10].

در دو پژوهش دیگر هم از روش‌های متفاوتی برای شناسایی رانش مفهوم استفاده می‌شود. آنها از خوشه‌بندی پیمایش‌ها و محاسبهٔ فاصلهٔ هر جفت رویداد در هر پیمایش استفاده کردند. درحقیقت ساختار فرآیند در قالب محل قرارگرفتن هر رویداد در پیمایش در روش آنها مورد توجه قرار گرفت. در روش آنها هم مانند روش بوز و همکاران اندازهٔ پنجره باید توسط کاربر تنظیم می‌شد [17,18].

در پژوهشی که توسط کارمونا و همکارن انجام شد، یک روش بلادرنگ با توانایی آموزش را برای شناسایی رانش مفهوم ارائه شد. در روش آنها اندازهٔ پنجره طبق یک الگوریتم یادگیری تخمین زده می‌شد؛ سپس یک پنجرهٔ انطباقی با استفاده از پارامتر به‌دست‌آمده نقاط رانش مفهوم را شناسایی می‌کرد [19].

در مقاله‌ای که خجسته و همکاران انجام دادند رانش مفهوم ناگهانی را در فرآیندهای کسب‌وکار با استفاده از روش یادگیری عمیق شناسایی کردند. روش پیشنهادی آنها مستقل از پنجره طراحی و پیاده‌سازی شده است. در این روش با مدل‌سازی پیمایش‌ها بر دنباله واژگان در زبان طبیعی و خوشه‌بندی پیمایش‌ها، همهٔ انواع روابط میان پیمایش ها را به‌صورت خودکار استخراج شده است [20].

۳- روش پیشنهادی

۳-۱- مفاهیم و تعاریف

هنگامی‌که یک فرآیند ایجاد می‌شود، مجموعه‌ای از کار[11]ها در فرآیند شکل می‌گیرد که روی هر کار مجموعه‌ای از رویدادها رخ می‌دهد. هر رویداد، شروع کار،

[1] Distributions of runs
[2] Activity
[3] Concurrency
[4] Causality
[5] Self-loop
[6] Length-two loop
[7] Heuristic Miner
[8] Genetic Process Mining
[9] Weighted heuristic miner
[10] Differential Evolution
[11] task

[Downloaded from jsdp.rcisp.ac.ir on 2026-09-12]
[DOI: 10.52547/jsdp.19.1.75]

پایان کار، انصراف، تعلیق و دیگر وضعیت‌های یک کار را بیان می‌کند.

نگارهٔ رویداد و دنباله‌ها، نقطه‌ی شروع مدیریت رانش مفهوم هستند. هر نگارهٔ رویداد شامل چند دنباله و هر دنباله شامل چند رویداد است. رویدادهای یک نمونه فرآیند دارای ترتیب هستند و می‌توانند حاوی مشخصاتی مانند برچسب زمانی و منابع مرتبط باشند. در ادامه مفاهیم اصلی استفاده‌شده در این مقاله تعریف شده‌اند:

تعریف ۱: رویداد، نگارهٔ رویداد، دنباله

فرض می‌کنیم L یک نگارهٔ رویداد روی مجموعهٔ برچسب‌های $\mathcal{L}$ است؛ همچنین فرض می‌کنیم E مجموعه‌ای از رخداد رویدادها و تابع $\lambda : E \to \mathcal{L}$ یک تابع برچسب‌گذاری است. دنبالهٔ $\sigma \in E^*$ ، به‌صورت ترتیبی از $i \in [0, n-1]$ و دنباله‌ای از رویدادهای با طول n تعریف می‌شود، به‌طوری‌که $\sigma = \langle e_0, e_1, e_2, \dots, e_n \rangle$. یک دنباله است که اثر اجرای یک نمونه از اجرای فرآیند را در نگارهٔ رویداد، نشان می‌دهد و بیان‌گر ترتیب کلی رخداد رویدادها در اجرای فرآیند است [8].

تعریف ۲: گونه (مسیر)

دو دنبالهٔ $\sigma_A = \langle e_{1_A}, e_{2_A}, e_{3_A}, \dots, e_{n_A} \rangle$ و $\sigma_B = \langle e_{1_B}, e_{2_B}, e_{3_B}, \dots, e_{n_B} \rangle$ را در نظر می‌گیریم. اگر تساوی $\lambda(e_{1_A}), \lambda(e_{2_A}), \dots, \lambda(e_{n_A}) = \lambda(e_{1_B}), \lambda(e_{2_B}), \dots, \lambda(e_{n_B})$ برقرار باشد (یعنی سلسله کارهای این دو دنباله یکسان باشد)، σ_A و σ_B در یک گونهٔ یکسان v قرار می‌گیرند. به عبارت دیگر، دنباله‌هایی که در یک گونه از فرآیند قرار دارند، سلسله‌ای از کارهای یکسان دارند و با دنباله‌های گونه‌های دیگر متفاوت هستند و همچنین مسیر یکسانی را در مدل فرآیند پیموده‌اند. به بیان دیگر، گونه $v \in \mathcal{L}^*$، شامل مجموعه‌ای از دنباله‌های مشابه است که همگی مسیر یکسانی را در مدل فرآیند طی می‌کنند. به‌عنوان مثال شکل (۲) بخشی از یک نگارهٔ رویداد فرآیند «پرداخت خسارت» [1] را نشان می‌دهد. در این شکل، پنج دنباله و دو گونه وجود دارد.

(شکل-۲): بخشی از یک نگارهٔ رویداد با ۵ دنباله و۲ گونه

(Figure-2): A part of an event log with 5 traces and 2 variants.

(جدول-۱): فهرست نمادها

(Table-1): list of notations

نماد	عنوان	توضیحات
σ	دنباله‌ای از رویدادها	از کنار هم قرار گرفتن چند رویداد که ترتیب زمانی آنها در اجرای فرآیند اهمیت دارد. $\sigma \in E^*$
v	دنباله‌ای از برچسب ها	دنباله‌ای از کارها که در مدل فرآیند مشخص شده است و برای هر کار یک پرچسب تعیین شده است به عبارت دیگر هر مسیر در گراف مدل فرآیند است. $v \in \mathcal{L}^*$
E	مجموعه رویدادها	مجموعه تمام رویدادهایی که در اجرای یک فرآیند سازمانی رخ می‌دهد
$\mathcal{L}$	مجموعه برچسب‌ها	مجموعه همه کارهایی (Task) که با یک پرچسب یکتا در مدل فرآیند تعریف شده است.
λ	تابع استخراج برچسب هر رویداد. $\lambda : E \to \mathcal{L}$	

۲-۳- ایدهٔ اصلی

رانش مفهوم در فرآیند بر اثر تغییر ساختار در مدل فرآیند صورت می‌گیرد. این تغییر ساختار می‌تواند بر اثر ایجاد یا حذف بخشی از فرآیند، ایجاد یک انشعاب جدید، اضافه‌شدن یک حلقه به فرآیند، جایگزینی و یا جابه‌جایی بخش‌های فرآیند صورت بگیرد. برای مثال در یک شرکت سفارش و تحویل غذا که هزینه را پس از تحویل غذا به مشتری دریافت می‌کرد، براساس تغییر رویه و قانون جدید شرکت، هزینه را قبل از تحویل غذا دریافت می‌کند. یا براساس سیاست جدید شرکت، قبل از فرآیند تکمیل سفارش، گرفتن مشخصات فردی سفارش‌دهنده که درقبل در فرآیند وجود نداشته، به مراحل سفارش اضافه می‌شود. اثر هر کدام از این تغییرات را می‌توان در نگارهٔ رویداد مشاهده کرد. به این صورت که برخی از این تغییرات،

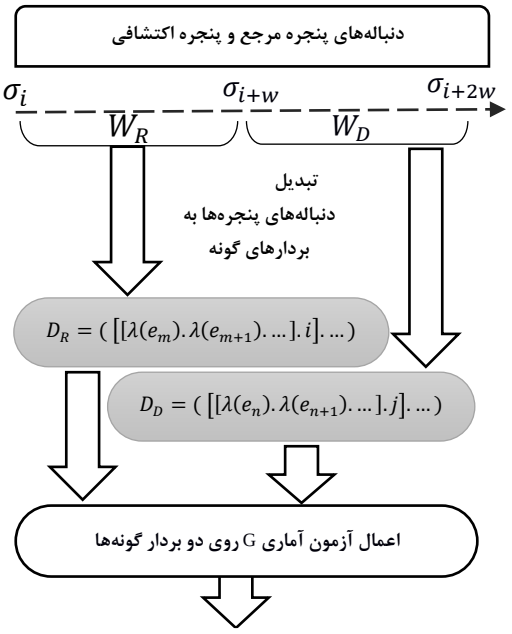
[Downloaded from jsdp.rcisp.ac.ir on 2026-09-12]

[DOI: 10.52547/jsdp.19.1.75]

منجر به ایجاد یک گونهٔ جدید می‌شوند، برخی دیگر، گونه‌ای که وجود داشته را حذف و گونه دیگری را جایگزین و برخی نیز توزیع گونه‌ها را دگرگون می‌کنند. تنها حالتی که باعث تغییر گونه در فرآیند نمی‌شود، تغییر در توزیع انشعاب‌های شرطی است که این تغییر، فراوانی دو گونه (مسیر) را تغییر می‌دهد. در این پژوهش با توجه به آن‌که تغییرات فرآیند بر روی مسیرهایی که فرآیند در طول زمان طی می‌کند، تأثیر گذار است، توزیع گونه‌ها به عنوان معیار و خصوصیت اصلی برای شناسایی رانش مفهوم در نظر گرفته شده است.

۳-۳- شناسایی رانش مفهوم

همان‌طور که گفته شد، در روش ارائه‌شده در این مقاله، برای تشخیص رانش مفهوم در فرآیند، از تغییر توزیع گونه‌ها در نگارهٔ رویداد استفاده می‌شود. درواقع نقطه‌ای را به‌عنوان رانش مفهوم در نظر می‌گیریم که توزیع گونه‌ها، قبل و بعد از این نقطه، تفاوت قابل ملاحظه‌ای با یکدیگر داشته باشند. الگوریتم (۱) روش کلی این پژوهش را بیان می‌کند. همچنین شکل (۳) نمایی کلی از روش ارائه‌شده را نشان می‌دهد که مراحل آن در ادامه با جزییات تشریح می‌شود.



(شکل-۳): چارچوب کلی از روش ارائه‌شده
(Figure-3): Framework for the proposed method

1	Initialize the event log L, the window size w, p-value threshold τ, step size s
2	**For** (i = 0 ; i < |L| - w*2 ; s)
3	W_R= W [i : i + w -1] // پنجره مرجع
4	W_D= W[i + w : i + w * 2-1] // پنجره اکتشافی
5	Generate two vectors D_R and D_D
6	gValue = gTest (D_R, D_D)
9	τ = *Significance_level* (0.05)
10	**For** (i = 0 : i < |gValue|)
11	Set flag = false.
12	**If** gValue [i] > τ and flag = false **then**
13	**If** gValue [i] > τ and flag = false **then**
14	Set flag = true
15	A drift detected.
16	**If** gValue[i] < τ
17	Set flag = false.

(الگوریتم-۱): شبه کد روش پیشنهادی
(Algorithm-1): Pseudocode of the proposed method

۱-۳-۳- دنباله‌های مرجع و اکتشافی

در مرحله نخست، نگارهٔ رویداد را به بخش‌های کوچک‌تر یا زیرنگاره تقسیم می‌کنیم (خطوط ۲ الی ۴ الگوریتم ۱). برای این‌کار از دو پنجرهٔ متوالی استفاده می‌شود که طول هرکدام از پنجره‌ها ثابت و برابر با w است. پنجرهٔ نخست شامل دنباله‌های قدیمی‌تر و پنجرهٔ دوم شامل دنباله‌های جدیدتر هستند که به دنباله‌های آن‌ها به‌ترتیب، دنباله‌های پنجرهٔ مرجع W_R و دنباله‌های پنجرهٔ اکتشافی W_D گفته می‌شود. همان‌طورکه در شکل (۳) دیده می‌شود، این دو پنجره مجاور یکدیگر هستند و یک پنجرهٔ ترکیبی به‌طول 2w از دنباله‌های نگاره تشکیل می‌دهند. درنتیجه $W_R = \sigma_i, \sigma_{i+1}, \dots \sigma_{i+w-1}$ دنباله‌های پنجرهٔ مرجع است و $W_D = \sigma_{i+w}, \sigma_{i+w+1}, \dots \sigma_{i+2w-1}$ دنباله‌های پنجرهٔ اکتشافی است.

۲-۳-۳- محاسبهٔ بردار گونه‌ها

پس از شکستن نگارهٔ رویداد با استفاده از پنجره‌ها، برای هر پنجره، دنباله‌های پنجره‌های مرجع و اکتشافی به بردار گونه‌های D_R و D_D تبدیل می‌شوند (شکل (۳) و خط ۵ الگوریتم (۱))، سپس بین هر دو پنجره مقایسه صورت گرفته و فاصلهٔ بین بردارهای آن‌ها محاسبه می‌شود.

برای محاسبهٔ بردار گونه‌ها، در هر پنجره دنباله‌های مشاهده‌شده در گونه‌های خود دسته‌بندی و شمارش می‌شوند. هر بردار شامل دوتایی‌های سلسله‌رویدادهای گونه و تعداد تکرار آن‌ها است؛ به این صورت که در $D_{R_i} = ([[\lambda(e_n), \lambda(e_{n+1}), \dots], m], \dots)$، مقدار عبارت $\lambda(e_n), \lambda(e_{n+1}), \dots$ بیان‌گر سلسله‌رویدادهای آن گونه و m تعداد آن در بردار به‌دست‌آمده از پنجره اکتشافی iم است. به‌عنوان مثال شکل (۴) بردار به‌دست‌آمده از دنباله‌های شکل (۲) را نشان می‌دهد.

$$\begin{pmatrix} [[\text{ثبت ،درخواست بررسی ،کامل ،تصمیم‌گیری پرداخت خسارت}]. 2]. \\ [[\text{ثبت ،درخواست بررسی ،جزیی ،تصمیم‌گیری رد خسارت}]. 3] \end{pmatrix}$$

(شکل-۴): ایجاد بردار گونه‌ها از دنباله‌های شکل (۲)
(Figure-4): Generating variant vectors from figure (2) traces

[Downloaded from jsdp.rcisp.ac.ir on 2026-09-12]

[DOI: 10.52547/jsdp.19.1.75]

بردارهای گونه‌های به‌دست‌آمده از دنباله‌های دو پنجره اکتشافی و مرجع می‌توانند طول یکسان نداشته باشند؛ زیرا ممکن است، گونه‌های دو پنجره مجاور یکسان نباشند.

۳-۳-۳- اعمال آزمون آماری G برای شناسایی رانش‌ها

در این مرحله، برای تشخیص وجود اختلاف قابل ملاحظه بین دو بردار پنجره اکتشافی و پنجره مرجع، آزمون آماری G اعمال می‌شود (خط ۶ الگوریتم (۱)). آزمون G، یک آزمون فرض آماری و از نوع غیر پارامتریک است که دانش اولیه‌ای از توزیع آماری ندارد و برای مقایسهٔ توزیع آماری مشاهده‌شده با توزیع مورد انتظار به کار می‌رود. با استفاده از این آزمون، فاصلهٔ بین دو بردار گونه‌های هر دو پنجره محاسبه می‌شود.

همان‌طورکه گفته شد، ممکن است، این دو بردار طول یکسانی نداشته باشند. راه حل در نظر گرفته‌شده در این مقاله به این صورت است که بین دو بردار اجتماع گرفته می‌شود؛ سپس برای هر بردار، تعداد گونه‌هایی که در آن وجود دارد و در بردار دیگر وجود دارد، صفر در نظر گرفته می‌شود. به این ترتیب طول هر دو بردار یکسان می‌شود و به‌عنوان ورودی به آزمون آماری داده می‌شوند.

در اینجا فرض صفر آماری این است که توزیع مشاهده‌شده از فراوانی دنباله‌های هر گونه در پنجرهٔ مرجع با توزیع مشاهده‌شده در پنجرهٔ اکتشافی یکسان باشد. روش محاسبهٔ *gValue* در آزمون آماری G بدین صورت است:

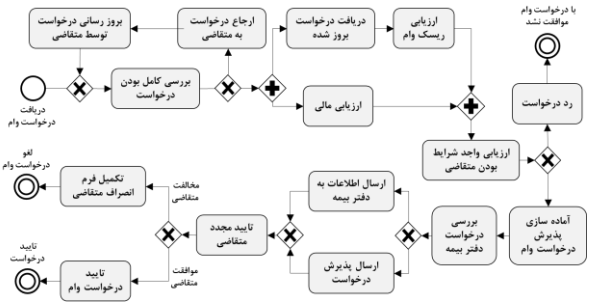
$$G = 2\sum_i O_i \cdot \ln\frac{O_i}{E_i} \qquad (۱)$$

که در آن O_i بردار گونه‌های مشاهده شده است (بردار مرجع D_R) و E_i گونه‌های مورد انتظار (بردار اکتشافی D_D) است؛ درصورتی‌که gValue به‌دست‌آمده از آزمون G از یک مقدار آستانه τ بیشتر باشد، به‌عنوان رانش مفهوم تشخیص داده می‌شود. از آنجایی‌که که آزمون G از توزیع آماری chi-square تبعیت می‌کند [11]، مقدار آستانه، مقدار سطح معنی‌دار[1] ۰/۰۵ در نظر گرفته می‌شود [12].

[1] significance level

۴- آزمایش‌ها و نتایج

در این بخش به ارزیابی روش ارائه‌شده برای شناسایی رانش‌های مفهوم در حالت ناگهانی پرداخته شده است. برای این منظور نتایج الگوریتم پیشنهادی را با الگوریتم‌های ارائه‌شده توسط ماراجی و همکاران [8]، بوز و همکاران [4] و سیلیگر و همکاران [10] مقایسه کرده و از بانک اطلاعاتی ماراجی و همکاران [8] ایجاد کرده‌اند، استفاده می‌کنیم. این بانک اطلاعاتی شامل ۷۲ نگارهٔ رویداد ساختگی است که از نمونه فرآیند درخواست وام [13] گرفته شده است و شکل (۵) مدل "پایه" آن را نشان می‌دهد که دارای ۱۵ رویداد، یک رویداد شروع و ۳ رویداد پایان است. این مدل شامل ساختارهای گوناگونی اعم از حلقه‌ها، انشعابات موازی و جایگزین است.



(شکل-۵): مدل پایهٔ فرآیند درخواست وام

(Figure-5): Base model of the loan application process

(جدول-۲): دوازده الگوی ساده تغییر[8]

(Table-2): 12 simple change patterns

کد	الگوی تغییر	گروه
re	ایجاد یا حذف بخشی از فرآیند	I
cf	شرطی‌کردن دو بخش یا پشت سر هم قراردادن آن‌ها	R
lp	حلقه داشتن یا نداشتن یک بخش	O
pl	موازی‌کردن دو بخش یا پشت سر هم قراردادن آن‌ها	R
cb	قابلیت پرش از قسمتی از فرآیند	O
cm	اضافه‌کردن یک بخش به انشعاب شرطی یا حذف یک بخش از آن	I
cd	همزمان‌سازی دو بخش	R
cp	تکثیر یک بخش	I
pm	موازی‌سازی یک بخش	I
rp	جایگزینی یک بخش	I
sw	جابه‌جایی دو بخش با یکدیگر	I
fr	تغییر فراوانی یک انشعاب در فرآیند	O

ماراجی و همکاران [8]، برای ایجاد رانش مفهوم در نگاره‌ها، مدل پایه را با استفاده از انواع تغییرات جریان کنترلی شامل دوازده الگوی ساده تغییر که در مقالهٔ وبر و همکاران [14] بیان شده است، دست‌کاری کرده و تغییر

[Downloaded from jsdp.rcisp.ac.ir on 2026-09-12]
[DOI: 10.52547/jsdp.19.1.75]

داده‌اند. این تغییرات که در جدول (۲) نشان داده شده‌اند، در سه گروه طبقه‌بندی شده‌اند: ایجاد[1] ("I")، ترتیب‌بندی مجدد[2] ("R") و قابلیت انتخاب[3] ("O")؛ سپس برای تشکیل الگوهای تغییر پیچیده‌تر، الگوهای ساده را با یکدیگر ترکیب کرده و نگاره‌های رویداد بیشتری ساختند ("IOR", "IRO", "OIR", "ORI", "RIO", "ROI") که در مجموع شش نگارهٔ رویداد ترکیبی و دوازده نگارهٔ رویداد پایه ساخته شده است.

نحوهٔ ترکیب این نگاره‌ها برای ایجاد رانش مفهوم، به این صورت است که ابتدا چهار نگاره به طول ۲۵۰، ۵۰۰، ۷۵۰ و ۱۰۰۰ دنباله از مدل "پایه" و همچنین برای هر کدام از ۱۸ مدل "تغییریافته"، ایجاد کرده و سپس پنج گروه از هرکدام از نگاره‌های پایه را با پنج گروه از هرکدام از نگاره‌های تغییریافته به‌صورت یکی در میان ترکیب کرده‌اند و درنتیجه برای هر کدام از هجده الگو، نگاره‌های رویدادی به‌طول ۲۵۰۰، ۵۰۰۰، ۷۵۰۰ و ۱۰۰۰۰ دنباله تشکیل شده است. به این صورت هر نگارهٔ رویداد دارای ۹ رانش مفهوم با فاصله‌های مشخص است که با داشتن این دانش، می‌توانیم صحت و دقت روش خود را ارزیابی کنیم [8].

برای ارزیابی روش خود، از معیار *F_1-score* استفاده می‌کنیم که یکی از معیارهای مهم شناسایی رانش مفهوم در داده‌کاوی است [15]. *F_1-score* میانگین هارمونیک بین دو مقدار *دقت*[4] $\boldsymbol{p}$ و *بازخوانی*[5] $\boldsymbol{r}$ است. معیار دقت، نسبت رانش‌های شناسایی شده صحیح (TP[6]) به کل رانش‌های شناسایی شده (TP + FP[7]) و معیار بازخوانی، نسبت رانش‌های شناسایی شده صحیح (TP) به کل رانش‌های واقعی (TP + FN[8]) است. به‌عبارت دیگر، مقدار دقت، میزان رانش‌های شناسایی شده‌ای که صحیح هستند و مقدار بازخوانی، میزان رانش‌های واقعی که کشف شده‌اند را نشان می‌دهند. روابط (۳) و (۴) محاسبهٔ این دو مقدار را نشان می‌دهد. مقدار *F_1-score* با استفاده از (۳) و (۴) طبق رابطهٔ (۵) محاسبه می‌شود:

$$p = \frac{TP}{TP + FP} \quad (۳)$$

$$r = \frac{TP}{TP + FN} \quad (۴)$$

$$F_1 = 2 \cdot \frac{p \cdot r}{p + r} \quad (۵)$$

برای محاسبهٔ مقدار آستانه کافی است در هیستوگرام مقادیر G که از رابطه (۱) به‌دست می‌آید، محلی که ۹۵ درصد مساحت زیر نمودار سمت چپ آن قرار می‌گیرد (با توجه به شکل (۶ الف)) را به‌دست بیاوریم. به‌عنوان مثال در شکل (۶ الف) در نقطه ۲۲۰/۲۸ در نمودار هیستوگرام ۹۵ درصد مقادیر G کمتر از این مقدار است و این بدان معناست اگر مقدار آستانه را برای شناسایی رانش مفهوم مقدار ۲۲۰/۲۸ در نظر بگیریم احتمال تشخیص غلط کمتر از پنج درصد خواهد بود.

برای مثال شکل (۶ب) رانش‌های مفهوم شناسایی‌شده در یکی از نگاره‌های رویداد استفاده‌شده در این مقاله را نشان می‌دهد. در این شکل هشت رانش شناسایی‌شده که هفت‌تای آن تشخیص درست (TP) و یکی از آن‌ها نادرست (FP) است. همچنین دو رانش دیگر در این نگاره وجود دارد که شناسایی نشده‌اند (FN). درنتیجه مقادیر $p = 0 \cdot 88$، $r = 0 \cdot 78$ و F1-score $= 0 \cdot 82$ هستند.

۱-۴- آزمایش شناسایی رانش مفهوم ناگهانی

در این آزمایش، به بررسی تأثیر اندازهٔ پنجره در شناسایی رانش‌های ناگهانی پرداخته شده است. همچنین نتایج به‌دست‌آمده را بین الگوهای تغییر مقایسه کرده و با پژوهش‌های قبلی صورت‌گرفته در این حوزه، مقایسه شده است.

۱-۱-۴- ارزیابی تأثیر تغییر در اندازه پنجره

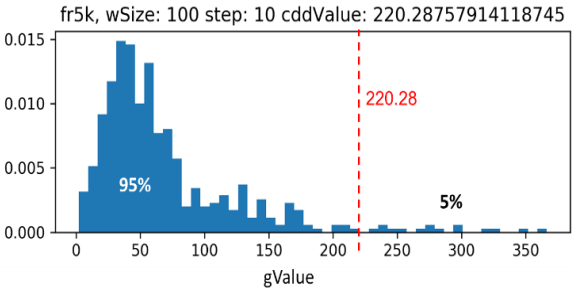
در آزمایش نخست، اندازهٔ پنجره‌ها تغییر داده شده است و تأثیر آن‌ها روی نتایج بررسی می‌شود. برای این منظور، روش ارائه‌شده را با پنجره‌های ثابت با طول‌های متفاوت روی هر ۷۲ نگارهٔ رویداد آزمایش کردیم. طول‌های پنجره در نظر گرفته‌شده بین ۲۵ تا ۲۵۰ با افزایش ۲۵تایی است. شکل (۷)، *F_1-score* به‌دست‌آمده از چهار دسته از نگاره‌ها (نگاره‌های به طول ۲۵۰۰، ۵۰۰۰، ۷۵۰۰ و ۱۰۰۰۰) را نشان می‌دهد که برای هر دسته، بین F-scoreهای همهٔ هجده الگوی تغییر میانگین گرفته شده است.

شکل (۷) نشان می‌دهد، در ابتدا با پنجره‌ای به‌طول ۲۵، کمترین مقدار F_1-score به‌دست آمده است (به‌طور میانگین ۰/۳۳). سپس با افزایش طول پنجره و

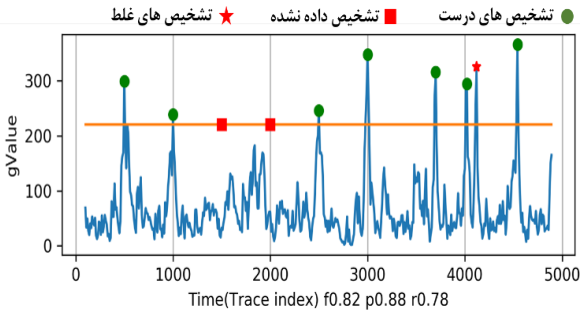
[1] Insertion
[2] Resequentialization
[3] Optionalization
[4] precision
[5] recall
[6] True Positive
[7] False Positive
[8] False Negative

[Downloaded from jsdp.rcisp.ac.ir on 2026-09-12]
[DOI: 10.52547/jsdp.19.1.75]

درنتیجه در اختیارداشتن داده‌های کافی، مقدار F_1-score بهبود یافته است. بهترین نتایج در پنجره به طول ۲۲۵ به مقدار ۹۸٪ به‌دست آمده به‌طوری‌که در دسته نگاره‌های ۲۵۰۰تایی در این پنجره، F_1-score به ۰/۹۹رسیده است.



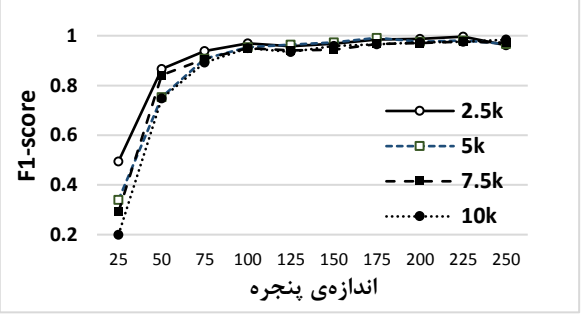
(الف) هیستوگرام مقادیر G



(ب) مقادیر G برای پنجره به طول ۱۰۰ روی نگاره رویداد fr با ۵۰۰۰ دنباله

(شکل-۶): نمونه‌ای از رانش‌های شناسایی شده، رانش‌های شناسایی نشده و تشخیص‌های غلط در یک نگارهٔ رویداد

(Figure-6): An example of identified, unidentified and incorrect identified drifts in an event log.



(شکل-۷): F1-score به‌دست‌آمده از طول پنجره‌های متفاوت

(Figure-7): F1-score obtained from different window lengths

به‌طورتقریبی در نیمی از پنجره‌ها (به‌خصوص پنجره‌های کوچک‌تر)، نگارهٔ رویداد ۲۵۰۰ تایی نتایج بهتری را نشان می‌دهد. همچنین نتایج نشان می‌دهد که اغلب در پنجره‌های کوچک‌تر، نگاره‌های با طول کمتر نتایج بهتری ارائه داده‌اند و از آنجا که فواصل رانش‌ها در نگاره‌های کوچک‌تر، کمتر است، می‌توان نتیجه گرفت که در نگارهٔ ۲۵۰۰تایی تعداد داده‌های موجود در پنجره‌ها برای تشخیص رانش‌ها به نسبت طول پنجره‌ها مناسب‌تر هستند. همچنین در پنجرهٔ ۲۵۰تایی، نتایج به‌طورکامل برعکس می‌شود؛ یعنی نگاره‌های بزرگ‌تر نتایج بهتری ارائه داده‌اند که نشان می‌دهد در این پنجره اطلاعات برای تشخیص درست رانش در نگارهٔ ۱۰هزارتایی مناسب‌تر است.

۲-۱-۴- ارزیابی و مقایسه الگوهای تغییر

در آزمایش دوم، F_1-score هر هجده الگوی تغییر را به دست آوردیم. برای این‌کار اندازهٔ پنجره را ثابت و به‌طول ۱۰۰ در نظر گرفتیم (مانند روش‌های پیشین) تا قابل مقایسه با پژوهش‌های پیشین باشد. در این آزمایش از میانگین F_1-score به‌دست‌آمده بین سه دسته از نگاره‌ها (نگاره‌های به‌طول ۵۰۰۰، ۷۵۰۰ و ۱۰۰۰۰) استفاده شده است. در این آزمایش، نتایج به‌دست‌آمده را با پژوهش‌های قبلی مقایسه کرده‌ایم که در شکل (۷) نمایش داده شده است. در روش سیلیگر [10] و ماراجی [8] از پنجره تطبیقی استفاده شده است که روش ایجاد پنجره‌های تطبیقی آن‌ها با یکدیگر متفاوت است و در روش بوز از پنجره ثابت به‌طول ۱۰۰ استفاده شده است. همان‌طورکه در شکل (۸) قابل مشاهده است، در سیزده الگوی تغییر، روش پیشنهادی F1-Score بهتری نسبت به روش‌های پیشین به‌دست آورده است. به‌جز چهار الگوی fr، cp، IOR و OIR، نتایج به‌دست‌آمده، بهتر یا به‌طورتقریبی برابر با روش‌های دیگر است. همچنین میانگین کلی به‌دست آمده از F_1-score در روش پیشنهادی ۰/۹۵۱ و بالاتر از روش‌های پیشین است.

همان‌طورکه در جدول (۲) مشهود است روش پیشنهادی در دوازده مورد از هجده الگوی تغییر به تشخیص صد درصدی رسیده است؛ درحالی‌که روش ماراجی [8] در دو مورد، سیلیگر در سه مورد و بوز فقط در پنج مورد به تشخیص صد درصدی رسیده است. از مهم‌ترین چالش‌هایی که در روش‌های پیشین وجود داشته (بجز روش ماراجی در تشخیص fr که جای تعجب است) و همچنان در روش پیشنهادی هم دیده می‌شود، تشخیص الگوی fr و cp است که در روش پیشنهادی به‌ترتیب به دقت تشخیص ۷۹ درصدی و ۷۲ درصدی رسیده است، که علت آن را می‌توان در عدم تشخیص تغییرات فراوانی مسیرهای انشعابی در مدل فرآیند در روش پیشنهادی دانست.

[Downloaded from jsdp.rcisp.ac.ir on 2026-09-12]

[DOI: 10.52547/jsdp.19.1.75]

۵- نتیجه‌گیری

در این مقاله روشی برای شناسایی رانش‌های مفهوم براساس تغییر در توزیع گونه‌های نگارهٔ رویداد ارائه شده است. هر کدام از گونه‌ها، سلسله‌ای از رویدادهای فرآیند را دربر می‌گیرد و تغییر در رفتار فرآیند، در توزیع گونه‌ها تأثیر می‌گذارد. در این روش با استفاده از پنجره‌های مرجع و اکتشافی، گونه‌های فرآیند استخراج و به‌وسیلهٔ آزمون آماری با یکدیگر مقایسه شدند. همچنین الگوریتم خود را روی یک نگارهٔ رویداد معرفی شده در سابقهٔ پژوهش آزمایش کردیم که نتایج به‌دست‌آمده نسبت به پژوهش‌های قبلی به میزان ٪۰/۰۶ بهبود داشته است. مزیت این روش، سادگی و پیچیدگی کمتر نسبت به روش‌های پیشین است.

در این مقاله روش ارائه‌شده روی رانش‌های ناگهانی اعمال و روی رانش‌های تدریجی، دوره‌ای و افزایشی کمتر مطالعه و پژوهش انجام شده است و می‌تواند از جمله پژوهش‌هایی باشد که در آینده مورد توجه پژوهش‌گران قرار گیرد؛ علاوه‌بر آن روش پیشنهادی روی دو الگوی تغییر fr و cp هنوز به دقت تشخیص بالای ۰/۸ نرسیده است و می‌توان با الهام از روش سیلیگر و همکاران [10] که در این دو نوع الگوی تغییر نتیجه‌ای خوبی داشته‌اند، بهبود قابل توجهی در تشخیص این الگوها ایجاد کرد. روش پیشنهادی در این مقاله قابل تعمیم برای تشخیص هر نوع رانش مفهومی در پایگاه دادگان دیگر نیز است و در اینجا برای شناسایی رانش‌های مفهوم در مدل فرآیند استفاده شده است.

6- References ۶- مراجع

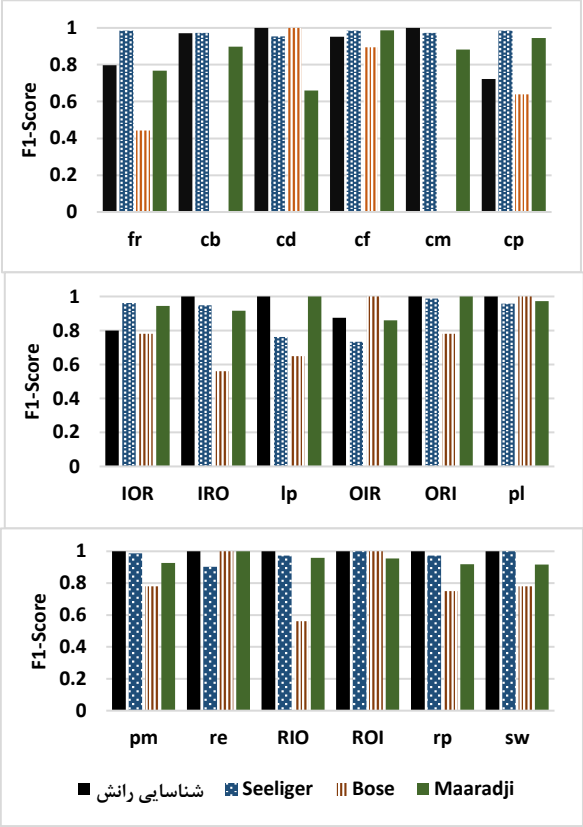
[1] W. Van Der Aalst, “Process mining: discovery, conformance and enhancement of business processes,” *Springer*, vol. 2, pp.157-158, 2011.

[2] W. M. P. Aalst, “Process Mining: Overview and Opportunities,” *ACM Trans. Manag*. Inf. Syst., vol. 3, no. 2, pp. 1–17, 2012.

[3] W. Van Der Aalst, A. Adriansyah, A. K. A. De Medeiros, F. Arcieri, T. Baier, T. Blickle, J. C. Bose, P. Van Den Brand, R. Brandtjen, J. Buijs, and others, “Process mining manifesto,” in International Conference on Business Process Management, 2011, pp. 169–194.

[4] R. P. J. C. Bose, W. M. P. van der Aalst, I. Zliobaite, and M. Pechenizkiy, “Dealing with concept drifts in process mining.,” *IEEE Trans. neural networks Learn.* Syst., vol. 25, no. 1, pp. 154–71, 2014.

[5] R. Klinkenberg and T. Joachims, “Detecting Concept Drift with Support Vector Machines.,”



(شکل-۸): مقایسه F1-score به‌دست‌آمده در روش پیشنهادی برای هر الگوی تغییر با روش ماراجی و همکاران [8]، بوز و همکاران [4] و سیلیگر و همکاران [10]

(Figure-8): The F1 score obtained for each pattern of change and comparison with Maaradji et al. [8], Bose et al. [4] and Seeliger et al. [10]

(جدول-۲): نتایج F_1-score به‌دست‌آمده در مقایسه با پژوهش‌های پیشین با اندازه پنجره ۱۰۰

(Table-2): The F_1-score results of the proposed alogoritm and comparison with others works by window size 100

ماراجی [8]	بوز [4]	سیلیگر [10]	روش پیشنهادی	
F_1-score				
0.9853	0.442	0.7677	**0.7965**	fr
0.9722		0.8985	**0.972**	cb
0.9546	1	0.6599	**1**	cd
0.9853	0.895	0.9868	**0.9523**	cf
0.9722		0.8833	**1**	cm
0.9853	0.6395	0.9444	**0.7228**	cp
0.9606	0.7805	0.9444	**0.8**	IOR
0.9487	0.5612	0.9167	**1**	IRO
0.7618	0.6484	1	**1**	lp
0.7331	1	0.8603	**0.875**	OIR
0.9869	0.7805	1	**1**	ORI
0.9575	1	0.9722	**1**	pl
0.9869	0.7805	0.9265	**1**	pm
0.9036	1	1	**1**	re
0.9722	0.5612	0.9591	**1**	RIO
1	1	0.9559	**1**	ROI
0.9722	0.75	0.9194	**1**	rp
1	0.7805	0.9167	**1**	sw
0.9465	**0.7887**	**0.9173**	**0.9510**	میانگین

[Downloaded from jsdp.rcisp.ac.ir on 2026-09-12]

[DOI: 10.52547/jsdp.19.1.75]

in ICML, 2000, pp. 487–494.

[6] B. R. P. J. C. . b, V. D. A. W.M.P.a, Ž. I.a, and P. M.a, "Handling concept drift in process mining," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2011, vol. 6741 LNCS, pp. 391–405.

[7] J. Martjushev, R. P. Jagadeesh Chandra Bose, and W. M. P. van der Aalst, "Change point detection and dealing with gradual and multi-order dynamics in process mining," in Lecture Notes in Business Information Processing, 2015, vol. 229, pp. 161–178.

[8] A. Maaradji, M. Dumas, M. La Rosa, and A. Ostovar, "Fast and accurate business process drift detection," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2015, vol. 9253, pp. 406–422.

[9] T. Li, T. He, Z. Wang, Y. Zhang, and D. Chu, "Unraveling Process Evolution by Handling Concept Drifts in Process Mining," in Proceedings - 2017 IEEE 14th International Conference on Services Computing, SCC 2017, 2017, pp. 442–449.

[10] A. Seeliger, T. Nolle, and M. Mühlhäuser, "Detecting Concept Drift in Processes using Graph Metrics on Process Graphs," in Proceedings of the 9th Conference on Subject-Oriented Business Process Management, 2017, pp. 1–10.

[11] P. Harremoës and G. Tusnády, "Information divergence is more χ 2-distributed than the χ 2-statistics," in 2012 IEEE International Symposium on Information Theory Proceedings, 2012, pp. 533–537.

[12] R. Nuzzo, "Scientific method: statistical errors," *Nat. News*, vol. 506, no. 7487, pp. 150, 2014.

[13] M. Dumas, M. La Rosa, J. Mendling, H. A. Reijers, and others, Fundamentals of business process management, vol. 1. Springer, 2013.

[14] B. Weber, M. Reichert, and S. Rinderle-Ma, "Change patterns and change support features--enhancing flexibility in process-aware information systems," *Data Knowl. Eng.*, vol. 66, no. 3, pp. 438–466, 2008.

[15] S.-S. Ho, "A martingale framework for concept change detection in time-varying data streams," in Proceedings of the 22nd international conference on Machine learning, 2005, pp. 321–327.

[16] A. Ostovar, M. Abderrahmane, M. La Rosa, A. H. ter Hofstede, and B. F. van Dongen., "Detecting drift from event streams of unpredictable business processes," in International Conference on Conceptual Modeling, 2016, pp. 330–346.

[17] R. Accorsi and T. Stocker, "Discovering workflow changes with time-based trace clustering," in International Symposium on Data-Driven Process Discovery and Analysis, 2011, pp. 154–168.

[18] B. Hompes, J. C. A. M. Buijs, W. M. P. van der Aalst, P. Dixit, and H. Buurman, "Detecting Change in Processes Using Comparative Trace Clustering.," in SIMPDA, 2015, pp. 95–108.

[19] J. Carmona and R. Gavalda, "Online techniques for dealing with concept drift in process mining," in International Symposium on Intelligent Data Analysis, 2012, pp. 90–102.

[20] F. Khojasteh, M. Kahani and B. Behkamal, "Concept drift detection in business process logs using deep learning" *Signal and Data Processing*, vol. 46, no. 4, pp. 33–48, 2021

[۲۰] فاطمه خجسته، محسن کاهانی و بهشید بهکمال "شناسایی رانش مفهومی در نگاره‌های فرایند کسب‌وکار با استفاده از یادگیری عمیق." پردازش علائم و داده‌ها، شماره ۴، نسخه ۴۶، صفحه ۳۳-۴۸، ۱۳۹۹

فرشته جوادزاده دانشجوی کارشناسی ارشد مهندسی کامپیوتر در دانشگاه گلستان، دورهٔ کارشناسی خود را از دانشگاه گلستان در رشتهٔ مهندسی نرم‌افزار کامپیوتر دریافت کرده است. زمینه پژوهشی ایشان فرآیندکاوی و رانش مفهوم در فرآیندهای کسب و کار است.

نشانی رایانامه ایشان عبارت است از:

fereshteh.jvz@gmail.com

مهدی یعقوبی عضو هیأت علمی دانشگاه گلستان، گروه مهندسی کامپیوتر، فارغ‌التحصیل مقطع دکترا گرایش هوش مصنوعی از دانشگاه صنعتی شاهرود، مقطع کارشناسی ارشد گرایش نرم‌افزار از دانشگاه صنعتی امیرکبیر است. در حال حاضر زمینه پژوهشی ایشان سامانه‌های اطلاعاتی مبتنی بر فرآیندهای کسب و کار، فرآیندکاوی و داده‌کاوی است.

نشانی رایانامه ایشان عبارت است از:

m.yaghoubi@gu.ac.ir

[Downloaded from jsdp.rcisp.ac.ir on 2026-09-12]

[DOI: 10.52547/jsdp.19.1.75]

سهیلا کرباسی عضو هیأت علمی دانشگاه گلستان، گروه مهندسی کامپیوتر، فارغ‌التحصیل مقطع دکترا گرایش سامانه‌های اطلاعاتی از دانشگاه دانشگاه پل ساباتیه در کشور فرانسه، مقطع کارشناسی ارشد گرایش نرم‌افزار از دانشگاه فردوسی مشهد است. در حال حاضر در زمینه سامانه‌های اطلاعاتی، جستجوی اطلاعات، داده‌کاوی پژوهش و مطالعه می‌کند.
نشانی رایانامه ایشان عبارت است از:
s.karbasi@gu.ac.ir

[Downloaded from jsdp.rcisp.ac.ir on 2026-09-12]

[DOI: 10.52547/jsdp.19.1.75]