مدیریت فناوری اطلاعات
دانشکدهٔ مدیریت دانشگاه تهران
دورهٔ ۷، شمارهٔ ۴
زمستان ۱۳۹۴
ص. ۸۴۴- ۸۲۵

کشف و یادگیری پدیده‌های استثنایی با به‌کارگیری تئوری استثنائات و تئوری رضایتمندی

مسعود عابسی[۱]، الهه حاجی گل‌یزدی[۲]

چکیده: منطق یادگیری از استثنائات چالش مهمی در حیطهٔ داده‌کاوی و کشف دانش است. در این پژوهش بر اساس تئوری استثنائات و رضایتمندی، الگویی نوین برای بهبود شایان توجه میزان اعتماد و اطمینان به کشف و یادگیری از استثنائات ارائه می‌شود. ابتدا به‌کمک رویکرد تلفیقی پیشنهادی بر اساس تئوری استثنائات، حدود رفتار نرمال و استثنایی داده‌ها مشخص می‌شود و پس از آن با به‌کارگیری تئوری رضایتمندی، راه‌حل‌های رضایت‌بخش به‌دست می‌آید. استخراج دانش از داده‌های نرمال و استثنایی به‌کمک رویکرد یادگیری پایین به بالا و به‌کارگیری الگوریتم پیشنهادی RISE ارتقایافته صورت می‌گیرد. به‌منظور تعیین کارایی الگوی پیشنهادی، کشف سهام استثنایی در پایگاه اطلاعاتی بازار بورس ایران هدف قرار گرفت. برتری نتایج روش پیشنهادی با نتایج به‌دست‌آمده از به‌کارگیری سایر الگوریتم‌های داده‌کاوی، روزنه‌ای برای توجه به رویکرد پیشنهاد شده است. همچنین با بهره‌مندی از شاخص g-means میزان دقت این روش سنجیده شد. نتایج نشان داد روش پیشنهادشده از قابلیت شناسایی و یادگیری از داده‌های استثنایی برخوردار است.

واژه‌های کلیدی: تئوری استثنائات، تئوری رضایتمندی، داده‌کاوی، یادگیری پایین به بالا.

۱. استادیار مدیریت صنعتی، دانشکدهٔ صنایع، دانشگاه یزد، یزد، ایران
۲. دانشجوی دکتری مهندسی صنایع، دانشکدهٔ صنایع، دانشگاه یزد، یزد، ایران

تاریخ دریافت مقاله: ۱۳۹۳/۰۸/۲۸
تاریخ پذیرش نهایی مقاله: ۱۳۹٤/۰۹/۰٦
نویسندهٔ مسئول مقاله: الهه حاجی گل‌یزدی
E-mail: elahehajigol@gmail.com

Archive of SID
www.SID.ir

مقدمه

یادگیری از داده‌های استثنایی، یکی از موضوعات مهم و پیچیدهٔ پژوهش است که کاربردهای فراوانی در دنیای امروز دارد. کشف و مطالعهٔ الگوهای استثنایی در پایگاه داده، از جذابیت خاصی برخوردار است که از کاربردهای آن می‌توان به کشف ناهنجاری‌های سیستم و تجزیه‌وتحلیل داده‌های پزشکی اشاره کرد. استثنایی به پدیده‌های موجود در پایگاه داده اطلاق می‌شود که رفتاری متفاوت از الگوهای اصلی و مورد انتظار از خود بروز می‌دهند. اغلب در فرایند داده‌کاوی، یادگیری از داده‌های استثنایی دشوار است؛ زیرا داده‌های غیرعادی حجم کمی از مجموع داده‌های پایگاه داده را دربرمی‌گیرند. همچنین مدل‌های معمول یادگیری کارایی کافی را برای شناسایی و یادگیری از این داده‌ها ندارند؛ زیرا این مدل‌ها با فرض توازن توزیع داده‌ها عمل می‌کنند. این فرضیه سبب می‌شود الگوریتم‌های یادگیری معمول، بر اساس معیارهای هزینهٔ خطا، دقت و توزیع دسته‌ها، در شناسایی داده‌های استثنایی عملکرد ضعیفی داشته باشند. بنابراین مطالعه و یافتن راهکارهای عملی برای غلبه بر مشکلات شناسایی و یادگیری از داده‌های استثنایی، ارزشمند خواهد بود.

مسئلهٔ کشف داده‌های استثنایی با عنوان کاوش دسته‌های استثنایی، کشف تصادفی[1]، کشف موارد جدید[2] و مسئلهٔ کاوش استثنائات[3] نیز شناخته می‌شود. داده‌های استثنایی از لحاظ ماهیت با داده‌های پرت بسیار تفاوت دارند. شناسایی داده‌های استثنایی به‌دلیل دانش نهفته‌ای که دارند، بسیار ارزشمند است. داده‌های پرت، اغلب به‌دلیل خطای انسانی، خطای ماشین (ابزار) یا خطا در سیستم رخ می‌دهند، در حالی‌که داده‌های استثنایی به‌دلیل انحراف طبیعی از میانگین (قد بلند)، رفتار تقلبی یا تغییر در رفتار سیستم، ایجاد می‌شوند. در واقع داده‌های استثنایی داده‌های مثبت پیش‌بینی‌نشده و داده‌های پرت داده‌های منفی پیش‌بینی نشده‌اند.

مبنای نظری پژوهش حاضر، به‌کارگیری تئوری استثنائات[4] و تئوری رضایتمندی[5] (سیمون ۱۹۷۴) است که به‌منظور مدیریت رخدادهای غیرعادی، شامل کشف حالت‌های استثنایی، استخراج دانش و یادگیری از فرایندهای استثنایی، به‌کار گرفته می‌شود. بر اساس تئوری استثنائات برخی از داده‌ها رفتار متفاوتی از رفتار مورد انتظار نشان می‌دهند. از آنجا که کشف استثنائات با تفسیر یافته‌های مشاهده‌شده در محتوای حوزهٔ مسئله ارتباط دارد، نقطهٔ آغاز مناسب

1. Chance discovering
2. Novelty detection
3. Exception mining
4. Abnormality Theory
5. Satisficing Theory

Archive of SID

www.SID.ir

برای توضیح استثنائات، مدل مفهومی انواع دانشی است که رفتار عـادی و غیرعـادی سیسـتم را شرح می‌دهند.

از دید تئوری رضایتمندی، بسیاری از مسائل پیرامون ما به بهینه‌سـازی نیـاز ندارنـد و کـافی است راه‌حل رضایت‌بخشی برای آنها کشف شود. رویکـرد رضـایتمندی در حـل مسـئله، هماننـد لبه‌های قیچی عمل می‌کند؛ زیرا فضای مسئله را به فضای کوچک‌تری که شـامل راه‌حـل‌هـای رضایت‌بخش می‌شود، تبدیل می‌کند. این رویکرد فضای مسئله را کوچک‌کرده و پیچدگی مسئله را بی‌آنکه تغییر عمده‌ای در کیفیت اطلاعات ایجاد کند، به‌طور شایان توجهی کاهش می‌دهد.

در این پژوهش برآنیم چارچوبی برای تلفیق تئوری اسـتثنائات و تئـوری رضـایتمندی ایجـاد کنیم و از آن برای کشف استثنائات بهره ببریم. ابتدا به‌منظـور شناسـایی قـوانین اسـتثنایی‌شـدن داده‌ها، مدل مفهومی رفتار غیرعادی سهام شکل می‌گیرد که دربردارندهٔ آستانه‌های قابـل قبـول برای استثنایی‌شدن داده‌هاست. پس از شناسایی داده‌های استثنایی، قـوانین بـه‌کمـک الگـوریتم RISE ارتقایافته[1] استخراج می‌شود. به‌منظور تعیـین کـارایی الگـوی پیشـنهادی، حیطـهٔ عملـی مطالعه رصد رفتار متفاوت انواع سهام برای شناسایی و تشکیل سبد سـهام اسـتثنایی بـر اسـاس تئوری استثنائات و تئوری رضایتمندی در نظر گرفته شده است. سبد سـهام اسـتثنایی سـهامی را پوشش می‌دهد که رفتارها و ویژگی‌هـای مثبـت غیرمعمـولی در بـازار از خـود نشـان مـی‌دهـد؛ به‌طوری که سبب کسب بیشترین ثروت شود. بنابراین، ساختن سیستم کارا بـرای انتخـاب سـبد استثنایی از سهام موجود در بازار بورس به تفسیر دو مسئلهٔ اصلی نیاز دارد. اول؛ سیستم باید یـاد بگیرد، بفهمد و به‌صورت خودکار مدلی از رفتارهـای غیرمعمـول و اسـتثنایی از سـهام در دسـت بررسی بسازد. دوم؛ سیستم باید بتواند از دانشی که یاد گرفته است استفاده کند و بـا نظـارت بـر رفتارهای فعلی سهام، هرگونه انحراف در الگوهای نرمال فعالیت را کشف کند و از آنها به‌منظـور ساختن سبد سهام استثنایی بهره‌مند شود.

پژوهش بدین شرح ادامه می‌یابد؛ نخسـت در بخـش پیشـینهٔ پـژوهش، سـوابق شناسـایی و یادگیری از داده‌های استثنایی، تئـوری اسـتثنائات و چگـونگی شناسـایی داده‌هـای اسـتثنایی بـا بهره‌مندی از این تئوری بررسی می‌شود و با تشریح رویکرد رضایتمندی در کشف استثنائات ادامه می‌یابد. سپس به‌منظـور اسـتخراج قـوانین پنهـان در پایگـاه داده، بـه معرفـی الگـوریتم RISE ارتقایافته پرداخته می‌شود. بخش روش‌شناسی پژوهش، به ارائـهٔ مـدل پیشـنهادی بـرای کشـف داده‌های استثنایی و استخراج قوانین موجود در پایگاه داده می‌پردازد. سپس یافته‌های پژوهش در

1. Enhanced RISE algorithm

Archive of SID

www.SID.ir

حیطهٔ شناسایی و یادگیری سهام استثنایی بیان می‌شود. بخش آخر نیز به خلاصه‌ای از نتایج اختصاص دارد.

پیشینهٔ پژوهش

مطالعه در زمینهٔ کشف و یادگیری داده‌های استثنایی با کاوش حول مسئلهٔ یادگیری از داده‌های نامتوازن پا به عرصهٔ وجود گذاشت. روند تاریخی پیشرفت سیستم‌های یادگیری از داده‌ها و حالت‌های نامتوازن، در تحقیقات چاولا، جاپکویچ و ایز (۲۰۰۷) و ویس (۲۰۰۴) بررسی شدند. چن، چن، سو و ژنگ (۲۰۰۸) رویکرد داده‌کاوی را بر اساس دانه‌های اطلاعاتی[1] برای کشف دانش از داده‌های نامتوازن به‌کار بردند. بورز و وندن پل (۲۰۰۹) موضوع از دست‌دادن مشتری را به‌منزلهٔ رخداد استثنایی در صنعت خدمات بررسی کردند که هدف از آن ارتقای کارایی روش‌های نمونه‌سازی با به‌کارگیری سنجه‌های ارزیابی مناسب‌تر بود. کو ، ژو و ژانگ (۲۰۰۸) با به‌کارگیری ابزار آماری به کاوش الگوی فعالیت‌های هدف (غیرعادی)، فعالیت‌های مغایر با هدف و فعالیت‌هایی با تأثیر معکوس پرداختند، در حالی‌که ساختار داده نامتوازن در نظر گرفته شد. مسئلهٔ کشف تقلب در سیستم‌های بانکی (وثوق، تقوی‌فرد و البرزی، ۱۳۹۳و محقر، لوکاس، حسینی و منشی، ۱۳۸۹) و در صنعت بیمه (تقوی‌فرد و جعفری، ۱۳۹۴) برای یافتن رفتارهای نادر پرخطر برای سیستم مورد بررسی، از مصادیق دیگر کشف استثنائات است.

روش‌های ارائه‌شده در مطالعات پیشین، فقط برای شناسایی موردها و حالت‌های غیرعادی اعم از داده‌های غیرعادی مثبت (استثنائات) و غیرعادی منفی (داده‌های پرت) به‌کار رفته‌اند و تاکنون روشی برای تشخیص و شناسایی داده‌های استثنایی از داده‌های پرت و یادگیری از داده‌های استثنایی ارائه نشده است. پژوهش حاضر به بحث و بررسی چالش‌های پروسهٔ کاویدن و کشف دانش از داده‌های استثنایی می‌پردازد و به‌دنبال یافتن راهکار مناسب برای یادگیری از داده‌ها و حالت‌های استثنایی به‌کمک تئوری استثنائات است.

تئوری استثنائات

رویکردهای متفاوتی در انواع زمینه‌های علمی و عملی برای استثنائات وجود دارد که از جملهٔ آنها می‌توان به استثنائات موضوعی[2]، استثنائات آماری[3]، استثنائات ژنتیکی[4]، استثنائات بیولوژیک[5]،

---

1. Information granuls
2. Subjective abnormality
3. Statistical abnormality
4. Genetic abnormality
5. Biological abnormality

Archive of SID

www.SID.ir

انحراف از استانداردهای اجتماعی[1] و رویکردهای تئوریک اشاره کرد. نخستین بار هافمن (۱۹۷۹) تئوری استثنائات را در علم ژنتیک مطرح کرد. پس از ظهور تئوری استثنائات در علم ژنتیک، مک‌کارتی در سال ۱۹۸۰ بحث تئوری استثنائات را در استنتاج آغاز کرد. مک‌کارتی با مثال معروفی به معرفی استثنائات پرداخت: این قانون خاص را در نظر بگیرید «پرندگان به‌صورت معمولی می‌توانند پرواز کنند» اگر x پرنده‌ای خاص باشد، AB(x) به این معناست که «x دستهٔ استثنایی از جامعهٔ پرندگانی است که نمی‌توانند بپرند» (مانند پنگوئن). مک‌کارتی با به‌کارگیری قوانین علی و معلولی به شناسایی استثنائات پرداخت. پس از رویکرد مک‌کارتی، رویکردهای تئوریک به مسئلهٔ شناخت استثنائات مطرح شد. رویکردهای تئوریک به استثنائات بر اساس ایجاد یا توسعهٔ تئوری توسط شخصی آغاز می‌شود. اگر برای مسئله بتوان حیطهٔ نرمالی تعریف کرد، استثنائات شکستی در توسعهٔ این تئوری در نظر گرفته می‌شوند.

تئوری مفهومی است که برای یافتن استثنائات به‌کار می‌رود و بسته به نوع دانش موجود عمل می‌کند. در واقع استثنائات با تفسیر یافته‌های مشاهده‌شده در محتوای حوزهٔ مسئله ارتباط دارند. نقطهٔ آغاز مناسب برای توضیح استثنائات، مدل مفهومی انواع دانشی است که در کاربردهای مختلف استثنائات اهمیت دارد. دانش ضمنی در سیستم یافتن استثنائات ممکن است بر پایهٔ توضیحی از ساختار عادی و رفتار وظیفه‌ای سیستم یا بیان رفتار غیرعادی سیستم باشد. یافته‌های جمع‌آوری‌شده‌ای که با رفتار نرمال سیستم مطابقت دارند «یافته‌های نرمال» نام دارند و در غیر این صورت به آنها «یافته‌های غیر نرمال» می‌گویند. این تئوری‌ها بر اساس انواع دانش کسب‌شده و یافته‌های مشاهده‌شده شکل می‌گیرند؛ مانند تئوری‌های «انحراف از ساختار و رفتار نرمال» و «انطباق با رفتار غیر نرمال».

تئوری کشف استثنائات بر اساس سازگاری

ری ریترز (۱۹۸۷) نخستین‌بار این تئوری را به‌منزلهٔ چارچوبی منطقی برای کشف عارضهٔ سیستم‌های فیزیکی با استفاده از مدل ساختار و رفتار نرمال سیستم مطرح کرد. ایدهٔ اصلی تئوری یادشده، حاصل پژوهش‌های ریترز و مک‌کارتی دربارهٔ مسئلهٔ استنتاج غیریکنواخت[2] است. بر اساس این تئوری، استثنائات بر مبنای مقایسهٔ داده‌های مشاهده‌شده با ساختار و رفتار نرمال سیستم و مغایرت با آن کشف می‌شوند.

1. Violation of socially accepted standards
2. Non-monotonic Reasoning

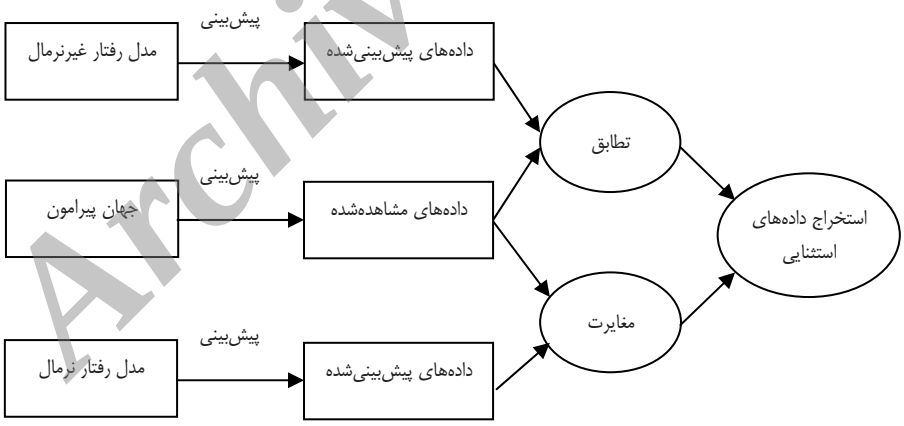
Archive of SID

www.SID.ir

تئوری کشف استثنائات براساس تطابق با رفتار غیرعادی

کشف استثنائات بر اساس تطابق با رفتار غیرعادی و با درنظرگرفتن دانش رفتارهای استثنایی سیستم عمل می‌کند؛ به نحوی‌که به شبیه‌سازی رفتار غیرعادی سیستم می‌پردازد. با فرض وجود عیب‌های مشخص، می‌توان برخی از حالت‌های قابل مشاهدهٔ غیرعادی را پیش‌بینی کرد. در عمل، حوزهٔ دانش به‌کاربرده‌شده در این سیستم، ارتباطات علت و معلولی است. تئوری‌های معتبر در این زمینه، تئوری‌های پوشش مجموعه‌ها[1] و استدلال قیاسی[2] است. جیمز رگیا، تئوری پوشش مجموعه‌ها را که بر اساس دانش علی به‌صورت ارتباطات ریاضی بیان می‌شود، ارائه کرد. توراسا و کنزوله نیز تئوری استدلال قیاسی را که برمبنای منطق است مطرح کردند. پژوهش حاضر از رویکرد پوشش مجموعه‌ها بهره می‌برد.

به‌منظور ارتقای دقت در کشف استثنائات، چارچوب تئوریک تلفیقی جدیدی مرکب از تئوری کشف استثنائات بر اساس سازگاری و تئوری کشف استثنائات بر اساس تطابق با رفتار غیرعادی به‌صورت شکل ۱ پیشنهاد می‌شود. استثنائات بر مبنای میزان سازگاری داده‌های مشاهده‌شده با مدل رفتار غیرعادی و مغایرت داده‌های مشاهده‌شده با رفتار عادی سیستم کشف می‌شوند. یافته‌های جمع‌آوری‌شده‌ای که با رفتار عادی سیستم مغایرت دارند یا مطابق با رفتار غیرعادی سیستم‌اند «یافته‌های استثنایی» نامیده می‌شوند.



شکل ۱. الگوی پیشنهادی برای شناسایی داده‌های استثنایی با به‌کارگیری تئوری استثنائات

1. Set-covering theory
2. Abduction theory

Archive of SID

www.SID.ir

رویکرد رضایتمندی در کشف استثنائات

در سال ۱۹۵۷ هربرت سیمون واژهٔ رضایتمندی را در زمینهٔ تصمیم‌گیری مسائل اقتصادی مطرح کرد و میر (۱۹۸۳) و وارد (۱۹۹۲) این مفهوم را توسعه دادند. پس از آن محققان زیادی (اسلات، ۱۹۸۹ و فلنر، گوت و مارتین، ۲۰۰۶) در زمینه‌های گوناگونی از این مفهوم استفاده کردند. برای مثال، پلوس (۱۹۹۳) از رویکرد رضایتمندی در علم روان‌شناسی و کانلیسک (۱۹۹۶) در علم اقتصاد بهره برد. مالاخو (۲۰۱۴) تئوری رضایتمندی را برای تجزیه‌وتحلیل مسائل اقتصادی خرید به‌کار برد. وی برای تعیین میزان اقتصادی اندازهٔ جست‌وجو قبل از خرید، رویکرد رضایتمندی را برگزید. شوارتز و همکارانش (۲۰۰۲) معتقدند مسئلهٔ اصلی در انتخاب جواب مطلوب، میزان رضایت کسب‌شده از آن است. این تئوری بر مبنای دسترسی‌نداشتن به اطلاعات کافی یا هزینهٔ زیاد یافتن جواب بهینه برای تصمیم‌گیری کارآمد است. رویکرد رضایتمندی[1] راهبرد تصمیم‌گیری با پیمایش در فضای حل مسئله است که جست‌وجو در مقادیر ممکن را تا جایی ادامه می‌دهد که به آستانه‌های قابل قبول[2] دست یابد. زمانی که دستیابی به راه‌حل بهینه هزینه‌بر باشد، رویکرد رضایتمندی در کنار تأمین جواب رضایت‌بخش بسیار سریع‌تر و راحت‌تر از روش‌های بهینه‌سازی عمل می‌کند. هر مسئلهٔ رضایتمندی به‌کمک تابع شاخص برای سنجش میزان رضایت کافی می‌تواند به‌عنوان مسئلهٔ بهینه‌سازی فرموله شود. اگر x نشان‌دهندهٔ فضای مسئله و S نشان‌دهندهٔ فضای رضایت‌بخش مسئله باشد، انتخاب راه‌حلی که بتواند محدودیت‌ها را ارضا کند، به‌صورت زیر است:

رابطهٔ ۱) $$\mathrm{Max} I_S(s)$$

رابطهٔ ۲) $$S \in X$$

رابطهٔ ۳) $$I_S(s) = \begin{cases} 1 & s \in S \\ 0 & s \notin S \end{cases}$$

به‌منظور شناخت فضای رضایت‌بخش مسئله، از تئوری استثنائات بهره می‌بریم. بر اساس چارچوب پیشنهادی، یافته‌های جمع‌آوری‌شده‌ای که مغایر با رفتار عادی سیستم یا مطابق با رفتار غیرعادی سیستم‌اند، یافته‌های استثنایی نامیده می‌شوند. بنابراین باید رفتارهای عادی و غیرعادی سیستم شناسایی شود. بدین منظور آستانه‌های قابل قبول رفتارهای عادی و غیرعادی استخراج می‌شود. پس از شناخت آستانه‌های قابل قبول برای سیستم، به جست‌وجو در فضای

1. Satisficing Solution
2. Acceptable Thresholds

Archive of SID

www.SID.ir

حل مسئله پرداخته می‌شود. این جست‌وجو از نقطهٔ ابتدایی آغاز می‌شود و تا زمانی که حالت‌های هدف محقق شود، در فضای مسئله ادامه می‌یابد. همزمان ظرفی برای جواب ایجاد می‌شود و زمانی که داده‌ای به حالت هدف رسید به ظرف جواب انتقال می‌یابد. پس از جست‌وجوی فضای جواب و رسیدن به معیار پایانی، ظرف جواب حاوی داده‌های غیرعادی است.

روش‌شناسی پژوهش

منطق یادگیری از استثنائات در حوزهٔ یادگیری ماشین، مسئلهٔ شایان توجهی است. در پژوهش حاضر چارچوب جدیدی بر اساس رویکرد تلفیقی تئوری استثنائات و تئوری رضایتمندی برای کشف پدیده‌های استثنایی و شناسایی الگوهای رفتاری پنهان آنها ارائه شده است. تئوری استثنائات با هدف کشف استثنائات بر پایهٔ راه‌حل‌های رضایت‌بخش به‌کارگرفته می‌شود و چرخهٔ شناسایی استثنائات، یادگیری رفتار آنها و به‌کارگیری قوانین شناسایی‌شده را تکمیل می‌کند. به‌منظور تشکیل فضای رضایت‌بخش از نظر خبرگان بهره برده شده است. بر اساس فضای رضایت‌بخش شناسایی‌شده، مسئلهٔ برنامه‌ریزی صفر و یک شکل می‌گیرد و با حل مسئله، داده‌های استثنایی شناسایی می‌شوند.

پس از شناسایی داده‌های استثنایی و تفکیک آن از داده‌های نرمال، باید دانش رفتار پدیده‌های عادی و استثنایی کشف شود. الگوریتم‌های یادگیری سنتی مانند درخت تصمیم بر پایهٔ جست‌وجوی حریص از رویکرد یادگیری بالا به پایین پیروی می‌کنند. این الگوریتم‌ها اغلب در شناسایی قوانین عمومی موجود در پایگاه داده عملکرد بهتری دارند. بدین ترتیب که یادگیری به‌کمک مؤثرترین متغیر و ایجاد قانون عمومی آغاز می‌شود و با درنظرگرفتن تک‌تک متغیرهای مهم با استفاده از ساختاری که به‌طور طبیعی در فضای فرضیه‌ها رخ می‌دهد، صورت می‌پذیرد. این قوانین، داده‌های متعلق به دسته‌های اصلی را پوشش می‌دهد و از داده‌های استثنایی صرف نظر می‌کند. بنابرین الگوریتم‌های یادگیری سنتی، عملکرد مناسبی در رفتار استثنائات ندارند. به‌منظور غلبه بر کاستی‌ها و مشکلات این الگوریتم‌ها در یادگیری از استثنائات، روش جدیدی مبتنی بر رویکرد یادگیری پایین به بالا با نام الگوریتم RISE ارتقایافته برای کشف دانش استثنائات پیشنهاد شده است. در این رویکرد، توجه و تمایل به ایجاد قوانین خاص با دقت زیاد است.

بر اساس الگوریتم RISE ارتقایافته، ابتدا داده‌ها به دو دستهٔ عادی و استثنایی طبقه‌بندی می‌شوند و سپس داده‌های عادی بر اساس میزان شباهتشان، به بیشترین تعداد خوشهٔ ممکن تقسیم می‌شوند. در نتیجه، داده‌ها بر اساس بیشترین شباهت در گروه‌های مشابه قرار می‌گیرند.

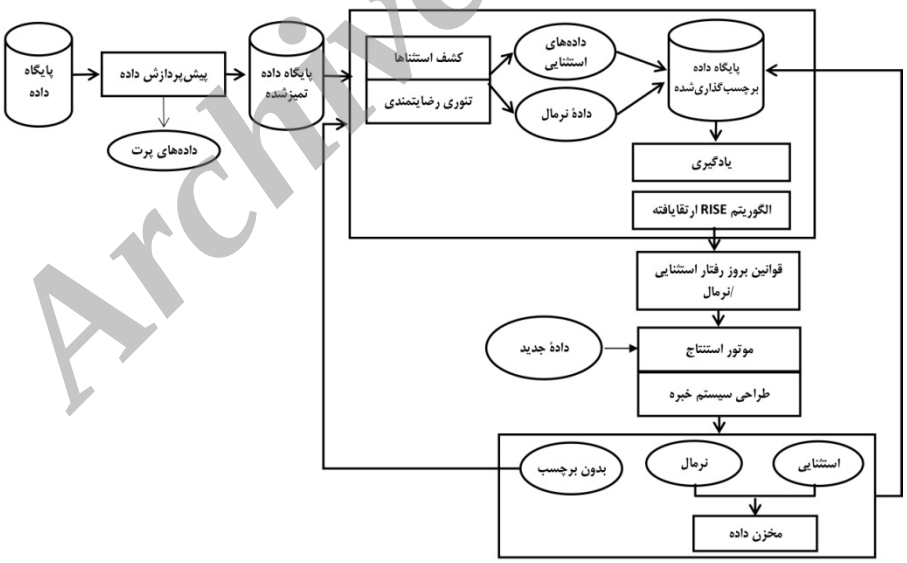
Archive of SID

www.SID.ir

فرایند خوشه‌بندی، دقت قوانین مستخرج را افزایش می‌دهد. فرایند شناسایی قوانین رفتاری داده‌های هر خوشه، با انتخاب تصادفی درصدی از داده‌ها آغاز می‌شود؛ به‌طوری که مطابق با هر نقطهٔ دادهٔ انتخابی، قانونی ساخته می‌شود. برای مثال اگر نقطهٔ دادهٔ $A = (a_1, a_2, \dots, a_n)$ به‌طور تصادفی از دستهٔ عادی انتخاب شده باشد، قانون R به‌صورت زیر ایجاد می‌شود:

رابطهٔ ۴) $R: IF(x_1 = a_1)\ and\ (x_2 = a_2) and\ \dots and\ (x_n = a_n)\ Then\ (x\ is\ normal)$

پس از ساخت قوانین اولیه، باید قوانین با توجه به نقاط همسایگی عمومی‌تر شوند. عمومی‌سازی به‌کمک تابع فاصلهٔ اقلیدسی صورت می‌گیرد؛ یعنی داده‌هایی که در همسایگی معینی وجود دارد، پوشش داده شوند و قوانین با حذف محدودیت روی متغیرها یا بازکردن بازه‌ها، عمومی‌تر شوند. مدل RISE ارتقایافته، مدل یادگیری خاص به عام برای یادگیری رفتار استثنائات است. از آنجا که استثنائات از ابتدای کار شناسایی شدند و در دستهٔ جداگانه‌ای قرار گرفتند و سپس قوانین رفتاری آنها استخراج شد، دقت یادگیری رفتار استثنائات افزایش می‌یابد. شکل ۲ فرایند کشف و یادگیری از رفتار استثنائات را نشان می‌دهد. به‌منظور به‌کارگیری قوانین در کشف استثنائات جدید، سیستم خبره‌ای طراحی شد که از قوانین ایجادشده در چارچوب تئوری استثنائات، برای شناسایی استثنائات جدید بهره می‌برد.



شکل ۲. چارچوب پیشنهادی کشف و یادگیری استثنائات

Archive of SID

www.SID.ir

در این قسمت به ارائهٔ چارچوب پیشنهادی برای شناسایی و یادگیری از داده‌های استثنایی با بهره‌مندی از تئوری‌های استثنائات و رضایتمندی پرداخته می‌شود که در ادامه به گام‌های آن اشاره شده است:

گام اول؛ شناسایی آستانه‌های قابل قبول برای بروز رفتار استثنایی: در این گام فضای رضایت‌بخش مسئله ترسیم می‌شود. این فضا به‌کمک اطلاعات استخراج‌شده از نظر خبرگان شناسایی می‌شود.

گام دوم؛ تشکیل مسئلهٔ صفر و یک.

گام سوم؛ حل مسئله: در گام سوم مجموعهٔ جواب‌های مسئلهٔ صفر و یک مشخص می‌شود. بدین منظور از ظرف جواب برای نگهداری جواب‌های رضایت‌بخش استفاده می‌شود. مجموعهٔ جواب رضایت‌بخش، پدیده‌های استثنایی شناسایی‌شده را شکل می‌دهد.

گام چهارم؛ دسته‌بندی داده‌ها: ابتدا داده‌ها بر اساس برچسب عادی و استثنایی دسته‌بندی می‌شوند و سپس با خوشه‌بندی دوبارهٔ داده‌های عادی، داده‌های مشابه در دستهٔ جداگانه‌ای قرار می‌گیرند.

گام پنجم؛ از داده‌های موجود در خوشه‌های ایجادشدهٔ گام چهارم، درصدی (α٪) به‌طور تصادفی انتخاب می‌شود و مطابق با هر داده در نمونهٔ تصادفی انتخاب‌شده، قانونی ساخته می‌شود.

گام ششم؛ داده‌هایی که نزدیک‌ترین فاصله را با قانون ایجادشده دارند و تاکنون قانون دیگری آنها را پوشش نداده است، به‌کار گرفته می‌شوند تا با حداکثر دقت و حداقل عمومی‌سازی، قانون را عمومی‌تر کنند. این کار توسط حذف شرایط یا بازکردن بازه‌ها برای متغیرهای عددی انجام می‌گیرد. به‌منظور سنجش کارایی فرایند عمومی‌سازی در هر مرحله از شاخص g-means و support استفاده می‌شود. در صورتی تغییرات لحاظ می‌شوند که شاخص g-means بدتر نشود و قانون ایجادشده درصد مشخصی از داده‌های موجود در آن دسته را پوشش دهد.

گام هفتم؛ پس از شناسایی قوانین بروز رفتار عادی و استثنایی، الگوهای رفتاری استخراج‌شده در چارچوب پیشنهادی تئوری استثنائات به‌منظور طراحی موتور استنتاجی برای کشف استثنائات جدید، به‌کار گرفته می‌شوند. استثنائات جدید را می‌توان بر اساس قوانین رفتاری عادی و استثنایی شناسایی کرد. بدین ترتیب که داده‌هایی که منطبق بر قوانین رفتار عادی نیستند یا مطابق قوانین رفتار استثنایی عمل می‌کنند، به‌عنوان پدیده‌های استثنایی شناسایی می‌شوند.

Archive of SID

www.SID.ir

جدول ۱. الگوریتم RISE ارتقایافته

گام عمومی سازی	گام ایجاد قوانین
اگر $R = (a_1, a_2, \dots, a_m, C_R)$ یک رول و $N = (e_1, e_2, \dots, e_m, C_n)$ نمونهٔ نزدیک باشد: function generalization (R,N) $a_i$is either true , $x_i = r_i$ or $r_{i,lower} \leq x_i \leq r_{i,upper}$ For attribute i-th IF a_i = Truethen do nothing else if $e_i > r_{i,upper}$then $e_i = r_{i,upper}$ else if $e_i < r_{i,lower}$ then $e_i = r_{i,lower}$	ES is the training set SS= select α % of ES randomly Let RS be SS For each rule R in RS N= the nearest neighborhoods E to R by $d < d^*$ let $\acute{R}$ = Generalization (R,N) Let $\acute{R}S = RS$ with R replaced with $\acute{R}$ if Acc($\acute{R}S$)>= $\beta\%\ Acc(RS)$ Then replace RS by $\acute{R}S$ if $\acute{R}$ is identical to another rule in RS then delete $\acute{R}$ from RS Until $Acc(RS) \geq \gamma$ Return RS

یافته‌های پژوهش

هدف اصلی پژوهش حاضر کشف و یادگیری رفتار استثنایی سیستم‌ها با به‌کارگیری تئوری استثنائات و تئوری رضایتمندی است. چارچوب پیشنهادی، چالش‌های پیش‌روی داده‌کاوی داده‌های استثنایی را برطرف‌کرده و الگوی رفتار غیرعادی سیستم را کشف می‌کند. حیطهٔ عملی پژوهش، رصد رفتار سهام به‌منظور شناسایی سهام استثنایی و تشکیل سبد سهام فوق مثبت در نظر گرفته شده است.

مطالعات پیشین در حوزهٔ شناسایی سهام استثنایی بر این باورند که سهام استثنایی، همان سهام با بازده زیاد است، اما در پژوهش حاضر به سهامی استثنایی گفته می‌شود که برای سرمایه‌گذار منفعت استثنایی به همراه داشته باشد. از مهم‌ترین مراحل اجرای پروژه، شناسایی و انتخاب متغیرهای دخیل در استثنایی‌شدن سهام است. متغیرهای انتخابی باید نسبت به استثنائات حساس باشند. به‌منظور اندازه‌گیری منفعت کسب‌شده از سرمایه‌گذاری، بر اساس مطالعات صورت‌گرفته و نظر خبرگان، متغیرهای تأثیرگذار در تشخیص سهام استثنایی استخراج شدند. با اجرای آزمون آماری t، از بین متغیرهای منتخب، متغیرهای سود هر سهم، سهام جایزه و افزایش سرمایهٔ شرکت برگزیده شدند. این متغیرها تفاوت معناداری را بین سهام عادی و استثنایی ایجاد می‌کنند. سرمایه‌گذاری استثنایی با رشد استثنایی ثروت برای سرمایه‌گذار همراه است. این رشد می‌تواند از طریق کسب سود استثنایی ایجاد شود. همچنین صدور سهام جایزه با حجم زیاد، سبب چند برابرشدن ثروت سهامدار می‌شود. اطلاعات مالی در این سه شاخص برای شرکت‌های فعال در بازار بورس ایران در ۱۳۲۷ رکورد گردآوری شده است.

Archive of SID

www.SID.ir

در گام بعد به‌دلیل اهمیت پاکسازی ورودی‌های مدل و تأثیر آلودگی داده‌ها بر کاهش عملکرد الگوریتم‌های داده‌کاوی، داده‌ها پیش‌پردازش می‌شوند. بدین ترتیب، اطلاعات سهام از طریق میانگین و انحراف معیار مجموعهٔ داده‌های گردآوری‌شده با به‌کارگیری مقدار Z، نرمالیز شدند. با توجه به رابطهٔ مستقیم متغیرهای منتخب با کسب سود، داده‌هایی که در فاصلهٔ بیشتر از ۳σ- میانگین قرار دارند، داده‌های پرت محسوب‌شده و از محاسبات خارج می‌شوند.

به‌منظور تشکیل پروفایلی از سهام یادشده و کشف استثنائات، باید از مشخصات سهام مجموعه‌ای ایجاد کرد که امکان نمایش داشته باشد. فرض کنید $x_{i,j}$ نشان‌دهندهٔ سهم iام در مشخصهٔ j باشد، در حالی‌که $i = \{1, 2, \dots, n\}$ معرف n سهم بررسی‌شده و $j = \{1, 2, \dots, m\}$ متغیرهای در نظر گرفته‌شده برای داده‌کاوی است. در این صورت M نشان‌دهندهٔ پروفایل سهام در دست بررسی است.

$$M = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{bmatrix}$$

فضای رضایت‌بخش مسئله با به‌کارگیری تئوری استثنائات ترسیم می‌شود. بر این اساس یافته‌های مشاهده‌شده‌ای که مغایر با رفتار عادی سیستم یا مطابق با رفتار غیرعادی سیستم‌اند «یافته‌های استثنایی» نامیده می‌شوند. بنابراین رفتار عادی و غیرعادی سهام باید شناخته شود. بدین منظور باید آستانه‌های قابل قبولی برای بروز رفتار غیرعادی تعیین کرد. آستانهٔ بروز رفتار استثنایی بر اساس عملکرد تک‌تک متغیرها مشخص می‌شود، به این ترتیب که حد قابل قبول برای بروز رفتار استثنایی، ۳σ+ از میانگین هر متغیر است. جدول ۲ اطلاعات سهام بررسی‌شده را نشان می‌دهد.

جدول ۲. مشخصهٔ سهام بررسی‌شده

مشخصه	سود هر سهم	تعداد سهام جایزه	افزایش سرمایهٔ شرکت
میانگین	۸/۵۲	۱۴/۱۶	۲۲/۳۷
انحراف معیار	۹۹/۹۸	۶۴/۳۳	۲۷۶/۴۷
حداقل	۰	۰	۰
حداکثر	۱۰۰۹۶	۹۹۰۰	۹۰۰
میانگین + ۳ × انحراف معیار	۳۵۱۵	۲۰۷	۸۵۲

Archive of SID

www.SID.ir

مسئلهٔ مطرح‌شده به پشتوانهٔ تئوری رضایتمندی و با بهره‌مندی از تابع شاخص برای سنجش میزان کافی رضایت‌بخشی، به‌صورت مسئلهٔ برنامه‌ریزی صفر و یک فرموله می‌شود. I_1، I_2 و I_3 متغیرهای صفر و یک برای ترسیم فضای رضایت‌بخش ایجادشده از آستانهٔ قابل قبول در مسئله‌اند. هدف، حداکثرشدن مجموع سه شاخص روی مجموعهٔ دادهٔ در دست بررسی است. اگر EPS نشان‌دهندهٔ سود هر سهم، BS و CA به‌ترتیب معرف تعداد سهام جایزه و میزان افزایش سرمایهٔ شرکت باشد، محدودیت‌های ۶ تا ۸ با توجه به آستانه‌های قابل قبول (جدول ۱) برای بروز رفتار استثنایی ایجاد می‌شوند. سهامی استثنایی است که دست‌کم یکی از این محدودیت‌های ۶ تا ۸ را ارضا کند (رابطه ۹).

رابطهٔ ۵) $Max\ I_1 + I_2 + I_3$

رابطهٔ ۶) $I_1(3515 - EPS) \leq 0$

رابطهٔ ۷) $I_2(207 - BS) \leq 0$

رابطهٔ ۸) $I_3(852 - CA) \leq 0$

رابطهٔ ۹) $I_1 + I_2 + I_3 \geq 1$

رابطهٔ ۱۰) I_1 و I_2 و $I_3 =$ 0 یا 1

راهبرد تصمیم‌گیری بر پایهٔ تئوری رضایتمندی با پیمایش در فضای حل مسئله، جست‌وجو در مقادیر ممکن را تا جایی ادامه می‌دهد که به آستانه‌های قابل قبول دست یابد. با جست‌وجو در فضای حل مسئله ۳۱ شرکت به‌عنوان سهام غیرعادی برگزیده شدند. جدول ۳ اطلاعات تعدادی از این سهام را نشان می‌دهد.

جدول ۳. سهام استثنایی کشف‌شده از طریق روش پیشنهادی

نام شرکت	سود سهم	سرمایهٔ شرکت	سهام جایزه	نام شرکت	سود سهم	افزایش سرمایه	سهام جایزه
پتروشیمی پردیس	۱۰۰۹۶	۰	۰	سالمین	۰	۰	۹۰۰
ارتباطات سیار ایران	۹۰۲۰	۰	۰	بیمهٔ دانا	۰	۰	۸۶۸
سیمان قائن	۸۰۰۰	۰	۰	قند اصفهان	۰	۰	۷۱۲
پالایش نفت تهران	۸۰۰۰	۰	۰	هنکل پال وش	۰	۰	۵۱۷
صنایع شیمیایی ایران	۸۰۰۰	۰	۰	بیمهٔ البرز	۰	۰	۵۰۰
پتروشیمی خارک	۷۵۰۰	۰	۰	داروسازی الحاوی	۰	۰	۵۰۰
پتروشیمی ایرانیان	۰	۹۹۰۰	۰	تجهیزات سنگین	۰	۰	۴۸۸

Archive of SID

www.SID.ir

پس از شناسایی سهام غیرعادی، گام بعد یادگیری رفتار استثنایی سهام و کسب توانایی برای پیش‌بینی سهام استثنایی است. به همین منظور از رویکرد یادگیری پایین به بالا بهره برده می‌شود. ابتدا داده‌های موجود در پایگاه داده بر اساس برچسب عادی یا غیرعادی دسته‌بندی می‌شوند و پس از آن، داده‌های متعلق به دستهٔ عادی با اجرای روش k-means خوشه‌بندی خواهند شد. این نوع خوشه‌بندی به‌دلیل راهبردی که برای جداسازی داده‌ها و امکان کنترل تعداد خوشه‌ها دارد، انتخاب شده است و تلاش می‌کند فضای مسئله را به‌گونه‌ای دسته‌بندی کند که داده‌های هر دسته ضمن داشتن کمترین فاصله و بیشترین شباهت به یکدیگر با دسته‌های دیگر بیشترین فاصله را داشته باشند. بنابراین سهام در چهار دسته شامل یک خوشه از داده‌های غیرعادی و سه خوشه از داده‌های عادی قرار گرفتند. خوشه‌بندی، داده‌های به‌نسبت مشابه را در یک گروه قرار می‌دهد و سبب افزایش دقت قوانین استخراج‌شده از آن خوشه می‌شود. اطلاعات خوشه‌ها در جدول ۴ خلاصه شده است.

جدول ٤. اطلاعات خوشه‌ها

خوشه	دسته	تعداد	میانگین			انحراف معیار		
			سود سهم	سهام جایزه	افزایش سرمایه	سود سهم	سهام جایزه	افزایش سرمایه
۱	غیرعادی	۳۱	۳۴۴۲	۲۱۹/۴	۲۹۱	۳۴۷۳/۲	۲۵۷/۹۷	۱۶۹۷/۸۳
۲	عادی	۱۱۰۵	۲۶۰	۲/۹	۱۶/۲	۲۸۱	۱۱	۵۸,۴
۳	عادی	۱۳۸	۲۰۷۵/۶	۰/۹۱	۵/۱۴	۷۸۲/۷	۸/۷۷	۳۱
۴	عادی	۵۳	۰	۱۴۱/۵۰	۲۴/۳	۰	۵۲/۵	۶۸/۴۳

از سهام موجود در هر خوشه درصدی (α ٪) به‌طور تصادفی انتخاب‌شده و به ازای هر نقطه داده قانونی منطبق بر آن ساخته شده است. شایان ذکر است، بسته به حجم داده‌های هر خوشه، مقادیر متفاوتی را می‌توان برای α در خوشه‌های مختلف لحاظ کرد. جدول ۵ تأثیر تغییر در مقدار α را بر دقت پیش‌بینی و تعداد قانون استخراج‌شده نشان می‌دهد. بر اساس بررسی‌های صورت‌گرفته، بهترین مقدار برای α در خوشه‌های ۱، ۲، ۳ و ۴ به‌ترتیب برابر با ۱۵ درصد، ۲ درصد، ۲/۵ درصد و ۵ درصد است. داده‌های خوشهٔ ۱ متعلق به سهام استثنایی است. از آنجا که سهام استثنایی حجم کمی از داده‌ها (۲/۳ درصد از حجم داده) را به خود اختصاص داده است، به قانون‌های اولیه بیشتری (۱۵ درصد معادل ۵ قانون) برای پوشش تمام سهام نیاز داریم. در این

Archive of SID

www.SID.ir

حالت، دقت پیش‌بینی به حداکثر رسیده است و مسئله از لحاظ پیچیدگی و زمان حل به کمترین میزان می‌رسد.

جدول ۵. تعیین تعداد بهینهٔ قوانین مستخرج

خوشه					
۱	٪۱۰	٪۵	٪۱۰	٪۱۲	٪۱۵
۲	٪۱۰	٪۷	٪۵	٪۲	٪۲
۳	٪۱۰	٪۷	٪۵	٪۵	٪۲,۵
۴	٪۱۰	٪۵	٪۵	٪۷	٪۷
دقت پیش‌بینی	٪۷۵	٪۷۸	٪۸۸	٪۹۶	٪۱۰۰
تعداد قانون استخراجی	۱۳۴	۹۰	۶۶	۳۴	۳۱

در مرحلهٔ بعد باید قوانین بر اساس داده‌های همسایه عمومی‌تر شوند؛ یعنی قوانین از تعداد محدودی از داده‌ها به تمام داده‌های موجود در پایگاه داده تسری یابد. به‌کمک تابع فاصلهٔ اقلیدسی، نقاط داده در همسایگی قوانین اولیه تعیین می‌شود. سهامی که نزدیک‌ترین فاصله را تا قانون ایجادشده دارند و تا به حال هیچ‌یک از قوانین آنها را پوشش نداده‌اند، به‌کار گرفته می‌شوند. بدین ترتیب با حداکثر دقت و حداقل عمومی‌سازی در قوانین اولیه، قوانین جدید ایجاد می‌شود. این روش از طریق حذف شرایط یا بازکردن بازه‌ها برای متغیرهای عددی ضمن حفظ میزان دقت از پیش تعیین‌شده برای قانون جدید، اجرا می‌شود. چنانچه اندازهٔ فاصلهٔ انتخابی کم باشد، تعداد تکرارها و زمان عمومی‌سازی افزایش می‌یابد و در صورتی که فاصله زیاد باشد، از دقت قوانین کاسته می‌شود. با در نظر گرفتن $k = ۰/۳$، پنج قانون برای تشخیص داده‌های استثنایی استخراج می‌شود که با به‌کارگیری قوانین همپوشانی، این پنج قانون در سه قانون زیر خلاصه خواهد شد:

- اگر سود هر سهم بیشتر از ۵۰۰۰ باشد، سهم غیرعادی است؛
- اگر تعداد سهام جایزه بیشتر از ۳۰۰ باشد، سهم غیرعادی است؛
- اگر افزایش سرمایهٔ شرکت بیشتر از ۹۵۰۰ باشد، سهم غیرعادی است.

با توجه به قوانین مستخرج از داده‌های استثنایی، درمی‌یابیم افزایش سرمایهٔ شرکت نسبت به متغیر سود هر سهم و تعداد سهام جایزه، اهمیت کمتری در بروز رفتار استثنایی سهام دارد. پس از استخراج قوانین حاکم بر رفتار عادی و استثنایی سهام، سیستم خبره‌ای طراحی شد که با ورود دادهٔ متعلق به سهام جدید، توانایی شناسایی سهام استثنایی را دارد. در مواجهه با دادهٔ جدید به دو صورت داده‌های استثنایی تشخیص داده می‌شود: ۱. سهامی که رفتارشان با هیچ‌یک از الگوهای

Archive of SID

www.SID.ir

عادی همخوانی ندارد و ۲. سهامی که رفتارشان با الگوی استثنایی مطابقت دارد. پس از تشخیص برچسب دادهٔ جدید، این داده بر پایگاه دادهٔ اصلی افزوده می‌شود و در آینده برای شناسایی داده‌های استثنایی جدید به‌کار می‌رود.

به‌منظور سنجش صحت عملکرد مدل پیشنهادشده برای کشف و یادگیری سهام استثنایی، آزمونی از این مدل روی داده‌های سه ماههٔ آخر سال ۱۳۹۳ به اجرا درآمد. همچنین از خبرگان درخواست شد سهام استثنایی این دوره را شناسایی کنند. نتایج مقایسهٔ خروجی مدل حاضر و نظر خبرگان بازار بورس، نشان داد دانش استخراج‌شده از سهام استثنایی با نظر خبرگان مطابقت دارد.

جدول ٦. سهام استثنایی شناسایی‌شدهٔ روش پیشنهادی و نظر خبرگان

ردیف	نام سهم	سود هر سهم	درصد افزایش سرمایهٔ جایزه
۱	شرکت پالایش نفت تهران	۰	۶۷۲/۲
۲	صنایع تولیدی اشتاد ایران (سهامی عام)	۰	۴۰۱/۴
۳	نئوپان ۲۲ بهمن	۰	۳۰۰
۴	گلوکوزان	۵/۹۷۰	۰

نتیجه‌گیری و پیشنهادها

کشف و پیش‌بینی رفتارهای استثنایی سیستم‌ها، مسئله‌ای چالش‌انگیز است. هدف از پژوهش حاضر، طراحی چارچوبی کارا برای مدیریت رخدادهای غیرعادی، شامل کشف داده‌ها و حالت‌های استثنایی از بین انبوه داده‌های موجود در پایگاه داده، استخراج دانش و یادگیری از فرایندهای استثنایی با به‌کارگیری تئوری‌های استثنائات و رضایتمندی است. از روش پیشنهادی به‌منظور رصد رفتار سهام مختلف و شناسایی سهام استثنایی بهره برده شده است. بدین ترتیب که با به‌کارگیری تئوری استثنائات حدود قابل قبول برای بروز رفتار استثنایی و رفتار عادی سهام شناسایی شد و تئوری رضایتمندی به‌منظور کشف سهام استثنایی به‌کار رفت. پس از تفکیک داده‌های غیرعادی از عادی، دانش پنهان سهام غیرعادی و عادی با اجرای رویکرد یادگیری پایین به بالا بر مبنای الگوریتم RISE ارتقایافته استخراج شد. بر اساس روش پیشنهادی به بررسی عملکرد ۱۳۲۷ سهم پرداخته شد که فقط ۲/۳ درصد از آنها رفتار فوق تصور خوبی از خود بروز دادند. از مجموع ۳۱ سهم غیرعادی کشف‌شده، پنج قانون به‌دست آمد و این پنج قانون در سه قانون خلاصه شد و به‌منظور پیش‌بینی رفتار سهام جدید به‌کار رفت.

از قوت‌های پژوهش حاضر، می‌توان به دیدگاه حاکم بر تحقیق در تعریف سهام استثنایی و به‌کارگیری ابزار کارآمد برای کشف آنها اشاره کرد. هدف ما حداکثرسازی ثروت سهامداران با

Archive of SID

www.SID.ir

شناسایی سهام استثنایی بر مبنای بروز رفتار فراتر از انتظار مثبت و تشکیل پرتفوی استثنایی بود؛ در حالی‌که تحقیقات مشابه تنها کسب بازده بالا را معیاری برای استثنایی‌شدن سهام دانسته و برای شناسایی و تشخیص آن به اجرای آزمون‌های آماری اقدام کردند. همچنین روش‌های ارائه‌شده در تحقیقات پیشین فقط برای شناسایی موارد و حالت‌های غیرعادی اعم از داده‌های غیرعادی مثبت (استثنائات) و غیرعادی منفی (داده‌های پرت) به‌کار رفته است؛ درحالی‌که روش پیشنهادی فقط به‌منظور تشخیص و شناسایی داده‌های استثنایی (داده‌های غیرعادی مثبت) ارائه شده است.

در پژوهش‌های بعدی می‌توان مجموعهٔ متغیرهای بیشتری که بروز رفتار استثنایی سهام را بیشتر توجیه می‌کنند، مانند سود تقسیمی، حق تقدم، درجهٔ نقدشوندگی، نوسان‌های قیمت و... در نظر گرفت. همچنین پیشنهاد می‌شود به‌منظور افزایش دقت تشخیص آستانهٔ رفتار استثنایی سهام از ریاضیات فازی استفاده شود.

References

Alavije, M.R., Askari, SH. & Paraste, S. (2015). Intelligent Online Store: User Behavior Analysis based Recommender System. *Journal of Information Technology Management*, 7(2): 385-406. (*in Persian*)

Albanis, G. & Batchelor, R. (2007). Combining heterogeneous classifiers for stock selection. *Intelligent Systems in Accounting, Finance and Management*, 15 (1-2): 1-27.

Barber, B.M. & Lyon, J.D. (1997). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics*, 43(3): 341-372.

Boshes, J. (2009). *Change point detection in cyber attack data*. PHD theses, Arizona state university.

Burez, J. & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3): 4626-4636.

Califf, M.E. & Mooney, R.J. (2003). Bottom-Up Relational Learning of Pattern Matching Rules for Information Extraction. *Journal of Machine Learning Research*, 4:177-210.

Cao, L., Zhao, Y. & Zhang, C. (2008). Mining Impact-Targeted Activity Patternsin Imbalanced Data. *IEEE Transactions on knowledge and data engineering*, 20(8): 1053 – 1066.

Archive of SID

www.SID.ir

Chang, T. S. (2011). A comparative study of artificial neural networks, and decision trees for digital game content stocks price prediction. *Expert Systems with Applications*, 38 (12): 14846–14851.

Chawla, N.V., Japkowicz, N. & lcz, A.K. (2004). Editorial: Special Issue on Learning from Imbalanced Data Sets. *Sigkdd Explorations*, 6(1):1–6.

Chen, M. C., Chen, L. S., Hsu, C.C. & Zeng, W.R. (2008). An information granulation based data mining approach for classifying imbalanced data. *Information Sciences*, 178(16): 3214-3227.

Chen, S. (2011). *Adaptive Learning from data flow and imbalanced data*. PhD Thesis, Stevens institute of technology, Hoboken.

Clark, E. (2014). Exploiting stochastic dominance to generate abnormal stock returns. *Journal of Financial Markets*, 20(1): 20-38.

Conlisk, J. (1996). Why bounded rationality? *Journal of Economic Literature*, 34(2): 669-694.

Fellner, G., Guth, W. & Martin, E. (2006). Satisficing or Optimizing? An Experimental Study, Max-Planck-Institute for Ökonomik. *Papers on Strategic Interaction*, www.econbiz.de/Record/satisficing-or-optimizing-an-experimental-study-fellner-gerlinde/10004870567.

Galván, J.R., Elices, A., Muñoz, A., Czernichow, T. & Sanz-Bobi, M.A. (1998). System for detection of abnormalities and fraud in customer consumption. *12th Conference on the Electric Power Supply Industry*, Thailand.

García, V., Sánchez, J.S. & Mollineda, R.A. (2012). On the effectiveness of preprocessing methods when dealing with different levels of class imbalance. *Knowledge-Based Systems*, 25(1): 13-21.

Gigerenzer, G. (2010). Moral Satisficing: Rethinking Moral Behavior as Bounded Rationality. *Topics in Cognitive Science*, 2(3): 528-554.

Gong, R.S. (2010). *A Segmentation and Re-balancing Approach for Classification of Imbalanced Data*. PHD theses, University of Cincinnati.

Hu, D. H., Zhang, X. X., Yin, J., Zheng, V. W. & Yang Q. (2009). Abnormal Activity Recognition Based on HDP-HMM Models. *the Twenty-First International Joint Conference on Artificial Intelligence*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.

Huang, K. Y. & Jane, C. J. (2009). A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories. *Expert Systems with Applications*, 36 (3): 5387-5392.

Archive of SID

www.SID.ir

Joshi, M.V. (2002). *Learning Classifier Models for Predicting Rare Phenomena*. PhD thesis, University of Minnesota, USA.

Kershaw, D. (2010). *Anomaly-based network intrusion detection using outlier subspace analysis approach*. PhD Thesis, Dalhousie University Halifax.

Kim, Y. & Sohn, S.Y. (2012). Stock fraud detection using peer group analysis. *Expert Systems with Applications*, 39(10): 8986–8992.

Kou, Y. (2006). *Abnormal Pattern Recognition in Spatial Data*. PHD theses, Faculty of Virginia Polytechnic Institute and State University.

Malakhov, S. (2014). Satisficing Decision Procedure and Optimal Consumption-Leisure Choice. *International Journal of Social Science Research*, 2(2): 138-151.

McCarthy, J. (1986). Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence*, 28(1): 89-116.

Mohaghar, A., Lucas, C., Hoseini, F. & Monshi, A. A. (2009). Use of business intelligence as a strategic information technology in banking: fraud discovery & detection. *Journal of Information Technology Management*, 1(1): 105-120. (*in Persian*)

Olson, D. & Mossman, C. (2003). Neural network forecasts of Canadian stock returns using accounting ratios. *International Journal of Forecasting*, 19(3): 453-465.

Reiter, R. (1987). A Theory of Diagnosis from First Principles. *Artificial Intelligence*, 32(1): 57-95.

Schwartz, B., Ward, A. Monterosso, J., Lyubomirsky, S., White, K. & Lehman, D. (2002). Maximizing Versus Satisficing: Happiness is a Matter of Choice. *Journal of Personality andSocial Psychology*, 83 (5): 1178-1197.

Simon, H.A. (1997). *Models of boundedrationality, Empirically grounded economic reason*. Vol. 3. Cambridge, Mass. MIT Press.

Slote, M. (1989). *Beyond Optimizing: A Study of Rational Choice, Cambridge, Mass*. Harvard University Press.

Taghavifard, M.T. & Jafari, Z. (2015). Fraud Detection Using a Fuzzy Expert System In Motor Insurance. *Journal of IT Management*, 7(2): 239-258. (*in Persian*)

Tezel, S. K. & Latecki, L. J. (2009). Improving SVM Classification on Imbalanced Data Sets in Distance Spaces. *Ninth IEEE International Conference on*. 6-9 Dec. DOI:10.1109/ICDM.2009.59.

Archive of SID

www.SID.ir

Vosough, M., Taghavifard, M.T. & Alborzi, M. (2015). Bank card fraud detection using artificial neural network. *Journal of Information Technology Management*, 6(4): 105-120. (*in Persian*)

Ward, D. (1992). The role of satisficing in foraging theory. Oikos, 63 (2): 312–317.

Weiss, G. (2004). Mining with rarity: A unifying framework. *SIGKDD Explorations Special Issue on Learning from Imbalanced Datasets,* 6(1):7–19.

Xiang, T. & Gong, S. (2008). Video Behavior Profiling for Anomaly Detection. IEEE Trans. *on Pattern Analysis and Machine Intelligence*, 30(5): 893–908.

Zhang, J. & Mani, I. (2003). *KNN approach to unbalanced data distributions*: *A case study involving information extraction.* ICML' 2003 Workshop on Learning from Imbalanced Datasets.

Zhu, P. & Hu, Q. (2013). Rule extraction from support vector machines based on consistent region covering reduction. *Knowledge-Based Systems*, 42(1): 1-8.

Archive of SID

www.SID.ir