مجله علمی پژوهشی «علوم و فناوری‌های پدافند غیرعامل»
سال چهارم، شماره۱، بهار ۱۳۹۲؛ ص ۶۳-۵۱

ارائه یک مدل داده‌کاوی جهت آشکارسازی ناهنجاری در پرتاب ماهواره

سینا دامی[۱]، حسین شیرازی[۲*]، سید مجتبی حسینی[۳]
۱- دانشجوی دکترا، ۲- دانشیار، ۳- استادیار دانشگاه صنعتی مالک اشتر
(دریافت: ۱۳۹۱/۰۵/۱۵،پذیرش: ۹۲/۰۳/۰۷)

چکیده

آشکارسازی ناهنجاری، یافتن الگوها در داده‌هایی است که از رفتار مورد انتظاری تبعیت نمی‌کنند. توسعه فناوری‌های آشکارسازی ناهنجاری و تشخیص خطا به‌صورت هوشمند، برای حامل پرتاب ماهواره به‌دلیل محیط سخت، دور و غیرقطعی، به‌عنوان یک مسئله کاملاً مهم و قابل توجه در صنعت هوافضا مطرح است. مدل پایش فعلی، با نظارت افراد خبره از طریق نمایش اطلاعات تله‌متری به‌کمک یک واسط گرافیکی انجام می‌شود. این رویکرد، علی‌رغم نیازمندی به تعداد زیادی افراد خبره، بسیار طاقت‌فرسا و زمانبر است. علاوه بر این، افراد همیشه قادر به تشخیص وضعیت‌های ناهنجار نیستند. در این مقاله به‌منظور پایش سلامت سیستم، یک چارچوب عیب‌شناسی نظام‌مند، شامل فرآیند داده‌کاوی توصیفی جهت آشکارسازی ناهنجاری ارائه شده است. نتایج حاصل از به‌کارگیری مدل‌های تشخیصی در چارچوب پیشنهادی، حاکی از این است که مدل‌های مذکور در ترکیب با مدل پایش فعلی، باعث بهبود امکانات عیب‌شناسی سنتی در تشخیص ناهنجاری می‌شوند. همچنین ضمن تسریع در فرآیند تصمیم‌گیری، می‌توانند ایمنی و قابلیت اعتماد را برای عملیات فضایی افزایش دهند.

کلیدواژه‌ها: داده کاوی،آشکارسازی ناهنجاری، پردازش تله‌متری، پایش سلامت ماهواره‌بر.

A Data Mining Model for Anomaly Detection of Satellite Launch Vehicle

S. Dami, H. Shirazi[*], S. M. Hoseini
Malek Ashtar University of Technology

(Received: 05/08/2012; Accepted: 28/05/2013)

Abstract

Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior. Development of advanced anomaly detection and failure diagnosis technologies for satellite launch vehicle (SLV) is a quite significant issue in the aerospace industry, because the space environment is harsh, distant and uncertain.Current SLV health monitoring and fault diagnosis practices involve around-the-clock limit-checking or simple trend analysis using text or graphical displays on large amount of telemetry data. This procedure, which requires large numbers of human experts, is of course cumbersome and time-consuming. Furthermore, humans are not always able to recognize anomalous situations. In this paper, a systematic and transparent diagnostic methodology will be proposed and developed within intelligent anomaly detectioon framework for SLV health monitoring. Experimental results show that the proposed method is capable of characterizing and monitoring interactions between multiple spacecraft parameters and can provide additional insight and valuable decision support for controllers and engineers.

Keywords: Data Mining, Anomaly Detection, Telemetry Processing, SLV Health Monitoring.

[*] Corresponding Author E-mail: shirazi@mut.ac.ir

۱. مقدمه

سنجش از دور[1]، علم جمع‌آوری داده از یک شیء، بدون تماس با آن و انجام تجزیه و تحلیل روی داده‌ها، برای حصول اطلاعات هدفمند است [۱]. ارتباطات مربوط به تله‌متری بین ایستگاه زمینی و حامل پرتاب ماهواره، برای انتقال داده‌هایی است که وضعیت سلامت زیرسیستم‌های حامل را مشخص می‌کند. در سیستم‌های زمینی پرتاب حامل، تله‌متری، رهگیری، ارسال فرمان‌های لازم و فاصله‌یابی اصلی‌ترین مرحله انجام کار در پایش سلامت حامل و تصحیح خطاهای ایجاد شده است که بر عهده زیرمجموعه TT&C[2] است [۲].

از زمان پرتاب حامل ماهواره تا تزریق در مدار، در هر لحظه باید وضعیت حامل کنترل شده و در صورت نیاز، فرمان‌های لازم صادر و ارسال شود. در این حین، حامل برای اعلام وضعیت زیرسیستم‌ها، اطلاعاتی را به‌عنوان داده‌های تله‌متری به ایستگاه زمینی ارسال می‌کند. بعد از پردازش تله‌متری، فرمان‌های لازم در قالب داده‌های دورفرمان[3] برای حامل ماهواره ارسال می‌شود. برای برقراری یک ارتباط کامل، می‌بایست برخی تمهیدات در نظر گرفته شود. نخست آنکه باید زاویه و ارتفاع حامل در هر لحظه از دید ایستگاه زمینی مشخص شود تا بتوان عملیات رهگیری را انجام داد. به‌طور کلی رهگیری به‌معنای پایش موقعیت حامل ماهواره در زمان‌های مشخص است که در این زمان‌ها، حامل توسط مرکز رهگیری خودکار قابل دسترسی است. در این روش، موقعیت حامل می‌تواند پایش و ذخیره شود و برای تعیین موقعیت بعدی، آنتن جهت رهگیری مورد استفاده قرار گیرد [۳]. بعد از دریافت داده‌های تله‌متری، اطلاعات مربوط به زیرسیستم‌های حامل ماهواره بر طبق پروتکل مورد توافق ایستگاه زمینی و حامل، استخراج می‌شود. پس از بررسی این اطلاعات، فرمان‌های لازم در قالب پروتکل دور فرمانبرای حامل، ارسال می‌شود. همچنین برای ایجاد هماهنگی بین بخش‌های مختلف ایستگاه زمینی و بررسی صحت عملکرد آنها، می‌بایست مدیریتی واحد بر آنها نظارت داشته باشد. هر چند نظارتی که خود نیازمند تعداد زیادی افراد خبره می‌باشد، بسیار طاقت‌فرسا و زمان‌بر است. در برخی مواقع، این کار قبل از اینکه داده‌ها پردازش شوند روزها زمان می‌برد، در حالی‌که در طی فازهای بحرانی مأموریت نظیر پرتاب، اطلاعات باید در ثانیه‌ها پردازش شده و تصمیم‌های لازم گرفته شود. علاوه بر این، انسان‌ها همیشه قادر به تشخیص وضعیت‌های غیرعادی نیستند، به‌ویژه زمانی که با روابط پیچیده‌ای در میان تعداد زیادی از متغیرها سر و کار داشته باشند [۴]. علاوه بر موارد ذکر شده، خطای کاربر نیز می‌تواند در این رویه نظارتی تأثیر منفی بگذارد. هدف این مقاله، ایجاد یک سیستم پایش خودکار جهت آشکار سازی ناهنجاری‌ها در حین بروز خطهای احتمالی، با توجه به الگوی داده‌های مأموریت‌های پیشین است.

در ادامه، ساختار مقاله به‌صورت زیر آورده شده است: پس از مرور تحقیقات گذشته در بخش دوم، به توضیح روش پیشنهادی شامل فرآیند داده کاوی توصیفی در بخش سوم پرداخته شده است. در بخش چهارم نتایج تجربی به‌همراه بحث بر روی نتایج آزمون مورد بررسی قرار گرفته و در نهایت، در بخش پنجم نتیجه‌گیری بیان شده است.

۲. پیشینه تحقیق

به‌طور کلاسیک، دو رویکرد عمده می‌تواند توصیف شود [۴]: روش‌های مبتنی بر مدل و روش‌های مبتنی بر داده (مستقل از مدل). روش‌های مبتنی بر مدل، از مدل‌های سخت‌افزاری و فرآیندهای فیزیکی برای پیگیری وضعیت سیستم و آشکارسازی انحرافات از رفتار اسمی استفاده می‌کند. گاهی این مدل‌ها برای تولید، بسیار پرهزینه خواهند بود زیرا تا حد زیادی به دانش خبره وابسته‌اند. علاوه بر این، ممکن است زمانی که برای سیستم‌های خیلی پیچیده نظیر پرتابگر فضاپیما به کار می‌روند، نتوانند موفق به تولید مجدد همه مدهای غیراسمی ممکن برای مدل‌های دقیقی که با فقدان زمان مواجه‌اند شوند.

از سوی دیگر، رویکردهای مبتنی بر داده، با استفاده از تکنیک‌های داده کاوی و یادگیری ماشین، بر اساس یک سیستم فیزیکی نمی‌باشند بلکه بیشتر با مدل‌هایی سروکار دارند که از داده‌های تله‌متری (مانند داده‌های سنسور دما) استنتاج می‌شوند. فعالیت‌های زیادی در این زمینه در چارچوب برنامه مدیریت یکپارچه سلامت حامل[4] (IVHM) از مرکز تحقیقات ایمز[5] ناسا برای نسل دوم حامل پرتابگر قابل استفاده مجدد[6] (RLV)، خدمه، و حامل‌های انتقال محموله انجام گرفته است.

برای روش‌های مبتنی بر مدل که به‌طور عمده با سیستم‌های خبره سنتی سروکار دارند، محدوده زمانی تحقیقات به سال‌های قبل از ۲۰۰۴ برمی‌گردد. درمقاله حاضر ابتدا مروری بر سیستم‌های به‌کار گرفته شده از این روش‌ها شده، سپس گذر مختصری بر روش‌های مبتنی بر داده، که به‌طور عمده از سال ۲۰۰۴ به بعد مورد استفاده قرار گرفته‌اند، خواهد شد.

۲-۱. روش‌های مبتنی بر مدل

یکی از پیشروان توسعه و کاربرد سیستم‌های خبره، سازمان‌های فضایی هستند که برای مشاوره و نیز بررسی شرایط پیچیده و صرفه‌جویی در زمان و هزینه چنین تحلیل‌هایی به این سیستم‌ها روی آورده‌اند. مرکز پرواز فضایی مارشال (MFSC)، یکی از مراکز وابسته به سازمان فضایی ناسا از سال ۱۹۹۴ است که در زمینه توسعه نرم‌افزارهای هوشمند کار می‌کند و هدف آن تخمین

[1] Remote Sensing
[2] Telemetry, Tracking & Control
[3] Telecommand
[4] Integrated Vehicle Health Management
[5] Ames Research Center
[6] Reusable Launch Vehicle

کم و کیف تجهیزات و لوازم مورد نیاز برای حمل به فضا است. این برنامه‌های کامپیوتری با پیشنهاد راهکارهایی در این زمینه، از بار کاری کارمندان بخش‌هایی چون ISS کاسته و به گونه‌ای طراحی شده‌اند که مدیریت‌پذیرند و بسته به شرایط مختلف، قابل تعریف هستند. مرکز فضایی MSFC، توسط فناوری ویژه خود موسوم به G2، به ایجاد برنامه‌های ویژه کنترل هوشمندانه و سیستم‌های پایش خطایاب می‌پردازد. سیستم خبره G2، جهت تحلیل داده‌های تله‌متری منتشر شده از فضاپیما و تعیین وضعیت آن به کار می‌رود. قابلیت اصلی G2، مدل کردن اشیای دنیای واقعی و پشتیبانی از تحلیل هوشمندانه داده‌های پیچیده است. علاوه بر این، قابلیت نمایش داده‌ها به کاربر نیز در فرمت مناسبی فراهم شده است.

توسعه سیستم‌های خبره، جهت ساخت و تولید نسل جدیدی از سیستم‌های توانمند است که برای حل مسائلی به‌کار گرفته شده که در حال حاضر تنها توسط انسان قابل حل می‌باشند. اولین کاربرد در این حوزه، یک سیستم خبره ناظر وضعیت پرواز[1] برای هواپیمای X-29 FSW بود. اولین پروژه در برنامه‌های کاربردی سیستم‌های خبره در مرکز تحقیقات ایمز ناسا، توسعه یک سیستم خبره ناظر وضعیت پرواز بود. سیستم خبره ناظر وضعیت پرواز، وظیفه پردازش خطای پائین‌گذر تله‌متری و کلمات وضعیت را با استفاده از پردازنده‌های زمینی بر عهده دارد. کلمات خطا بر اساس یک مدل مبتنی بر قاعده[2] از سیستم مدیریت خطا جهت ارزیابی مستقل وضعیت سیستم کنترل پرواز پردازش می‌شوند. تمام خطاهایی که توسط این فرآیند آشکار می‌شوند به‌وسیله مجموعه‌ای از قواعد دیگر برای ارزیابی سطح بالای سلامت و وضعیت جسم پرنده تفسیر می‌شود [۵].

باتوجه به لزوم فناوری‌های جدید در پایش سیستم‌های پرواز و طراحی مفهومی یک سیستم خبره ناظر وضعیت پرواز بررسی شده است [۶]. از طرفی، یک ابزار اکتساب دانش برای سیستم خبره ناظر وضعیت پرواز مطرح شده است [۷]. یک سیستم هواپیمایی واقعی با موفقیت تمام، توسط این ابزار تدوین شده و استفاده بی‌درنگ آن نیز با این ابزار تسهیل شده است. توصیف یک نسخه آزمایشی از سیستم خبره ناظر وضعیت پرواز (EESFSM) که در مرکز تحقیقات ایمز ناسا توسعه یافته است، صورت گرفته است [۸]. توسعه طرح‌ریزی شده این سیستم خبره شامل ۳ فاز است.

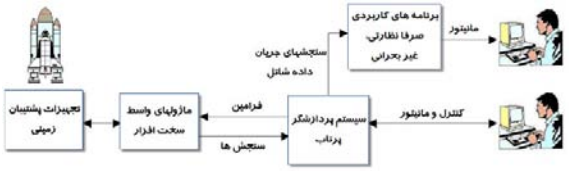
در یک گزارش [۹]، به‌توصیف محصولی تحت عنوان جعبه ابزار هوشمند مأموریت[3] (IMT) پرداخته شده است. IMT یک سیستم فرماندهی و کنترل است که بر روی یک سیستم خبره اعمال شده است. عمده توابع آن جهت ارسال فرمان‌هایی به فضاپیما و همچنین پردازش تله‌متری به‌کار می‌رود. در گزارش دیگری [۱۰]، که تحت مجوز شورای شبکه ملی (NGC) منتشر شده‌است، به توسعه دو سیستم خبره برای پردازش تله‌متری پرداخته شده است. هر دو مورد سیستم، جهت کاهش جریان اطلاعاتی به مهندس ناظر به‌کار گرفته شدند. ابتدا، سیستم خبره تحلیل خطا و کنترل هشدار (AHFA) به‌عنوان یک ابزار برخط برای مرکز کنترل پرواز طراحی شد که به تنهایی قادر به تشخیص خطاهای همزمان از سوئیچینگ علایم به‌صورت بی‌درنگ بود. در این فرآیند، تشخیص خطاها در قالب پیام‌های تله‌متری با تغییرات پله‌ای گزارش می‌شوند. بدین معنی که پنجره زمانی ثابتی برای دریافت همه پیام‌ها وجود ندارد و نیازی به یک فرآیند شبیه‌سازی نیز نمی‌باشد. دوم آنکه، سیستم خبره تحلیل خودکار رکورد خطا (AFRA) جهت حمایت از تحلیل برون‌خط رکوردهای خطای فرعی طراحی شد. AFRA ابتدا شکل‌های موج شبیه به هم برای شواهد و زمان خطا را بررسی کرده و سپس آن را دسته‌بندی می‌کند. گزارش دیگری [۱۱]، یک توسعه مقدماتی از روش پایش بر سیستم‌های پیشرانه و تعیین درستی روش‌های تخلیه اقلام بازیافتی را با استفاده از تکنیک‌های هوش مصنوعی ارائه می‌دهد که از آن در موتور F404 برای پشتیبانی از برنامه حامل F-18 استفاده شده است. اهداف عمده این پروژه سیستم خبره، افزایش ایمنی پرواز و کمک کردن به مهندسین پیشرانه در فرآیندهای زمانبر است .

در حال حاضر کار تحقیقات و برنامه‌های کاربردی در هوش مصنوعی توسط چندین گروه مختلف در مرکز فضایی جانسون (JSC) در ناسا در حال انجام است [۱۲]. سیستم خبره ONAV[4]، به‌عنوان یک مشاور بی‌درنگ در مرکز کنترل مأموریت JSC (MCC) جهت استفاده کنترلرهای هوایی ONAV توسعه یافته است. این سیستم خبره مبتنی بر دانش، جهت پایش سیستم ناوبری بر بورد شاتل فضایی، شناسایی خطاها و آگاه کردن پرسنل عملیات پرواز استفاده شده که وظیفه جمع‌آوری و آماده‌سازی داده‌ها، تجزیه و تحلیل خروجی فیلتر و کنترل پردازش را به‌صورت خودکار بر عهده دارد. این سیستم خبره در مأموریت‌های STS 51-F وSTS 51-I با استفاده از هدایت و کنترل تاکتیکی هوایی (TACAN) با جزئیات داده‌های واقعی، در زمانی نزدیک به زمان بی‌درنگ، با موفقیت آزمایش شده است. این کاربرد اولین سیستم مبتنی بر دانشی است که درتله‌متری و تراجکتوری داده از کامپیوتر عملیات مأموریت (MOC) استفاده شده است.

بخش سیستم‌های پردازش فضاپیما در مرکز فضایی کندی (KSC)، ابزار مبتنی بر عاملی را جهت نظارت جریان تله‌متری پردازش زمینی شاتل توسعه داده است. هدف از این کار، افزایش آگاهی‌های موقعیتی بوده است. برای این منظور بالغ بر چند صد قاعده جهت آگاه‌سازی مهندسین شاتل نوشته شده است. عامل‌های تله‌متری مبتنی بر قاعده جهت استفاده در پردازش زمینی شاتل فضایی بررسی شده است [۱۳-۱۵]. KSC مسئول بررسی زمینی

[1] Flight Status Monitor
[2] Rule-Based
[3] Intelligent Mission Toolkit
[4] Onboard Navigation

شاتل فضایی قبل از پرتاب است. در KSC، سیستم پردازش پرتاب[1] (LPS)، امکاناتی از قبیل فرمان، کنترل و پایش سیستم‌های حامل فضایی را از آغاز آزمایش شاتل تا پایان فازهای مختلف از جمله شمارش معکوس نهایی و پرتاب به‌عهده دارد. LPS به‌طور مداوم بر شاتل و تجهیزات پشتیبانی از زمین خود نظیر سخت‌افزاری که پیشران‌ها را بارگذاری می‌کند، نظارت دارد. زیرسیستم‌هایی مسئول برقراری ارتباط اطلاعات به‌صورت خودکار و مستقیم از سیستم‌های کامپیوتری شاتل می‌باشند. برخی از زیرسیستم‌ها نیز مسئول برقراری ارتباط با تجهیزات پشتیبانی زمینی از طریق ماژول‌های واسط سخت‌افزاری می‌باشند. هر ماژول می‌تواند باحدود ۲۴۰ سنسور به واحد کنترل وصل شود. به‌طور کلی، تقریباً ۵۰،۰۰۰ پارامتراز قبیل درجه حرارت، فشار، نرخ جریان، سطوح مایع، سرعت توربین، ولتاژها، جریان‌ها، موقعیت دریچه، وضعیت سوئیچ و بسیاری از پارامترهای دیگر باید نظارت و کنترل شوند. شکل (۱) عملکرد کنترل و پایش زمینی را در KSC نشان می‌دهد.

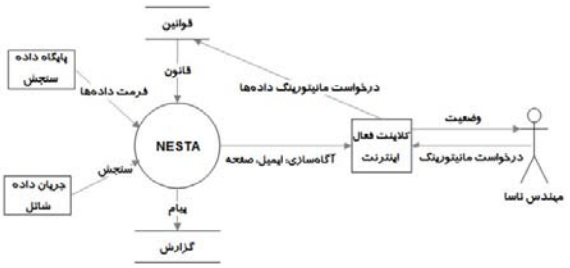


شکل ۱. عملیات پایش و کنترل زمینی در KSC [۱۳]

بالغ بر ۲۵ سال است که مهندسین ناسا از LPS به‌منظور بررسی آمادگی پرواز شاتل فضایی و کنترل شمارش معکوس پرتاب استفاده کرده‌اند. پس از فاجعه شاتل فضایی کلمبیا (۱ فوریه ۲۰۰۳)، انجمن بررسی حوادث کلمبیا پیشنهاداتی را به‌منظور بهبود امنیت از هر دو منظر سازمانی و فنی مطرح نمودند [۱۳].

در نیمه دوم سال ۲۰۰۳، ابزار نرم‌افزاری NESTA به‌عنوان یک عامل تله‌متری شاتل برای مهندسی ناسا پدیدار شد. لازم به ذکر است که داده‌های پردازش شده توسط LPS بر روی یک شبکه محلی توزیع می‌شود. داده‌های توزیع شده به‌عنوان جریان داده شاتل[2] (SDS) شناخته شده هستند که حاوی داده‌های پردازش زمینی بی‌درنگ حامل می‌باشند و تنها برای اعمال انجام پایش قابل استفاده هستند. هدف اصلی NESTA پایش مداوم و خودکار بر این جریان تله‌متری (SDS) است تا به‌محض آشکارسازی ناهنجاری، هشداری را نزدیک به زمان بی‌درنگ، به‌طور خودکار به مهندسین ناسا ارسال کند. ضوابط از پیش تعریف شده جهت پایش، شامل رویدادهای عملیاتی مورد انتظار (مثل به‌کار انداختن برق حامل) و رویدادهای غیرمنتظره (نظیر اختلاف‌های بزرگ بین مقادیر افزونه سنسور) یا خطاها می‌باشند [۱۵]. شکل (۲) نمودار مفهومی NESTA را نشان می‌دهد.

تا دسامبر ۲۰۰۵، NESTA به‌مدت دو سال در حال استفاده بوده که در طی آن، مهندسین ناسا صدها قاعده نوشته‌اند و NESTA تا به‌حال هزاران هشدار به کاربران متعدد در KSC و سایر سایت‌های دوردست ارسال کرده است.



شکل ۲. نمودار مفهومی NESTA

۲-۲. روش‌های مبتنی بر داده

به‌طور کلی، دو مسئله مهم در مواجهه با روش‌های سنتی آشکارسازی ناهنجاری در حوزه پرتاب حامل ماهواره وجود دارد:

۱. رفتار سیستم چگونه مدل و بازنمایی شود؟

۲. دانش لازم چگونه باید آماده (کسب) شود؟

به‌طور اساسی سیستم‌های خبره از افراد خبره به‌عنوان منبع دانش خود یاد می‌کنند. ولی روش‌های مبتنی بر داده، همان‌طور که پیش از این ذکر شد، با مدل‌هایی سروکار دارند که از داده‌های تله‌متری استنتاج می‌شوند. در واقع داده‌های تله‌متری به‌عنوان منبع دانشی برای این نوع از سیستم‌ها محسوب می‌شود.

توسعه اخیر در تکنیک‌های داده‌کاوی برای آشکارسازی ناهنجاری، آن را برای استفاده سودمند داده‌های بایگانی شده سیستم، جهت تولید برنامه‌های کاربردی پایش سلامت سیستم در سطح بالاتر و پیچیده‌تری ممکن کرده است. این برنامه‌های کاربردی مبتنی بر داده، قادر به توصیف و پایش تعاملات بین پارامترهای مختلف هستند و همچنین می‌توانند تصمیم‌گیری با ارزشی را نیز برای کنترلرهای مأموریت و مهندسین فراهم کنند. نرم‌افزارهای مبتنی بر داده به‌طور موفقیت‌آمیزی در عملیات فضایی شامل شاتل فضایی و ایستگاه فضایی بین‌المللی به‌کار گرفته شده‌اند. این ابزارها برای تحلیل مهندسی داده‌های فضاپیما جهت آشکارسازی حوادث غیر معمول در داده‌ها و پایش سلامت سیستم در اتاق کنترل پرواز به‌کار گرفته شده است. با افزودن ابزار پایش، نرم‌افزار سنتی کنترل مأموریت قادر به فراهم کردن کنترلرهایی با بینش وسیع‌تری نسبت به سلامت و عملکرد سیستم‌های فضایی با توجه به دید آنها خواهند بود. همچنین اضافه کردن قواعد اکتشافی بر اساس روش‌هایی که دانش به‌دست آمده از کنترلرهای مأموریت را کد گذاری می‌کنند، می‌تواند جهت کاهش منابع انسانی باتجربه مفید واقع شود [۱۶]. از نحوه به‌کارگیری چنین تکنیک‌هایی در عملیات کنترل مأموریت ناسا و

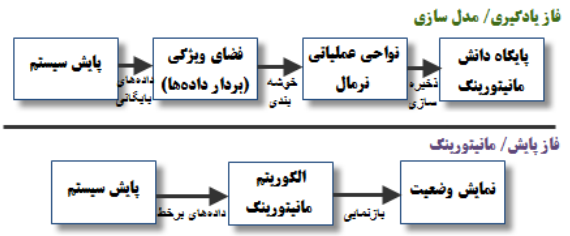
[1] Launch Processing System
[2] Shuttle Data Stream

طراحی سیستم‌های نرم‌افزاری پایش سلامت سیستم کنترل مأموریت آینده ارائه شده است [۱۷].

چندین ابزار نرم‌افزاری مبتنی بر داده، با موفقیت جهت عملیات مأموریت برای شاتل فضایی و ایستگاه فضایی بین‌المللی به‌کار گرفته شدند. یک الگوریتم آشکارساز بدون ناظر به نام Orca بر اساس رویکرد نزدیکترین همسایه معرفی شد که برای آزمایش داده‌های موتور اصلی شاتل فضایی و موتور راکت به‌کار گرفته شده است [۱۸]. Orca [۱۹] یک ابزار داده کاوی است که نقاط داده‌های ناهنجار[1] یا برون‌هشته‌ها[2] را در مجموعه داده‌های چند متغیره با محاسبه فاصله هر نقطه داده از نقاط همسایه‌اش کاوش می‌کند. برون‌هشته‌ها نقاطی هستند که رویداد آنها در یک مدل داده کم است. این نقاط می‌توانند در نتیجه اثر داده‌های نویزی، مضمون‌های بدخواهانه یا جمع‌آوری ناقص به‌وجود آیند. از منظر آماری یک برون‌هشته [۲۰]، داده مشاهده شده‌ای است که فاصله دورتر از سایر داده‌ها داشته باشد. حضور برون‌هشته‌ها در داده‌های سیستم فضاپیما برای کنترلرهای مأموریت دارای اهمیت هستند زیرا امکان دارد منجر به شناسایی اجزای سیستمی شود که عملکرد مطلوبی ندارند.

یک سیستم پایش استقرایی[3] (IMS) جهت آشکارسازی رفتارهای غیراسمی پیشنهاد شده است [۲۱]. داده‌های پرواز مربوط به مأموریت‌های گذشته، به‌عنوان داده‌های آموزشی برای یک الگوریتم خوشه‌بندی (یعنی، K-means و خوشه‌بندی مبتنی بر تراکم) مورد استفاده قرار گرفته‌اند که نواحی رفتارهای اسمی (خوشه‌ها) را در فضای داده n بعدی شناسایی می‌کنند، به‌طوری که n، تعداد ورودی‌های سنسور است. خوشه‌ها که پایگاه دانش IMS را بازنمایی می‌کنند قابلیت استفاده برای آشکارسازی بی‌درنگ رفتارهای غیرعادی در یک پرواز جدید را دارا می‌باشند. زمانی که یک بردار سنجش دریافت می‌شود، پایگاه دانش خوشه‌ای را که بردار بدان متعلق است، برمی‌گرداند. اگر بردار مذکور نتواند به یک خوشه خاص تعلق گیرد، فاصله با نزدیکترین خوشه به‌معنای یک انحراف از رفتار اسمی سیستم خواهد بود. این توصیف یا مدل از عملیات نرمال، در یک پایگاه دانش ذخیره می‌شود تا بتواند برای تحلیل یا پایش برخط حوادث ضبط شده مورد استفاده قرار گیرد. شکل (۳) خلاصه‌ای از روش IMS را نشان می‌دهد.

داده‌های دریافتی از سنسورهای موجود برای حامل AresI-X در KSC با استفاده از سه الگوریتم مبتنی بر قاعده، مبتنی بر مدل، و مبتنی بر داده تحلیل شده‌اند [۲۲]. از آنجایی که Ares I-X یک حامل جدید محسوب می‌شد، بنابراین فاقد داده‌های پرواز و داده‌های آزمایش بود. برای این منظور از داده‌های قبلی شاتل فضایی برای آموزش IMS و آزمایش هر سه روش استفاده شد. چراکه مرحله اولیه Ares I-X بسیار شبیه به بوسترهای راکت جامد (SRB) روی شاتل فضایی است.



شکل ۳. خلاصه‌ای از روش IMS. IMS رفتار نرمال را از داده‌های بایگانی یا شبیه‌سازی شده سیستم یاد می‌گیرد و پس از ایجاد یک مدل از عملیات نرمال آن را جهت پایش برخط در پایگاه دانش ذخیره می‌کند

الگوریتم خوشه‌بندی K-means [۲۳] نیز بر روی فضای ویژگی‌های استخراج شده از سری‌های زمانی[4] جمع‌آوری شده از مأموریت‌های گذشته به کار گرفته شد که در آن تلاش برای یافتن روابط خاص بین رخداد خطا و روند پارامترهای استنتاج شده از قواعد وابسته به داده‌ها انجام گرفت.

یک روش جهت تشخیص ناهنجاری بر اساس کمترین مربعات ماشین بردار پشتیبان[5] (LS-SVM) ارائه شده که از آن به‌منظور افزایش سلامت و کاهش زمان از کار افتادگی فضاپیما در مدار استفاده شده است [۲۴]. این فرآیند شامل مراحل زیر است: ۱- جمع‌آوری و پیش‌پردازش داده‌ها نظیر بار، ولتاژ، دما، ارتعاش و غیره، ۲- استخراج ویژگی[6]، ویژگی‌ها با استفاده از روش‌های آماری از سیگنال‌های سنسورها توصیف می‌شوند، ۳- انتخاب ویژگی[7]، ازتحلیل مؤلفه اصلی[8](PCA) برای انتخاب زیرمجموعه‌ای از ویژگی‌ها که حاوی اطلاعات مفیدترند، استفاده شده است و ۴- تشخیص ناهنجاری، از LS-SVM جهت شناسایی رفتارهای غیرهنجار فضاپیما در مدار استفاده شده است. همچنین زمان ملاحظه و نیز عامل ناهنجاری از سطح زیرسیستم تا سیستم مشخص شده است.

توصیف چهار الگوریتم Orca، GritBot، IMS و OCSVM[9] و همچنین تعداد نه ناهنجاری آشکار شده توسط چهار الگوریتم مورد بررسی قرار گرفته است [۲۵]. سه الگوریتم IMS، Orca و OCSVM نیز در گزارش دیگری مورد ارزیابی قرار گرفته است[۲۶]. نتایج حاصل نشان داده که IMS بهترین عملکرد را از نقطه نظر پیچیدگی محاسباتی و دقت در مقایسه با دو روش دیگر دارد. هر چند اگر تنها فاکتور دقت در نظر گرفته شود OCSVM نیز با IMS قابل مقایسه خواهد بود.

[1] Anomaly
[2] Outlier
[3] Inductive Monitoring System
[4] Time Series
[5] Least Squares Support Vector Machine
[6] Feature Extraction
[7] Feature Selection
[8] Principal Component Analysis
[9] One Class Support Vector Machine

در پشتیبانی از فعالیت تحلیل مهندسی در اتاق ارزیابی مأموریت (MER) واقع در JSC، ابزاری برای داده‌های سیستم آشکارساز ضربه لبه انتهایی بال (WLEIDS) از شاتل فضایی جهت اثرات بالقوه ضربه به‌کار گرفته شده است. این ابزار همچنین در اتاق کنترل پرواز ISS نیز جهت برنامه‌های پایش سلامت برخط برای ژیروسکوپ لحظه‌ای کنترل ISS به‌کار گرفته شده است. سیستم WLEIDS در پاسخ به فقدان مدارگرد کلمبیا در مأموریت STS-107 توسعه یافته است. در حین پرتاب STS-107، بیرون ریختن قسمتی از حباب مخزن سوخت خارجی شاتل باعث ضربه‌ای به لبه انتهایی بال چپ مدارگرد شده که سبب به خطر افتادن سیستم حفاظت گرمایی شده است. این خرابی در طی مرحله دخول مجدد حامل درجو[1]، به علت جوش آمدن و شکست ساختار داخلی بال منجر به از دست دادن فجیع حامل و خدمه شد [۲۷].

تحلیل اولیه WLEIDS توسط مهندسین MER با بررسی بصری گراف‌های سه بعدی از خلاصه داده‌هایی که موقعیت شتاب‌سنج و شدت ارتعاش بدنه را در طول محور زمان نشان می‌دهد، صورت گرفته است. ابزارهای Orca و IMS به منظور پشتیبانی از تحلیل WLEIDS بر روی سه حامل شاتل به‌کار گرفته شده‌اند. هدف فراهم کردن یک پویش سریع از خلاصه فایل‌های WLEIDS جهت یافتن نقاط داده غیر معمول و کمک به تمرکز تحلیل‌گران MER است.

Orca به‌منظور جستجوی برون‌هشته‌های داخل داده‌های جمع‌آوری شده در حین پرتاب است. IMS قبل از پرتاب، به‌منظور تحلیل داده‌های نرمال از پرتاب‌های پیشین جهت مشخص کردن الگوهای ارتعاش برای هر گروه از شتاب‌سنج‌ها مورد استفاده قرار گرفته است. داده‌های پرتاب فعلی با این توصیف مقایسه می‌شوند و الگوهای ارتعاش غیر معمول را که توسط حوادث ضربه ایجاد شده‌اند، شناسایی می‌کنند. نتایج نشان می‌دهد که استفاده از این ابزار بسیار سودمند بوده و تمامی حوادث را به درستی شناسایی کرده است. استفاده از این ابزار در تمامی پرتاب‌ها، حتی قادر به شناسایی اثرات ارتعاش انرژی پائین‌تر نیز بوده است.

جدول ۱. لیست ویژگی‌ها

نماد	ویژگی	شماره
f_{max}	مقدار بیشینه	۱
f_{min}	مقدار کمینه	۲
f_m	میانگین	۳
f_{sd}	انحراف معیار	۴
f_{fft1}	طیف توان فوریه مرتبه اول	۵
f_{fft2}	طیف توان فوریه مرتبه دوم	۶
f_{rms}	خطای rms	۷
f_s	میزان چولگی	۸
f_k	میزان کشیدگی	۹

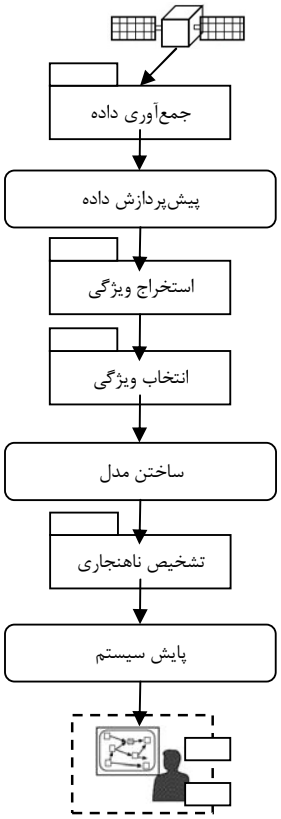
[1] Reentry Phase

۳. روش پیشنهادی شامل فرآینده‌داده‌کاوی توصیفی

فعالیت‌های داده کاوی به‌طور کلی در راستای دو هدف قرار می‌گیرند که شامل پیش‌بینی و توصیف می‌باشند [۲۸]. پیش‌بینی بر مبنای ویژگی‌های داده‌های موجود، سعی در پیش‌بینی خصیصه‌ها یا ویژگی‌های مورد نظر نمونه‌های جدید داده‌ای دارد و توصیف نیز در پی یافتن الگوهای جدید مستتر در داده‌های موجود است. در شکل (۴)، نمودار مدل پیشنهادی شامل فرایند داده‌کاوی توصیفی نشان داده شده است.

۳-۱. جمع‌آوری داده

این مرحله روی چگونگی تولید و جمع‌آوری داده‌ها متمرکز شده است. به‌طورکلی دو روش مجزا برای این کار وجود دارد. اولین روش زمانی است که فرایند تولید داده زیر نظر خبره مسئله یا طراح مسئله صورت می‌گیرد، این روش به آزمایش مخصوص[2] شهرت دارد. دومین دومین روش زمانی است که خبره مسئله نمی‌تواند تأثیری بر فرایند تولید داده بگذارد، این روش به روش مبتنی بر مشاهده[3] شهرت دارد [۲۹]. برای این منظور از دو دیتاست ایجاد شده از هر دو روش، به‌ترتیب برای آموزش با ناظر[4] و بدون ناظر[5] بهره گرفته شده است.



شکل ۴. چارچوب پیشنهادی شامل فرآیند داده‌کاوی برای پایش سلامت سیستم

[2] Designed Experiment
[3] Observational Approach
[4] Supervised Learning
[5] Unsupervised Learning

برای دیتاست ایجاد شده به روش آزمایش مخصوص، از داده‌های مربوط به شاتل فضایی از انبار داده یادگیری ماشین دانشگاه کالیفرنیا [۳۰] استفاده شده است (دیتاست ۱). این داده‌ها شامل دو مجموعه داده برچسب دار[1]، به‌ترتیب برای فازهای یادگیری و آزمایش است. در روش مبتنی بر مشاهده نیز، از داده‌های بالک الکترومکانیکی[2] (FLEA) به‌دست آمده از آزمایشگاه مجازی DASHlink ناسا [۳۱] استفاده شده است (دیتاست ۲). این داده‌های بدون برچسب[3] شامل دو دقیقه از سناریوهای یک مجموعه بزرگ از پروفایل‌های بار[4] و حرکت[5] است. داده‌های مذکور شامل داده‌هایی با سرعت کم[6] (بار، دما، موقعیت، و فشار) هستند، به این معنی که با سرعت نمونه برداری پایین (معمولاً ۱ KHz) ضبط می‌شوند. این سناریوها در نوامبر ۲۰۱۱ به‌منظور آزمایش الگوریتم‌های تشخیصی طراحی شدند.

۳-۲. پیش پردازش داده

در این مرحله برای مقابله با نمونه داده‌هایی که دارای مقادیر غیر معمول بوده و با بیشتر مشاهدات سازگار نیستند، از استراتژی توسعه روش‌های مدل‌سازی قدرتمند و حساس به نمونه‌های مجزا (که در ادامه ذکر خواهد شد)، در مقابل شناسایی و حذف آنها به‌عنوان بخشی از فاز پیش پردازش داده بهره گرفته شده است.

در دیتاست ۲ علاوه بر مرحله شناسایی این نمونه‌ها، جهت بهره‌برداری بهتر از داده‌ها، میزان تغییرات داده‌ها نیز محاسبه شده است. یعنی در ورودی مرحله بعد، به‌جای به‌کارگیری داده‌ها از میزان انحراف داده‌ها به‌عنوان مقادیر نرمالیزه شده استفاده شده است. جدول (۲) به‌ترتیب تغییرات میانگین و انحراف معیار مقادیر سنسور دمای موتور Y را در هر ۱۰ داده‌متوالی از دیتاست ۲ نشان می‌دهد.

روش‌های مبتنی بر استخراج ویژگی: این روش‌ها یک فضای چند بعدی را به یک فضای با ابعاد کمتر نگاشت می‌کنند. در واقع با ترکیب مقادیر ویژگی‌های موجود، تعداد کمتری ویژگی به‌وجود می‌آورند به‌طوری که این ویژگی‌ها دارای تمام (یا بخش اعظمی از) اطلاعات موجود در ویژگی‌های اولیه باشند.

پس از آماده سازی داده‌ها، ویژگی‌های لیست شده در جدول (۱) از هر ۱۰ نمونه متوالی یک سری زمانی (در دیتاست ۲) با استفاده از روش‌های آماری تعیین می‌شوند. به‌طور مسلم، محاسبه این کمیت‌های آماری سربار زیادی را بر روی سیستم تحمیل خواهد کرد. به‌منظور کاهش حجم محاسباتی و نیز سربار زمانی در پردازش برخط، از توابع ساده حوزه زمان نظیر بیشینه و میانگین استفاده می‌شود. این دو تابع نتایج بهتری را در مقایسه با توابع دیگر (مثل کمینه، میانه، انحراف معیار، انرژی، و غیره) به دنبال داشته است [۱۹]. برای این منظور، در این مقاله (تنها در دیتاست ۲) از تغییرات دو تابع میانگین و بیشینه در مرحله استخراج ویژگی استفاده شده است. از آنجایی که در دیتاست ۱، ترتیب زمانی نمونه‌ها با توجه به دسته‌بندی آنها صورت گرفته و نیز بخشی از نمونه‌ها به‌منظور ارزیابی از دیتاست حذف شده است، این ترتیب به‌صورت تصادفی در نظر گرفته می‌شود. بنابراین عملیات استخراج ویژگی بر روی این دیتاست اعمال نشده است.

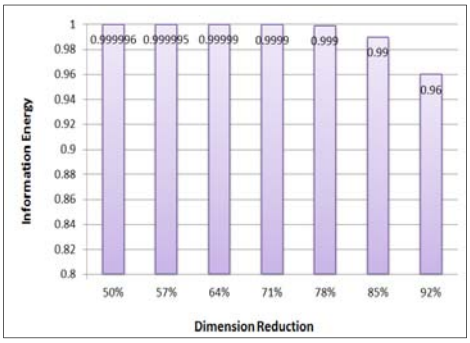
جدول ۲. تغییرات میانگین دمای موتور Y در هر ۱۰ داده‌ی متوالی از دیتاست ۲ با برش زمانی مشخص شده

زمان	میانگین دمای موتور Y	تغییرات میانگین دمای موتور Y
۱	۲۲/۴۵	
۲	۲۲/۶۶	۰/۲۱
۳	۲۱/۷۹	-۰/۸۷
۴	۲۲/۸۹	۱/۱۰
۵	۲۱/۷۹	-۱/۱۰
۶	۲۲/۳۳	۰/۵۴
۷	۲۲/۴۳	۰/۱۰
۸	۲۳/۰۱	۰/۵۸

روش‌های مبتنی بر انتخاب ویژگی: این روش‌ها سعی می‌کنند با انتخاب زیرمجموعه‌ای از ویژگی‌های اولیه، ابعاد داده‌ها را کاهش دهند.

PCA [۳۲]، بهترین روش برای کاهش ابعاد داده‌ها به‌صورت خطی است [۳۳]. یعنی با حذف ضرایب کم اهمیت به‌دست آمده از این روش اطلاعات از دست رفته نسبت به روش‌های دیگر کمتر است. از آنجایی‌که PCA در حین عمل انتخاب ویژگی توجهی به برچسب نمونه‌های آموزشی ندارد و همچنین تعداد ویژگی‌های دیتاست ۱ تنها شامل ۹ بعد است، بر این اساس در فاز پیش پردازش برروی دیتاست ۱ هیچ‌گونه عملیات استخراج و انتخاب ویژگی اعمال نشده است. ولی در دیتاست ۲ علاوه بر مرحله استخراج ویژگی که در بخش ۳-۲-۱ توضیح داده شد، به‌دلیل ابعاد بیشتر این نوع از دیتاست که فاقد برچسب است، از PCA جهت عملیات انتخاب ویژگی بهره گرفته شده است. شکل (۵) نمودار مربوط به‌میزان افزایش انرژی اطلاعاتی نسبت به‌میزان کاهش ابعاد را برای داده‌های دیتاست ۲ نشان می‌دهد. برای آنکه دقت عملیات خوشه بندی به‌صورت قابل توجهی تحت تأثیر این عملیات قرار نگیرد با حفظ ۹۹.۹۰٪ انرژی اطلاعاتی، تقریباً ۷۵٪ از تعداد ویژگی‌ها (ابعاد داده‌ها) کاهش یافتند. این امر بیانگر این است که می‌توان با از دست دادن مقدار کمی از انرژی داده‌ها (۰.۱٪)، به درصد خوبی از کاهش ابعاد داده‌ها دست یافت.

[1] Labeled Data
[2] Flyable Electro-Mechanical Actuator
[3] Unlabeled Data
[4] Load
[5] Motion
[6] Low Speed Data



شکل ۵. نرخ کاهش ابعاد داده‌ها باافزایش انرژی اطلاعاتی (درصد) در داده‌های دیتاست ۲

۳-۳. ساختن مدل

وظیفه اصلی این مرحله، انتخاب و پیاده سازی یک تکنیک داده‌کاوی مناسب است که این یک فرایند ساده نیست و در عمل، پیاده سازی مبتنی بر چند مدل است و انتخاب بهترین مدل یک وظیفه دیگر است. در بخش ۳-۴، سهم داده‌کاوی مناسب جهت تشخیص ناهنجاری، شامل الگوریتم K- نزدیکترین همسایه، درخت‌های تصمیم و الگوریتم خوشه بندی فازی C-means ارائه می‌شود. دو مدل اول، شامل تکنیک‌های دسته بندی بر روی دیتاست ۱ و مدل آخر، شامل تکنیک خوشه بندی بر روی دیتاست ۲ آزمایش می‌شوند که به‌ترتیب در بخش‌های ۳-۴-۱ و ۳-۴-۲ به آنها پرداخته خواهد شد.

۳-۴. تشخیص ناهنجاری

دسته‌بندی: مدل مشتق شده از فرایند دسته‌بندی می‌تواند به شکل‌های گوناگونی نمایش داده شود. برای این منظور از دو الگوریتم K- نزدیکترین همسایه و درخت‌های تصمیم استفاده شده است.

الگوریتم دسته بندی K- نزدیکترین همسایه: الگوریتم k- نزدیکترین همسایه[1] (K-NN) یکی از ساده‌ترین تکنیک‌های دسته‌بندی بر پایه یادگیری نمونه است. در این روش تصمیم‌گیری اینکه یک مورد جدید در کدام دسته قرار گیرد با بررسی تعدادی (k) از شبیه‌ترین موارد یا همسایه‌ها انجام می‌شود. تعداد موارد برای هر کلاس شمرده می‌شوند و مورد جدید به دسته‌ای که تعداد بیشتری از همسایه‌ها به آن تعلق دارند نسبت داده می‌شود [۳۴]. در اینجا از مقدار k=1، یعنی یک همسایه مجاور برای الگوریتم استفاده شده است. بررسی نتایج مربوط به مقادیر مختلف k به‌عنوان کارهای آتی در نظر گرفته شده است.

این الگوریتم قادر به نمایش پیچیده‌ترین مرزهای تصمیم گیری است و نیاز به آموزش خاصی ندارد. از مزایای دیگر این الگوریتم، توانایی انسان در فهم چگونگی و علت این نوع تعیین کلاس یک نمونه پرسشی است. در بسیاری از الگوریتم‌ها مانند شبکه‌های عصبی و یا دسته بندی کننده‌های آماری، فهم اینکه چگونه یک الگوریتم تصمیم گیری می‌کند برای انسان مشکل است [۳۵]. ازآنجایی که در K-NN برای هر مورد جدید نیاز به محاسبه جدید است، برای افزایش سرعت آن معمولاً تمام داده‌ها در حافظه نگهداری می‌شوند. که این خود نیازمند حافظه زیادی برای نگهداری نمونه‌ها است. همچنین K-NN بار محاسباتی زیادی را روی رایانه قرار می‌دهد زیرا زمان محاسبه به‌صورت فاکتوریلی از تمام نقاط افزایش می‌یابد. درحالی که به‌کاربردن درخت‌های تصمیم یا شبکه عصبی برای یک نمونه جدید فرایند سریعی است. هرچند ساختارهایی مانند KD- Tree [۳۶] برای افزایش سرعت وجود دارند اما در مقابل مشکلاتی همچون به‌روز رسانی داده‌ها را به وجود می‌آورند.

انتخاب تابع فاصله: الگوریتم‌های بر پایه نزدیکترین همسایه برای تصمیم‌گیری در مورد نزدیکترین نمونه‌ها به یک بردار ورودی، از یک تابع فاصله استفاده می‌کنند. بنابراین تصمیم‌گیری در مورد تابع فاصله در هر الگوریتمی که از این نوع دسته‌بندی استفاده می‌کند، انجام می‌شود و این معیار، تأثیر چشمگیری بر روی سیستم یادگیری دارد. معمولاً در این الگوریتم‌ها از فاصله اقلیدسی و ترکیبات آن استفاده می‌شود [۳۷]. این تابع به‌صورت زیر بیان می‌شود:

$$d(x,y)=\sqrt{\sum_{i=1}^{m}(x_i-y_i)^2}, \qquad (۱)$$

که x و y دو بردار ورودی و m تعداد ویژگی‌ها است. و x_i و y_i مقادیر ورودی برای ویژگی iام هستند. این شکل از تابع اقلیدسی در مواردی که همه ویژگی‌ها عددی بوده و میزان گستردگی آنها تقریباً یکسان باشد استفاده می‌شود. در حالتی که ویژگی‌ها به‌طور اساسی بازه‌های متفاوتی دارند و با نرمال‌سازی می‌توان گستردگی را در همه ویژگی‌ها یکسان کرد. روش‌های گوناگونی مانند تقسیم فاصله هر نمونه در هر ویژگی بر بازه آن ویژگی و یا بر انحراف معیار برای نرمال سازی وجود دارند [۳۵].

درخت‌های تصمیم: برای این منظور از الگوریتم درخت دسته بندی و رگرسیون[2] (CART) که جزء روش‌های طراحی بالا به پایین محسوب می‌شود، استفاده شده است. یادگیری در CART از نوع غیر افزایشی است. یعنی الگوریتم، درخت تصمیم گیری مورد نظر را در یک بار آموزش با داده‌های آموزشی یاد می‌گیرد. مسائل اصلی در این روش شامل موارد زیراست:۱- انتخاب معیار تقسیم گره‌ها، ۲- قوانین توقف و ۳- انتساب برچسب به گره‌های پایانی.

پس از ایجاد درخت موردنظر، درخت مزبور بر اساس یک ترتیب بهینه هرس می‌شود. در این الگو، ابتدا شاخه‌هایی هرس می‌شوند که باعث کاهش هزینه (نرخ خطای جانشینی مجدد[3]) شوند.

هر مسیر در درخت تصمیم تا یک برگ معمولاً قابل فهم است. از این لحاظ یک درخت تصمیم می‌تواند پیش‌بینی‌های خود را توضیح دهد که این یک مزیت مهم است. مشکل استفاده از درخت‌های

[1] K-Nearest Neighbor

[2] Classification and Regression Tree

[3] Resubstitution Method

تصمیم آن است که به صورت نمـایی بـا بـزرگ شـدن مسـئله رشـد می‌کنند که این امر قدری غیرهوشمندانه است. اما مسئله مهم‌تر این استکه امکان اُورفیت[1]شدن با مجموعه داده‌هـا را دارنـد. در [۳۸] بـه توضیح راهکارهایی برای مقابله با این مسئله پرداخته شده است.

۳-۴-۲. خوشه‌بندی

مهمترین چالش در خوشه‌بندی، استخراج ویژگی‌ها و بازنمایی الگوها اسـت. محققـین در زمینـه‌هـای بازشناسـی الگـو، از ایـن مرحلـه چشم‌پوشی کرده و فرض می‌کنند که بازنمایی الگوها به‌عنوان ورودی الگـوریتم‌هـای خوشـه‌بنـدی در دسـترس اسـت. دومـین مرحلـه در خوشه‌بندی، محاسبه شباهت است. روش‌های مختلفی برای محاسـبه شباهت میان دو الگو مورد اسـتفاده قـرار گرفتـه اسـت. گـروه بنـدی، مرحله بعدی در خوشه‌بندی است. به‌طور کلی گروه بندی از دو جنبه قابل بررسی است: سلسله مراتبـی و افـرازبنـدی. روش‌هـای سلسـله مراتبی چند منظوره‌تر و روش‌های افرازبندی کم هزینه‌تر هستند. در برخی برنامه‌ها نظیر بازیابی اطلاعات[2]، بهتر است نتایج خـوش‌بنـدی یک افراز نباشد، یعنی خوشه‌ها می‌توانند با یکدیگر هم‌پوشانی داشته باشند. خوشه‌بندی فازی برای حل چنین مسائلی مناسب است [۲۹].

الگوریتم خوشه‌بندی فازی C-means: الگوریتم خوشه‌بندی فازی C-means (FCM) [۳۹]، به‌طور گسترده روش خوشه‌بنـدی فـازی را در بازشناسی الگو که متعلق بـه دو یـا بیشـتر خوشـه باشـد، بـه‌کـار مـی‌بـرد. الگـوریتم خوشـه‌بنـدی FCM مجموعـه نقـاط داده‌هـای Xj(j= 1, 2, …, n) را بـه خوشـه‌هـای Ci(i= 1, 2, …, c) بـا حـداقل کردن تابع هدف زیر افراز می‌کند [۳۹]:

$$J_m = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}{}^{m}\left\|V_i - X_j\right\|^2, \qquad (۲)$$

که در آن، $\left\|V_i - X_j\right\|$ فاصله اقلیدسی بین مجموعـه داده X_j و مرکـز خوشه V_i(i= 1, 2, …, c)، u_{ij} درجه عضویت تعلق X_j به خوشه C_i، m شاخص فازی بودن، $m\geq 1$، n تعداد نقاط داده‌ها و c تعداد خوشـه‌هـا است. رویه‌های الگوریتم خوشه‌بنـدی FCM از [۳۹] بـه‌صـورت زیـر بازبینی شده است:

مرحله ۱: درجه عضویت u_{ij} را به‌صورت تصادفی مقداردهی می‌شـود، به‌طوری‌که $0\leq u_{ij}\leq 1$، $\sum_{i=1}^{c} u_{ij}=1$، $1\leq i\leq c$ و $1\leq j\leq n$ است.

مرحله ۲: مرکز خوشه V_i مربوط به کلاستر C_i محاسبه می‌شود:

$$V_i = \frac{\sum_{j=1}^{n}(u_{ij})^m \times X_j}{\sum_{j=1}^{n}(u_{ij})^m}, \qquad (۳)$$

به طوری که $1\leq i\leq c$ است.

مرحله ۳: درجه عضویت u_{ij} را به روز می‌شود به‌طوری که:

$$u_{ij} = \frac{1}{\sum_{d=1}^{c}\left(\frac{\|V_i - X_j\|}{\|V_d - X_j\|}\right)^{\frac{2}{m-1}}} \qquad (۴)$$

$1 \leq i \leq c$ و $1 \leq j \leq n$.

مرحله ۴: مرحله ۲ و مرحله ۳ را تا زمانی که مقدار J_m در معادله (۲) بیشتر از حدی کاهش نیافت، می‌بایست دو مرتبه تکرار شود.

همچنان الگوریتم‌های جدید زیادی در مقالات ارائـه مـی‌شـود کـه هریک از آنها در توصیف داده‌ها نسبت به الگوریتم‌های موجود، بهبود یافته‌اند. برای کاربرانی که قصد استفاده از الگوریتم‌های خوشه‌بنـدی را دارند، ضروری است که تنها به دانش یک تکنیک خـاص معطـوف نشوند، بلکه با جزئیات فرایند جمع‌آوری داده‌ها نیـز آشـنایی داشـته باشند. همچنین این دانش حوزه می‌تواند کیفیت استخراج ویژگی‌هـا، محاسبه شباهت‌ها، خوشه‌بندی و نمایش خوشه‌هـا را بهبـود بخشـد. تکنیک خوشه بندی خاصی وجود ندارد که بتواند به طور کلـی تمـام ساختارهای موجود در داده‌ها را استخراج نماید. زیرا این الگـوریتم‌هـا اغلب براساس معیارهای شباهت به‌کار رفتـه در آن الگـوریتم، شـامل فرضیاتی در مورد قالب و شکل خوشه‌ها می‌باشند [۲۹].

یکی از پارامترهای مهمی کـه بایـد بـرای الگـوریتم خوشـه‌بنـدی مشخص کرد تعیین تعداد خوشه‌ها است. به‌ویـژه در ایـن حـوزه کـه تعداد خوشه‌های تعیین شده می‌تواند تأثیر بسیار زیـادی در عملکـرد آشکارسازی ناهنجاری داشته باشد. تعداد خوشه‌های مناسب بسته به انتخاب هر مجموعه از سنسورهای (داده‌ها) انتخابی یـک زیرسیسـتم خاص (مثلاً سنسورهای مربوط به زیرسیستم گرمایی)، می‌تواند تغییر کند. در این مقاله برای تعیین پارامتر تعداد خوشه‌ها از روش آزمون و خطا استفاده شده است. تعیین پارامتر تعداد خوشه‌ها به طور خودکار و متناسب با هر مجموعه سنسور در یک زیرسیستم به‌عنوان کارهای آتی در نظر گرفته شد.

۴. تفسیر مدل و استخراج نتایج

در نهایت، مـدل‌هـای ارائـه شـده مـی‌بایسـت فـرد خبـره را در امـر تصمیم‌گیری یاری نمایند، از اینرو لازم است که ایـن مـدل‌هـا قابـل تفسیر باشند. باید توجه کرد که دقت مدل و دقت تفسیر مدل قدری با یکدیگر در تناقض هستند. معمولاً مدل‌های ساده بیشتر قابل قبول اسـت امـا دقـت آنهـا پـایین مـی‌باشـد. بنـابراین انتظـار مـی‌رود تکنیک‌های داده‌کاوی مدرن نتایجی با دقت بالاتر در اختیار مـا قـرار دهند. در مقابل، تفسیر ایـن مـدل‌هـا نیـز یـک مسـئله مهـم دیگـر است که به‌عنوان یک عملیات مجـزا بـا تکنیـک‌هـایی خـاص جهـت ارزیابی نتایج مورد توجه قرار گرفته است [۲]. درحال حاضر، خروجی تولید شده توسط تکنیـک‌هـای تشـخیص ناهنجـاری بـه یکـی از دو صورت زیر می‌باشند [۴۰]:

۴-۱. برچسب‌ها[3]

در این نوع خروجی، به هر نمونه یک برچسب که مخـتص آن کـلاس می‌باشد نسبت داده می‌شود. جهـت ارزیـابی الگـوریتم دسـته بنـدی K- نزدیکترین همسـایه و درخـت رگرسـیون از ایـن روش اسـتفاده

[1] Overfit
[2] Information Retrieval
[3] Labels

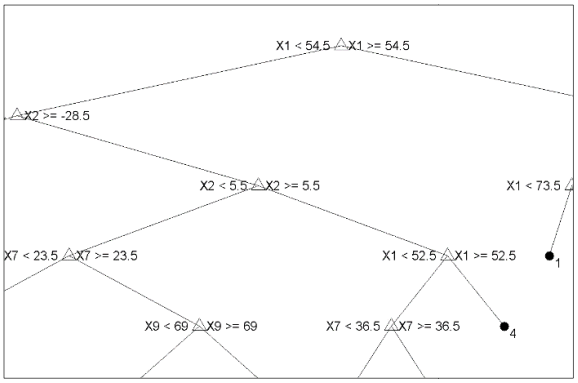
می‌شود که قابل اعمال بر روی دیتاست ۱ است. دیتاست ۱ شامل داده‌هایی با ابعاد ۹ و تعداد ۷ کلاس می‌باشد که تقریباً ۸۰٪ داده‌ها متعلق به کلاس ۱ هستند. از اینرو دقت پیش‌فرض حدود ۸۰٪ است. هدف در اینجا افزایش دقت دسته‌بندی می‌باشد. جدول (۳) نتایج حاصل از دسته بندی دیتاست ۱ را با استفاده از الگوریتم K- نزدیکترین همسایه نشان می‌دهد. همان‌طور که مشاهده می‌شود، با به‌کارگیری K-NN دقت دسته بندی موردنظر به ۹۹.۸۸٪ رسیده است و زمان محاسبه شده برای فرآیندهای یادگیری و آزمایش پس از اجرای ۱۰ بار الگوریتم و متوسط گیری بر روی آن به‌دست آمده است.

با اعمال الگوریتم CART به نمونه‌های آموزشی دیتاست ۱، درخت تصمیم نشان داده شده در شکل (۶) ایجاد خواهد شد. تعداد ۵۵ قانون از این درخت به‌دست آمده است. جدول (۴) نتایج حاصل از اعمال قوانین دسته بندی به‌دست آمده از درخت تصمیم را بر روی داده‌های آزمایشی نشان می‌دهد. همان‌طور که قابل مشاهده است، دقت دسته‌بندی موردنظر در مقایسه با K-NN،به ۹۹.۹۶٪ افزایش یافته است.

برخلاف دسته‌بندی کننده‌های تک مرحله‌ای رایج نظیر K-NN که هر نمونه از داده‌ها روی تمام دسته‌ها امتحان می‌شود، در CART یک نمونه فقط روی زیرمجموعه‌های خاصی از دسته‌ها امتحان می‌شود و به این نحو محاسبات غیرلازم حذف می‌شوند. از اینرو همان‌طور که در جدول (۴) نشان داده شد، سربار زمانی در مرحله آزمایش کاهش چشمگیری پیدا کرده است. شکل (۷) روند کاهش هزینه (درصد خطا) درخت‌های تصمیم را همراه با افزایش گره‌ها برای داده‌های آموزش نشان می‌دهد.

جدول ۳. نتایج دسته بندی با استفاده از K-NN

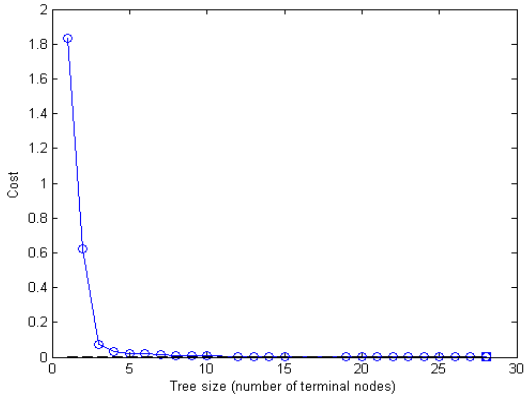
داده‌های آزمایشی	داده‌های آموزشی	دیتاست ۱
۱۴۵۰۰	۴۳۵۰۰	تعداد نمونه‌ها
۱۱۴۷۰	۳۴۱۰۸	تعداد مثبت درست
۶	۰	تعداد مثبت غلط
۳۰۱۶	۹۳۹۲	تعداد منفی درست
۸	۰	تعداد منفی غلط
۰/۹۹۸۸	۰/۷۲۴۶	دقت دسته بندی
۰/۰۰۱۲	۰/۲۷۵۴	نرخ خطا
۲/۳۸۳۶	۲/۳۸۳۶	زمان صرف شده (s)



شکل ۶. نمایی از درخت تصمیم ایجاد شده با استفاده از الگوریتم CARD

جدول ۴. نتایج دسته بندی با استفاده از CART

داده‌های آزمایشی	داده‌های آموزشی	دیتاست ۱
۱۴۵۰۰	۴۳۵۰۰	تعداد نمونه‌ها
۱۱۴۷۳	۳۴۱۰۸	تعداد مثبت درست
۱	۰	تعداد مثبت غلط
۳۰۲۱	۹۳۹۲	تعداد منفی درست
۵	۰	تعداد منفی غلط
۰/۹۹۹۶	۰/۷۲۴۶	دقت دسته بندی
۰/۰۰۱۴	۰/۲۷۵۴	نرخ خطا
۰/۰۰۳۶	۰/۶۷۰۱	زمان صرف شده (s)



شکل ۷. هزینه تولید درخت تصمیم با افزایش گره‌ها

۴-۲. امتیازها[1]

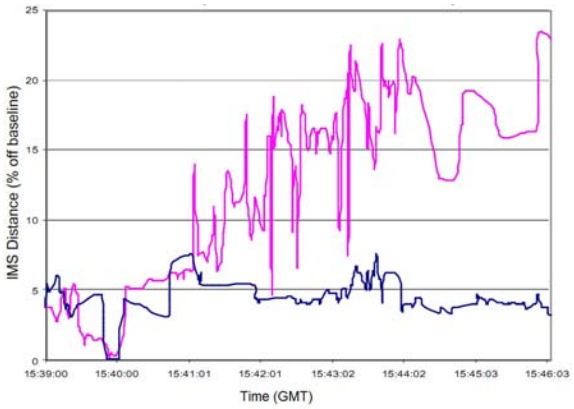
تکنیک‌های مبتنی بر امتیازدهی، یک درجه ناهنجاری به هر نمونه از داده‌های آزمایشی را با توجه به فاصله آنها از الگوی نرمال نسبت می‌دهند. در این گونه موارد ممکن است تحلیل‌گر از یک مقدار آستانه (خاص حوزه) جهت تشخیص ناهنجاری استفاده کند. به‌منظور ارزیابی عملکرد الگوریتم FCM از این نوع خروجی استفاده می‌شود

[1] Scores

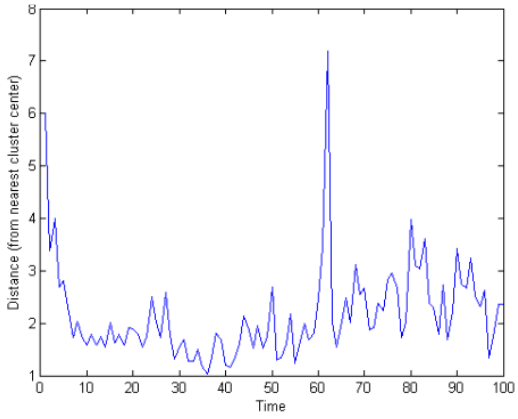
به‌طوری که پس از عملیات پیش پردازش، ابتدا با اعمال الگوریتم بر روی دیتاست ۲ مراکز خوشه‌های تولید شده محاسبه می‌شود و پس از آن، نتیجه اعمال الگوریتم بر روی داده‌های آزمایش را به فرم یک منحنی که نشان دهنده میزان فاصله از وضعیت نرمال سیستم است نمایش داده می‌شود. یعنی فاصله هر نمونه تا نزدیکترین مرکز خوشه به‌دست آورده می‌شود. چنین تحلیلی از WLEIDS مربوط به فاجعه کلمبیا در مأموریت STS-107 توسط IMS صورت گرفته است [۲۱]. شکل (۸) نتایج حاصل از این تحلیل را برای پرتاب کلمبیا با تعداد سه سنسور برای هر بال نشان می‌دهد. محور افقی نشان دهنده زمان با شروع از لحظه پرتاب است و محور عمودی اندازه انحراف IMS را از رفتار نرمال نشان می‌دهد. این اندازه با درصدی از بیشینه انحراف تحت پوشش داده‌های آموزشی مقیاس‌بندی شده است. نتایج مربوط به بال چپ و راست به‌ترتیب با خطوط روشن‌تر و تاریک‌تر نشان داده شده است. همان‌طور که در محور عمودی مشخص است، زمان ۱۵:۴۰:۲۲، لحظه رخداد ضربه حباب را نشان می‌دهد که سبب از کار انداختن سیستم حفاظت گرمایی در بال چپ شده است. پس از این نقطه از زمان، منحنی مربوط به بال چپ به‌وضوح شروع به واگرایی می‌کند، در حالی که منحنی مربوط به بال راست همچنان در بازه معقولی از رفتار اسمی قرار دارد.

جهت ارزیابی داده‌های دیتاست ۲، با استفاده از الگوریتم K- Fold Cross Validation [۳۸]، بخشی از داده‌ها (1/k) در هر تکرار به عنوان داده‌های آزمایش در نظر گرفته شده‌اند. شکل (۹) تحلیل انجام شده توسط FCM را بر روی داده‌های آزمایش ایجاد شده از دیتاست ۲ با استفاده از این الگوریتم نشان می‌دهد. همان‌طور که مشاهده می‌شود منحنی ایجاد شده رفتار نرمالی را برای سیستم نشان می‌دهد. لازم به ذکر است که مقدار زیاد فاصله در زمان ۶۰ به‌دلیل تغییر وضعیت از بالک Y (Actuator Y) به بالک X (Actuator X) در زمان ۱ دقیقه است. اگر چه این تحلیل به‌صورت برون‌خط و با استفاده از داده‌های بایگانی شده صورت گرفته ولی تکنیک‌های ارائه شده در این مقاله قابلیت پیاده‌سازی در پایش بی‌درنگ را نیز دارا می‌باشند.

پیچیدگی محاسباتی برای انتخاب یک تکنیک تشخیص ناهنجاری جنبه‌ای کلیدی محسوب می‌شود. اگرچه الگوریتم‌های CART و FCM دارای هزینه زمانی بالایی برای فاز یادگیری هستند، ولی در مرحله آزمایش معمولاً سریع عمل می‌کنند. به‌ویژه در CART که یک نمونه فقط روی زیرمجموعه‌های خاصی از دسته‌ها امتحان شده و روی تمام دسته‌ها امتحان نمی‌شود. معمولاً این امر قابل قبول است زیرا مدل‌ها می‌توانند در حالت برون‌خط [1] آموزش دیده و در مرحله آزمایش برای کاربردهای برخط [2] مناسب باشند. از اینرو استفاده از این الگوریتم‌ها می‌تواند نیازمندی‌های لازم برای پردازش برخط و نزدیک به زمان واقعی [3] را در عملیات فضایی برآورده نمایند. در مقابل، تکنیک‌هایی نظیر K-NN که فاقد فاز آموزش می‌باشند دارای مرحله آزمایش پرهزینه‌ای هستند که این امر به‌عنوان یک محدودیت جدی در محیط واقعی تلقی می‌شود.



شکل ۸. نتایج حاصل از تحلیل IMS از پرتاب STS-107 کلمبیا



شکل ۹. نتایج حاصل از تحلیل FCM در دیتاست ۲

۵. نتیجه‌گیری

در این مقاله، یک سیستم پایش خودکار جهت آشکارسازی ناهنجاری‌ها در زیرسیستم‌های فضایی با استفاده از روش‌های داده کاوی ارائه شده است. یک ویژگی قدرتمند تکنیک‌های داده کاوی و یادگیری ماشین، قابلیت تحلیل همزمان بر روی چند متغیر می‌باشد که به آنها اجازه می‌دهد تا تعاملات بین پارامترهای مرتبط به هم را که در نظر گرفتن آنها به طور مجزا در حین پایش مشکل می‌باشد، کشف و مدل نماید. با توجه به نتایج به‌دست آمده مشخص گردید که اینگونه روش‌ها قادر به بهبود امکانات عیب‌شناسی سنتی یا ایجاد یک مدل سیستمی جدید برای تشخیص ناهنجاری می‌باشند. این روش‌ها در ترکیب با رویکردهای سنتی می‌توانند ایمنی و قابلیت اعتماد را برای عملیات فضایی افزایش دهند.

به‌تازگی یادگیری گروهی به‌عنوان یک روش قدرتمند بر پایه به‌کارگیری چند الگوریتم برای تحلیل داده‌ها و ترکیب نتایج آنها به

[1] Offline
[2] Online
[3] Near-Real Time

منظور دست یافتن به نتایج با کیفیت‌تر مورد توجه قرار گرفته است. از قابلیت‌های این روش، تولید خوشه‌هایی با کیفیت و قدرت بیشتر، استخراج ساختارهای ناشناخته و گوناگون داده‌ها و مقیاس‌پذیری آن است. همچنین، در این روش نیازی به داشتن اطلاعات زمینه‌ای در مورد الگوریتم مورد استفاده و یا داده‌ها وجود ندارد. به‌طورکلی مشاهده شده که الگوریتم‌های خوشه‌بندی ارائه شده در مقاله‌ها، وابسته به مجموعه داده مورد استفاده و تابع شباهت به‌کارگرفته‌شده درالگوریتم هستند و بنابراین قابلا عمال بر روی هرگونه مجموعه داده‌های آموزشی نیستند. یادگیری گروهی بر استفاده از چند الگوریتم متفاوت برای تولید خوشه‌های گوناگون و ترکیب نتایج آنها و استخراج یک خوشه‌بندی قدرتمند استوار است. یکی از مسائلی که یادگیرنده‌های گروهی با آن مواجه هستند، زمان پردازش بالای این یادگیرنده‌ها است که این امر به‌دلیل اجراهای گوناگون الگوریتم پایه است. باتوجه به مقیاس‌پذیری روش‌های ترکیبی می‌توان از موازی‌سازی برای توسعه الگوریتم معرفی شده در [۲۹] جهت خوشه‌بندی داده‌های جریانی بهره جست. آزمایش این الگوریتم بر روی مجموعه داده‌های جریانی استاندارد، نتایج امیدوارکننده‌ای از کارایی و دقت لگوریتم را نشان‌داده است. به‌کارگیری این تکنیک‌ها بر روی داده‌های جریانی می‌تواند به‌عنوان بخشی از کارهای آتی در نظر گرفته شود.

۶. مراجع

[1] Mobasheri, M. R. “Foundations of Physics in Remote Sensing and Satellite Technology”; K. H. Toosi Univ. of Tech., Tehran, 2006.

[2] Goodarzi, M. “Implementation and Simulation of a Satellite Tracking Algorithms Based on Step-Track for Satellite Ground Stations”; MSc. Thesis, Malek-ashtar Univ. of Tech., Tehran, 2008.

[3] Dehghan, M. “Design and Simulation of the TT&C Station for LEO Satellites”; MSc. Thesis, Malek Ashtar Univ. of Tech., Tehran, 2008.

[4] Girimonte, D.; Izzo, D. “Artificial Intelligence for Space Applications”; In Schuster, A. J. Editors, Intelligent Computing Everywhere, Springer London, 2007, 235-253.

[5] Duke, E. L.; Regenie, V. A. “Expert Systems Development and Application”; NASA Ames Research Center, Dryden Flight Research Facility, Edwards, California, October 1985.

[6] Regenie, V. A.; Duke, E. L. “Design of an Expert-System Flight Status Monitor”, NASA Ames Research Center, Dryden Flight Research Facility, Edwards, California, Augusta 1985.

[7] Disbrow, J. D.; Duke, E. L.; Regenie, V. A. “Development of a Knowledge Acquisition Tool for an Expert System Flight Status Monitor”; NASA Ames Research Center, Dryden Flight Research Facility, Edwards, California, January 1986.

[8] Duke E. L.; Regenie, V. A. “Description of an Experimental Expert System Flight Status Monitor”; NASA Ames Research Center, Dryden Flight Research Facility, Edwards, California, October 1985.

[9] Norcross, S.; Grieser, W. H. “Spacecraft Command and Control Using Expert Systems”; In NASA. Goddard Space Flight Center, 3rd International Symposium on Space Mission Operations and Ground Data Systems, July 22, 1994, Part 2, 735-740.

[10] Esp, D. G.; Ekwue, A. O.; Macqueen, J. F. “Expert Systems for Telemetry Processing”; National Grid Company (NRC), 1993.

[11] Disbrow, J. D.; Duke, E. L.; Ray, R. J. “Preliminary Development of an Intelligent Computer Assistant for Engine Monitoring”; NASA Ames Research Center, Dryden Flight Research Facility, Edwards, California, August 1989.

[12] Healey, K. J. “Artificial Intelligence Research and Applications at the NASAJohnson Space Center”; AI Magazine (© AAAI), 1986, 7, 146-152.

[13] Semmel, G. S.; Davis, S. R.; Leucht, K. W.; Rowe, D. A.; Smith, K. E.; Bölöni, L. “NESTA: NASA Engineering Shuttle Telemetry Agent”, American Association for Artificial Intelligence (AAAI), 2005.

[14] Semmel,G. S.; Davis, S. R.; Leucht, K. W.; Rowe, D. A.; Kelly, A. O.; Bölöni, L. “Launch Commit Criteria Monitoring Agent”; Association for Computing Machinery (ACM), 2005.

[15] Semmel, G. S.; Davis, S. R.; Leucht, K. W.; Rowe, D. A.; Smith, K. E.; Bölöni, L. “Space Shuttle Ground Processing with Monitoring Agents”; IEEE Intelligent Sys., 2006.

[16] Iverson, D. L. “Data Mining Applications for Space Mission Operations System Health Monitoring”; In Proc. of the Space Ops 2008 Conf., ESA, EUMETSAT, AIAA, Heidelberg, Germany, May 2008.

[17] Iverson, D. L. “System Health Monitoring for Space Mission Operations”; In Proc. 2008 IEEE Aerospace Conf. Big Sky, Montana, 2008, 1–8.

[18] Schwabacher, M. “Machine Learning for Rocket Propulsion Health Monitoring”; SAE Trans. 2005, 114, 1192-1197.

[19] Joshi, A.; Gavriloiu, V.; Barua A.; Garabedian, A.; Sinha, P.; Khorasani, K. “Intelligent and Learning-based Approaches for Health Monitoring and Fault Diagnosis of RADARSAT-1 Attitude Control System”; 2007, 3177-3183.

[20] Barnett, V.; Lewis, T. “Outliers in Statistical Data”; John Wiley &Sons., 3rd Ed., 1994.

[21] Iverson, D. L. “Inductive System Health Monitoring”; In Proc. of The 2004 Int. Conf. on Artificial Intelligence (IC-AI '04), CSREA, Las Vegas, Nevada, June 2004.

[22] Schwabacher, M.; Waterman, R. “Pre-Launch Diagnostics for Launch Vehicles”, In Proc. of the IEEE Aerospace Conf. Big Sky, MT, March 2008.

[23] Vecchio, E.; Lazzerini, B.; Foley, S.; Donati, A. “Spacecraft Fault Analysis Using Data Mining Techniques”; In Proc. of the 8th Int. Symposium on Artificial Intelligence, Robotics and Automation in Space, Munchen, Germany, 5-8 September 2005. Published in CDROM.

[24] Long, X.; Hao-Dong, M.; Hong-Zheng, F.; Ke-Xu Z.; Da-Wei, Y. “Anomaly Detection of Spacecraft Based on Least Squares Support Vector Machine”; In Proc. of IEEE Prognostics & System Health Management Conf. Shenzhen, 2011.

[25] Rasmussen, E. “Clustering Algorithms”; In Frakes, W. B.; Baeza-Yates, R. Editors, Information Retrieval: Data Structures and Algorithms, Prentice Hall, Englewood Cliffs,1992, 419-42.

[26] Martin, R. A. “Evaluation of Anomaly Detection Capability for Ground-Based Pre-Launch Shuttle Operations”; NASA Ames Research Center, U.S.A., 2009.

[27] “Columbia Accident Investigation Board Report”; 2003, 1.

[28] Kantardzic, M. “Data Mining: Concepts, Models, Methods, and Algorithms”; John Wiley & Sons, 2003.

[29] Mokhtari, V. “Stream data Clustering using Paralleling Hybrid Algorithms, MSc Thesis, Azad Univ., Qazvin, 2007.

[30] Frank, A.; Asuncion, A. “UCI Machine Learning Repository”; Univ. of California, Irvine, School of Information and Computer Sci., 2010.

[31] Balaban, E.; Narasimha, S.; Roychoudhury, I; Saxena A. “Flyable Electro-Mechanical Actuator (FLEA) Testbed Datasets”; 2 Ed. NASA DASHlink Lab, https://c3.nasa.gov/dashlink/resources/503/, 2012.

[32] Bishop, M. “Pattern Recognition and Machine Learning”; Singapore: Springer, 2006.

[33] Yairi, T.; Kawahara, Y.; Takata, N. “Spacecraft Telemetry Data Monitoring by Dimensionality Reduction Techniques”; in SICE Annual Conf. Taipei, Taiwan, 2010, 1230-1234.

[34] Hand, D.; Mannila, H.; Smyth, P. “Principles of Data Mining”; MIT Press, 2001.

[35] Moosavi, S. M. R; Javan, M. F.; Ghodratnama, S.; SadrAldini, M. H. “An Innovative Sampling Method for Massive Data Reduction in Data Mining”; The 3rd Iran Data Mining Conf., Tehran, 2009.

[36] Moore, A. W. “Effcient Memory-Based Learning for Robot Control (An Intoductory Tutorial on kd-Trees)”; In Computer Laboratory: Cambridge, 1999.

[37] Wilson D. R.; Martinez, T. R. “Reduction Techniques for Instance-Based Learning Algorithms”; Machine Learning. 2000, 38, 257-286.

[38] Mitchel, T. M. “Machine Learning”; McGraw-Hill, 1997.

[39] Bezdek, J. C. “Pattern Recognition with Fuzzy Objective Function Algorithms”; NY: Plenum Press, 1981.

[40] Chandola, V.; Banerjee, A.; Kumar, V. “Anomaly Detection: A Survey”; ACM Computing Surveys (CSUR). 2009, 41, 1-72.