

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
إِنَّمَا أَمْرُهُ إِذَا أَرَادَ شَيْئًا أَنْ يَقُولَ لَهُ كُنْ فَيَكُونُ

سوره یس، آیه ۸۲

به نام خداوند بخشاینده و مهربان
"کنترل کننده تمام"، که چون اراده نماید، میان اراده و آنچه باید شد، فاصله ای نیست...



دانشگاه صنعتی شاهرود
دانشکده مهندسی برق و رباتیک
گروه کنترل

بهبود کارایی احتراق بویلر با به کارگیری کنترل هوشمند
مبتنی بر داده کاوی

دانشجو: سیده سارا اقوامی

استاد راهنما:

دکتر حیدر طوسی‌ان شاندیز

استاد مشاور:

دکتر محمدرضا جاهد مطلق

پایان نامه جهت اخذ درجه کارشناسی ارشد

فروردین ۱۳۹۰

تقدیم به:

مادرم؛ مهربانترین و پدرم؛ بزرگوارترین آموزگاران زندگی ام...

با سپاس فراوان از

استاد محترم آقای دکتر حیدر طوسیان شانندیز به خاطر راهنمایی‌های ارزنده‌شان؛

استاد بزرگوار آقای دکتر محمد رضا جاهد مطلق به خاطر همراهی‌های پرمهرشان؛

مدیر گروه کنترل و ابزار دقیق شرکت صنایع آذرآب، آقای مهندس علی‌رضا مهابادی

و مدیریت فنی و بهره برداری واحد بخار مجتمع پتروشیمی مارون، منطقه ویژه اقتصادی ماهشهر.

چکیده

به موازات پیشرفت سخت افزاری و نرم افزاری سیستم‌های ذخیره‌سازی و مدیریت داده، کاربرد دانش داده‌کاوی در بسیاری از زمینه‌های مختلف در حال گسترش است. حوزه فرآیندهای صنعتی و استفاده از بایگانی داده‌های عملیاتی ثبت شده از فرآیندها برای تحلیل و کنترل آنها از جمله این زمینه‌هاست. در این پژوهش پس از معرفی یک بویلر صنعتی به عنوان یک فرآیند نمونه، اطلاعاتی از تاریخچه عملیاتی آن مورد بررسی قرار می‌گیرد. این اطلاعات شامل مقادیر ۱۵ متغیر مهم ثبت شده در بازه‌های زمانی مشخص هستند. هدف پژوهش استفاده از روش داده‌کاوی برای افزایش کارایی بویلر است. از میان ۱۵ متغیر ثبت شده، دمای گاز خروجی و میزان اکسیژن موجود در گاز خروجی به نوعی معیارهایی از میزان کارایی فرآیند احتراق هستند. داده‌های عملیاتی مورد پیش پردازش و آماده‌سازی قرار گرفته و ۱۳ متغیر برای مدل سازی دو متغیر تعیین کننده کارایی در بازه کاری ۸۰-۵۰٪ بار بویلر مورد استفاده قرار می‌گیرند. روش‌های به کار رفته برای مدل سازی شامل دو روش رگرسیون جزئی حداقل مربعات (PLS) و شبکه‌های عصبی است. روش PLS به دلیل داشتن قابلیت کاهش ابعاد مجموعه داده که امتیاز مهمی در بررسی پایگاه‌های بزرگ داده محسوب می‌شود مورد توجه قرار گرفته و از شبکه‌های عصبی به عنوان ابزاری عمومی برای مدل سازی و تخمین مقادیر پاسخ استفاده شده است.

در مرحله بعد، داده‌ها به دو دسته داده‌های آزمایش و آموزش تقسیم می‌شوند. با معرفی معیاری برای سنجش کارایی بویلر که شامل دو پارامتر یاد شده دما و اکسیژن است، به ازای هر داده از مجموعه آزمایش، مجموعه‌ای از داده‌های آموزش که کارایی بهتری نسبت به داده مذکور دارند تشکیل می‌شود. در این مجموعه با استفاده از الگوریتم K-Means اصلاح شده، الگوهای عملیاتی برای داشتن راندمان بیشتر ساخته می‌شوند و نزدیک‌ترین الگو به داده مذکور به عنوان الگوی هدف انتخاب می‌شود. مقدار متغیرهای کنترل پذیر داده آزمایش با مقادیر متناظر در الگوی هدف جایگزین می‌شود. داده‌های تغییر یافته گروه آزمایش به مدل به دست آمده از مرحله قبل وارد شده و نتیجه، کاهش دمای گاز خروجی و میزان اکسیژن موجود در آن را نشان می‌دهد. ب این ترتیب با تغییر متغیرهای کنترل پذیر از مجموعه ۱۳ متغیر در اختیار، فرآیند احتراق بویلر به سمت قرار گرفتن در وضعیت‌های کاری متناظر با راندمان بیشتر هدایت می‌شود.

کلید واژه‌ها: داده‌کاوی، شبکه عصبی، رگرسیون جزئی چند متغیره، احتراق، بهبود کارایی

فهرست مقاله‌های ارائه شده

1. "Modeling the Performance of an Industrial Process Based on Neural Networks and Data Mining", ICCRC 2011, will be published into conference proceedings by IEEE, and indexed by Thomson ISI.
2. "Efficiency Enhancement through Decision Support Based on Data Mining", ICCRC2011, will be published into conference proceedings by IEEE, and indexed by Thomson ISI.

فهرست تصاویر و جداول

شماره	توصیف	صفحه
شکل (۱-۲)	نمودار فرآیند بویلر	۹
شکل (۲-۲)	چرخش آب در مسیر تبخیر	۱۳
شکل (۱-۳)	نمودار تغییرات دمای هوا	۲۴
شکل (۲-۳)	نمودار تغییرات دبی سوخت و دبی هوا	۲۵
شکل (۱-۴)	دسته بندی انواع مدل ها	۳۰
شکل (۱-۵)	متغیرهای اصلی	۳۶
شکل (۲-۵)	اجزاء اساسی	۳۶
شکل (۳-۵)	مدل اکسیژن در گاز خروجی با روش PLS	۴۰
شکل (۴-۵)	مدل دمای گاز خروجی با روش PLS	۴۱
شکل (۱-۶)	مدل دمای گاز خروجی با روش شبکه عصبی	۵۰
شکل (۲-۶)	مدل اکسیژن در گاز خروجی با روش شبکه عصبی	۵۰
شکل (۱-۸)	تقسیم بندی روش های خوشه سازی	۶۴
شکل (۲-۸)	معیار تک خط	۶۹
شکل (۳-۸)	معیار خط کامل	۷۰
شکل (۴-۸)	معیار خط مرکز	۷۰
شکل (۵-۸)	معیار خط میانگین	۷۱
جدول (۱-۹)	تفکیک متغیرها	۷۹
شکل (۱-۹)	پاسخ سیستم به ورودی تغییر یافته-دما	۸۴
شکل (۲-۹)	پاسخ سیستم به ورودی تغییر یافته-اکسیژن	۸۵
شکل (۱-۱۰)	مقایسه PLS و شبکه عصبی در مدل سازی دما	۸۷
شکل (۲-۱۰)	مقایسه PLS و شبکه عصبی در مدل سازی اکسیژن	۸۷

بخش اول: فرآیند صنعتی بویلر و مدل سازی آن

فصل ۱ - مقدمه.....	۴
۱-۱- معرفی مسأله.....	۴
۲-۱- پیشینه پژوهش های مرتبط.....	۵
۳-۱- معرفی فصل ها.....	۷
فصل ۲ - فرآیند صنعتی بویلر.....	۹
۱-۲- معرفی.....	۹
۲-۲- انواع بویلرها.....	۱۰
۱-۲-۲- دسته بندی بویلرها برحسب نوع سوخت.....	۱۰
۲-۲-۲- دسته بندی بویلرها برحسب ساختمان و کاربرد.....	۱۱
۳-۲-۲- اجزاء ساختمان بویلر.....	۱۴
۳-۲- کنترل خودکار بویلر.....	۱۶
۱-۳-۲- معرفی.....	۱۶
۲-۳-۲- حلقه کنترل سوخت و هوا.....	۱۷
۳-۳-۲- حلقه کنترل آب و بخار.....	۱۸
۴-۳-۲- حلقه کنترل دمای بخار.....	۱۸
۵-۳-۲- سایر حلقه های کنترلی.....	۱۹
۴-۲- محاسبه راندمان بویلر.....	۱۹
فصل ۳ - بویلر نمونه و داده های آن.....	۲۱
۱-۳- معرفی.....	۲۱
۲-۳- اعتبارسنجی و پیش پردازش داده ها.....	۲۳
۱-۲-۳- داده های ناخوانا یا ثبت نشده.....	۲۶
۲-۲-۳- نقاط پرتاب.....	۲۷
۳-۲-۳- نرمالیزه کردن داده ها.....	۲۷
فصل ۴ - مدل سازی و تخمین توابع.....	۲۹
فصل ۵ - مدل سازی آماری و کاوش داده ها.....	۳۱
۱-۵- معرفی.....	۳۱
۲-۵- رگرسیون خطی چندگانه.....	۳۲
۳-۵- آنالیز اجزاء اساسی.....	۳۵
۱-۳-۵- معرفی.....	۳۵
۲-۳-۵- استخراج اجزاء اساسی.....	۳۷

۳۷	۴-۵	رگرسیون جزئی حداقل مربعات
۴۰	۵-۵	نتایج مدل سازی به روش PLS
۴۲	۶-۶	شبکه های عصبی
۴۲	۱-۶	معرفی
۴۴	۲-۶	شبکه پرسپترون
۴۵	۳-۶	روشهای بر پایه گرادیان
۴۹	۴-۶	نتایج مدل سازی با استفاده از شبکه های عصبی

*بخش دوم: داده کاوی و بهبود کارایی فرآیند

۵۱	۷-۷	آشنایی با داده کاوی
۵۱	۱-۷	معرفی
۵۲	۲-۷	تفاوت دانش داده کاوی و آمار
۵۳	۳-۷	قابلیت های داده کاوی
۵۵	۴-۷	ساختارهای داده ای
۵۸	۵-۷	گامهای لازم برای داده کاوی
۵۹	۶-۷	کاربردهای داده کاوی
۶۱	۸-۸	الگوریتم های داده کاوی
۶۱	۱-۸	آشنایی با الگوریتم های داده کاوی اجزا آنها
۶۲	۲-۸	الگوریتم های خوشه سازی
۶۵	۱-۲-۸	روش های افزار
۶۸	۲-۲-۸	روش های سلسله مراتبی
۷۳	۳-۸	کلاس بندی و الگوریتم های درخت تصمیم گیری
۷۵	۴-۸	الگوریتم شناسایی قوانین متحد
۷۷	۹-۹	الگوریتم کنترل هوشمند مبتنی بر داده کاوی
۷۷	۱-۹	معرفی
۷۸	۲-۹	بررسی داده ها و معیار راندمان
۸۱	۳-۹	بدنه الگوریتم کنترل هوشمند
۸۳	۴-۹	محاسبه فاصله
۸۳	۵-۹	نتیجه به کارگیری الگوریتم کنترل هوشمند
۸۶	۱۰-۱۰	نتیجه گیری و پیشنهاد
۸۶	۱-۱۰	نتایج بخش مدل سازی

۸۸	۱۰-۲- نتایج بخش کنترل
۹۰	۱۰-۳- پیشنهادات
۹۱	فهرست منابع
۹۳	پیوست

فصل ۱ - مقدمه

۱-۱ - معرفی مسأله

فرآیند احتراق پایه بسیاری از فرآیندهای صنعتی است؛ از جمله فرآیند مادر تولید بخار، که خود به نوبه خود قلب تپنده بسیاری از صنایع کوچک و بزرگ، از نیروگاهها و پالایشگاههای بزرگ تا کارخانجات کوچک تولیدی است. در کنار این گستردگی و اهمیت کاربرد، تعریف یک مدل دقیق و مفهومی ریاضی از این پدیده در مقیاس صنعتی بسیار دشوار است. علاوه بر ویژگیهای غیرخطی، احتراق یک پدیده غیرایستاست که مشخصات آن با زمان تغییر می‌کند و مجموعه این شرایط، به ناکارآمدی مدل‌سازی با روشهای کلاسیک و راهبردهای کنترلی از پیش تعیین شده منجر می‌شود. به موازات این مسأله، دانش داده‌کاوی به عنوان دانش نو ظهوری که همراه با پیشرفت امکانات سخت‌افزاری و نرم‌افزاری ذخیره‌سازی داده رشد کرد و به مرور به بسیاری از حوزه‌های دانش و تجربه راه یافت و به حوزه فرآیندهای صنعتی نیز داخل شد [۱]. کاربرد گسترده سیستم‌های کنترل توزیع شده (DCS)^۱ در کنترل، حفاظت و مدیریت فرآیندهای صنعتی و امکانات پیشرفته این سیستم‌ها در حفظ داده‌های عملیاتی فرآیند و تنظیم و ذخیره آنها به صورت گزارش‌ها، منحنی‌ها و فهرست وقایع، مجموعه ارزشمندی از اطلاعات رفتاری فرآیند را در اختیار می‌گذارند. در واقع به

^۱ Discrete Control System

تعبیری هر فرآیند با "زبان" مقادیر پارامترهای خود و آهنگ تغییرات آنها در طول زمان حرف می‌زند؛ و برای تحلیل، مدل‌سازی، کنترل و حتی پیش‌بینی شرایط ناخواسته‌ای مانند خروج تجهیزات از حالت عادی عملکرد خود، خرابی، وقوع شرایط اضطراری در عملکرد فرآیند و یا بهبود بخشیدن به شرایط کاری آن شاید "گوش دادن" به خود فرآیند کارتر از طراحی و تحلیل بر اساس معادلات و روش‌های ریاضی و بر روی کاغذ باشد. علاوه بر این مسائلی مانند شرایط خاص محیطی و تأثیرات آن، ملاحظات مربوط به نصب و یا شرایط بهره‌برداری در مراحل طراحی نظری منظور نمی‌شوند، اما نقش آنها از بایگانی اطلاعات عملکرد فرآیند قابل استخراج است.

با توجه به این موضوعات و همچنین با اهمیت روزافزونی که مدیریت منابع انرژی و استفاده بهینه از سوخت پیدامی‌کند، بررسی امکان مدل‌سازی، کنترل هوشمند و هدایت فرآیند احتراق بویلر به سمت رفتاری منجر به عملکرد با سطح بیشتری از کارایی بر اساس داده‌های واقعی فرآیند، مساله مهمی به نظر می‌رسد.

۱-۲- پیشینه پژوهش‌های مرتبط

فعالیت‌های انجام شده در راستای بهبود بخشیدن به کارایی فرآیند احتراق بر سه روش استفاده از مدل‌های تحلیلی یعنی قواعد ترمودینامیک و معادلات ریاضی، استفاده از مدل‌های تخمینی عصبی و فازی و استفاده از روشهای ترکیبی هوشمند متمرکز هستند. به عنوان مثالهایی از فعالیت‌های

انجام شده بر اساس دو روش اول، لی و همکارانش^۱ [۲] یک روش جستجوی برخط برای پیدا کردن نسبت بهینه سوخت و هوا بر اساس معادلات کارایی حرارتی پایه گذاری کردند. هاولنا^۲ و همکارانش [۳] از یک مدل چندمتغیره کنترل پیش‌بین برای بهبود کارایی و کاهش آلاینده NOx در گاز احتراق استفاده کردند و کس^۳ و همکارانش یک مدل شبکه‌عصبی به همین منظور، یعنی بهبود کارایی و کاهش آلاینده NOx پیشنهاد کردند [۴]. مثال‌های مشابه بسیار دیگری نیز در این رابطه وجود دارد؛ اما روش سوم، یعنی به‌کارگیری روشهای ترکیبی هوشمند مانند داده‌کاوی به‌تازگی مورد توجه بسیاری قرار گرفته است. کوسیاک^۴ از داده‌کاوی برای بهبود کارایی فرآیند احتراق و بهبود بخشیدن به عملکرد سیستم بویلر-توربین استفاده کرده است [۶]، [۵] و [۷]. وی به طور خاص روش خوشه‌سازی را مورد توجه قرار داده است. زینگ^۵ و همکارانش از ترکیب اکتشاف قوانین متحد بین داده‌ها و فازی سازی برای بهبود احتراق در یک بویلر با سوخت ذغال سنگ استفاده کرده اند [۸]. جیانگ کیانگ^۶ و همکارانش [۹] بر اساس آنالیز همبستگی بین پارامترها به دنبال افزایش کارایی در یک نیروگاه بودند و تائو جین^۷ و همکارش از روش PLS به عنوان یکی از روش‌های داده‌کاوی برای مدل‌سازی احتراق استفاده کرده اند [۱۰].

¹ Li

² Havlena

³ Cass

⁴ Kusiak

⁵ Xining

⁶ Jian-qiang

⁷ Tao jin

۱-۳- معرفی فصل‌ها

با توجه به اهمیت مورد اشاره موضوعات، این پژوهش از دو بخش اصلی تشکیل شده است. در بخش اول فرآیند احتراق در بویلر به طور عمومی معرفی شده و سپس بویلر مورد بحث و بایگانی داده‌های عملکرد آن و مراحل پیش پردازش داده‌ها معرفی می‌شوند. به منظور ارزیابی تأثیر به‌کارگیری روش خوشه‌سازی در بهبود عملکرد بویلر و با توجه به اینکه پیاده‌سازی نتایج حاصل در شرایط واقعی و روی بویلر در حال تولید بخار امکان‌پذیر نبود؛ مدل‌سازی بویلر مطرح می‌شود. از مجموعه روش‌های مرسوم برای مدل‌سازی و تخمین، ابتدا روش PLS به دلیل قابلیت ماهوی اش در کاهش دادن ابعاد پایگاه داده و تخمین پاسخ با استفاده از تعداد متغیرهایی کمتر از متغیرهای مورد مشاهده مورد توجه قرار گرفته است. سپس روش شبکه‌های عصبی به عنوان یک روش عمومی جعبه-سیاه و با قابلیت پوشش خوب ویژگی‌های غیرخطی مورد استفاده قرار می‌گیرد. در این بخش علاوه بر معرفی اجمالی شبکه‌های عصبی، با نگاهی نسبتاً عمیق به روش PLS پرداخته می‌شود و در پایان، نتایج مدل‌سازی راندمان بویلر در قالب مقدار دو متغیر دمای گاز خروجی و میزان اکسیژن موجود در آن با دو روش ارائه شده و مورد مقایسه قرار می‌گیرد.

در بخش دوم ابتدا دانش داده‌کاوی به تفصیل و به طور پایه‌ای مورد بحث قرار می‌گیرد. خانواده الگوریتم‌های خوشه‌سازی به طور کامل معرفی شده و الگوریتم‌های درخت تصمیم‌گیری و شناسایی قوانین متحد مورد اشاره قرار می‌گیرند. سپس الگوریتم مورد استفاده برای کنترل هوشمند معرفی شده و نتایج حاصل از به‌کارگیری آن در مجموعه داده آزمایش با استفاده از مدل معرفی شده در بخش قبل راستی‌آزمایی می‌شود.

فصل ۲- فرآیند صنعتی بویلر

۱-۲- معرفی

استفاده از قدرت بخار، پیشینه‌ای شاید به قدمت استفاده از آتش در تاریخ زندگی انسان دارد. اما استفاده صنعتی و کنترل‌شده از این قدرت نهفته در آب آتش دیده، به سالهای آغازین قرن هجدهم میلادی و اختراع اولین دیگهای بخار بازمی‌گردد [۱۱]. روند تکامل ساخت دیگهای بخار با به‌وجود آمدن صنعت نورد پیگیری شد و در سالهای پایان قرن هیجدهم با اختراع گاورنر توسط جیمز وات کنترل خودکار فرآیند بخار نیز متولد شد.

در روند تکامل تولید دیگهای بخار، ابتدا افزایش درجه حرارت، افزایش فشار و افزایش میزان بخار خروجی هدف قرار گرفت و سپس موضوعاتی مانند افزایش راندمان، سهولت کنترل و در نهایت کاهش هزینه‌های ساخت و بهره‌برداری و افزایش طول عمر بویلر مورد توجه قرار گرفت تا به تدریج انواع اولیه دیگهای بخار به بویلرهای امروزی تبدیل شدند. نمودار عملیاتی پایه دیگ بخار در شکل

(۱-۲) نشان داده شده است.



شکل (۱-۲): نمودار فرآیند بویلر

۲-۲- انواع بویلرها

بویلرها، منابع تولید بخاری هستند که در طیف گسترده‌ای از انواع کاربردها طراحی و تولید می‌شوند. بنا به کاربرد مورد نظر، انواع بسیار متفاوتی از بویلرها وجود دارد که از دیدگاهها مختلف قابل دسته‌بندی هستند.

۲-۲-۱- دسته بندی بویلرها بر حسب نوع سوخت

استفاده از سوخت‌های فسیلی برای تولید بخار، عوارض زیست محیطی بسیاری به دنبال دارد. علاوه بر این تکیه بر ذخایر سوخت زمین و استخراج بی‌رویه آن به اتمام این ذخایر خواهد انجامید. با توجه به راندمان بالای نیروگاههای هسته‌ای و همچنین توجه به اینکه نفت و گاز را می‌توان به جای استفاده به صورت سوخت برای تولید مواد پتروشیمی و بسیاری از فرآورده‌های صنعتی دیگر به کاربرد، در درازمدت بویلرهای با سوخت هسته‌ای در نیروگاهها جایگزین بویلرهای با سوخت فسیلی خواهند شد.

در تاسیسات نیروگاههای هسته‌ای یک سیال ثانویه بین راکتور و بویلر جریان می‌یابد و در بویلر حرارت از سیال ثانویه به آب منتقل می‌شود. این سیال در یک مدار بسته عمل کرده و می‌تواند یکی از مواد گاز CO_2 ، آب، آب سنگین، هلیوم گازی و یا سدیم مذاب باشد. پس از انتقال حرارت به آب تشکیل بخار، بخار حاصل برای استفاده به مولدهای برق انتقال داده می‌شود.

تولید بخار، علاوه بر کاربرد در صنعت برق در بسیاری از صنایع دیگر نیز قلب تپنده تولید است. در تمام بویلرهای صنعتی از سوخت فسیلی استفاده می‌شود. سوخت فسیلی که طراحی بویلر براساس آن انجام می‌شود؛ بنابر ملاحظات اقتصادی مجموعه تولیدی از یک یا چند نوع از انواع سوخت‌های جامد مانند ذغال سنگ، سوخته‌های مایع مانند نفت کوره و سوخت‌های گازی مانند گاز طبیعی انتخاب می‌شود.

۲-۲-۲- دسته بندی بویلرها بر حسب ساختمان و کاربرد

به طور عمومی و در یک دسته‌بندی کلی، بویلرها به دو گروه بویلرهای با فشار بالا و بویلرهای با فشار پایین تقسیم‌بندی می‌شوند [۱۱]. به بویلرهایی که در فشارت بیش از پانزده ^۱ psig کاری کنند بویلرهای با فشار بالا گفته می‌شود. با افزایش فشار کاری، به دلیل افزایش چگالی سیالات ابعاد بویلر کاهش پیدامی‌کند. بویلرهای با فشار پایین بیشتر برای کاربردهای تأسیساتی مورد استفاده قرار می‌گیرند.

۲-۲-۲-۱- بویلرهای با لوله آتش^۲

در بسیاری از واحدهای کوچک تا متوسط صنعتی از این دسته از بویلرها استفاده می‌شود. در این بویلرها گاز حاصل از احتراق در مسیرهای لوله‌ای حرکت می‌کند که با آب احاطه شده‌اند و به این

^۱ psig = 0.07 Bar

^۲ Fire Tube

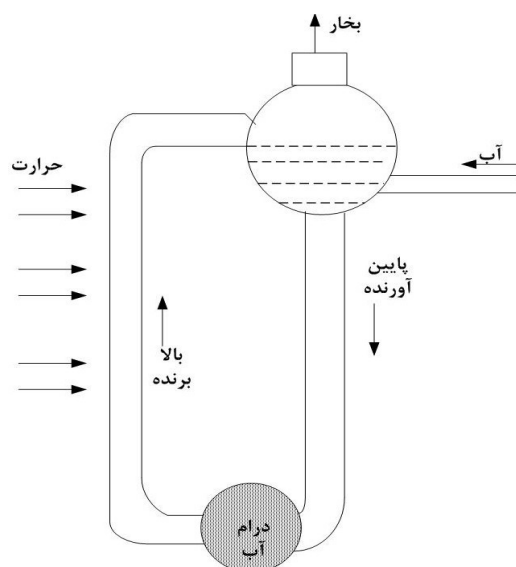
ترتیب تبادل حرارت صورت می‌پذیرد. بویلرهایی که در موتورخانه‌های ساختمانهای مسکونی مورد استفاده قرار می‌گیرند و یا نیروی محرکه در لوکوموتیوها را تامین می‌کنند از این دسته هستند.

۲-۲-۲-۲- بویلرهای با لوله آب^۱

همان‌گونه که اسم گویای آن است، برخلاف بویلرهای با لوله آتش، این بار آب در لوله‌ها جریان دارد. این لوله‌ها معمولاً بین دو مخزن استوانه‌ای شکل یا درام^۲ قرار می‌گیرند که درام پایینی درام آب و درام بالایی درام بخار نام دارد. گل و لای و رسوبات باقیمانده از تبخیر آب معمولاً در درام پایینی باقیمانده و از آن خارج می‌شود. در فضای کوره، حرارت به آب موجود در لوله‌ها منتقل شده و در ضمن چرخش آب در این مسیر بسته که به طور طبیعی و به دلیل کاهش جرم حجمی آب گرمتر و حرکت آن به سمت بالا و درام بخار انجام می‌شود، آب به مرور به بخار تبدیل شده و از فرآیند خارج می‌شود. این نوع از بویلرها برای طیف وسیعی از کاربردها، از فشار ۵۰ تا ۵۰۰۰ psi^۳ قابل طراحی و استفاده هستند. چرخش آب در یک بویلر با لوله آب به صورت نموداری در شکل (۲-۲) نشان داده شده است.

^۱ Water Tube

^۲ Drum



شکل (۲-۲): چرخش آب در مسیر تبخیر

۲-۲-۳- بویلرهای یکبارگذر^۱

در فشاری نزدیک به ۲۷۵۰ psig، حجم مخصوص آب و بخار بسیار نزدیک به هم است. در بویلرهایی که برای کار در این فشار طراحی می‌شوند، نیازی به وجود درام نیست و تمام آب تغذیه‌کننده موجود در لوله‌های دیواره کوره، پس از طی کردن محوطه احتراق به بخار تبدیل شده و از بویلر خارج می‌شود. در ساخت این نوع از بویلرها از فناوری پیشرفته‌تری استفاده می‌شود و کنترل آنها نیز در مقایسه با بویلرها دار دشوارتر است.

انواع دیگری از بویلرها نیز بر حسب نوع کاربرد در صنعت مورد استفاده قرار می‌گیرند که بویلرهای سیکل ترکیبی^۲، بویلرهای بازیافت^۳ و بویلرهای زباله‌سوز^۴ نمونه‌هایی از این بویلرها هستند.

¹ Once Through

² Heat Recovery

³ Recovery Boiler

⁴ Incinerator Boiler

در بویلرهای سیکل ترکیبی از انرژی گازهای خروجی توربین‌های گازی برای انتقال انرژی به آب استفاده شود. بویلرهای بازیافت در بخش کوره‌های ذوب فلزات مورد استفاده قرار می‌گیرند و بویلرهای زیاله‌سوز برای از بین بردن زیاله‌های شهری و استخراج انرژی از آنها به کار می‌روند.

۲-۲-۳ - اجزاء ساختمان بویلر

اجزاء ساختمان بویلر را معمولاً به دو دسته اجزاء تحت فشار^۱ و اجزاء غیر تحت فشار^۲ تقسیم بندی می‌کنند. اجزاء تحت فشار بخش‌هایی هستند که در داخل آنها آب یا بخار جریان دارد و فشار داخل آنها نسبت به محیط اطراف بسیار بیشتر است. به طور کلی مسیر آب از پمپ تغذیه آب بویلر تا خروجی بخار، مسیر اجزاء تحت فشار است. این مسیر شامل بخش‌های زیر است:

۱. لوله اصلی و پمپ آب تغذیه کننده: آبی که شرایط آن از نظر مواد معدنی محلول، اکسیژن

و PH به دقت کنترل شده است، توسط پمپ از مسیر خود به درون درام تزریق می‌شود.

۲. اکونومایزر^۳: تجهیز است شامل لوله‌های مار پیچ که آب تغذیه از درون آن عبور می‌کند.

این تجهیز در مسیر حرکت گاز خروجی از کوره به سمت دودکش قرار دارد و با عبور آب

از مجاورت گاز خروجی بخشی از دمای گاز به آب منتقل شده و باعث گرم شدن آن

¹ Pressure part

² Non-Pressure Part

³ Economizer

می‌شود. به این ترتیب، بخشی از حرارت که در راه خروج از کوره است توسط آب وارد شونده به درام جذب شده و از تلف شدن آن جلوگیری می‌شود.

۳. درام بخار و لوله‌ها: درام بخار مخزنی به شکل استوانه افقی است که در بالای کوره قرار می‌گیرد و جداسازی نهایی فاز بخار و فاز مایع آب در درام انجام می‌شود. دیواره کوره از لوله‌های اتصال دهنده درام آب به درام بخار پوشیده شده است. بخشی از این لوله‌ها بالابرنده آب و بخش دیگر پایین آورنده هستند که هر دو بخش در کنار هم مسیر چرخش طبیعی آب بین دو درام را فراهم می‌کنند.

۴. سوپرهیتر^۱ و دی سوپرهیتر^۲: بخار خروجی از درام، به تجهیز ساخته شده از لوله‌های مارپیچی که درون گرمترین بخش کوره قرار دارد وارد می‌شود و مجدداً تا رسیدن به فشار و دمای بالاتر در حین عبور از این مسیر مارپیچ در معرض حرارت قرار می‌گیرد. پس از خروج از سوپرهیتر، با توجه به اینکه دمای حاصل ممکن است بیش از دمای مورد نیاز طراحی شده باشد، بخار به دی سوپرهیتر وارد می‌شود. دی سوپرهیتر در واقع شیری است که برای کاهش دمای بخار، به صورت اسپری آب به درون مسیر آن می‌پاشد.

¹ Superheater

² DeSuperheater

علاوه بر اجزاء مسیر تحت فشار، مسیر ورودی هوای احتراق، فن دمنده هوا و دمپر کنترل کننده جریان هوا، کوره، مسیر گاز خروجی و دودکش اجزاء دیگر ساختمان بویلر را تشکیل می دهند. مجموعه این اجزاء مسیر حرکت هوا و دود را ایجاد می کند که به بخش غیر تحت فشار موسوم است. همچنین مسیر ورود سوخت به کوره و احتراق آن توسط لوله های سوخت، شیرهای کنترلی و حفاظتی و مشعل ها بسته می شود.

۲-۳- کنترل خودکار بویلر

۲-۳-۱- معرفی

فرآیند تبدیل آب به بخار دارای ثابت زمانی ای است که وابسته به مشخصات ساختار و نصب بویلر است. مشخصاتی از قبیل ذخیره حرارتی سیستم، ضریب انتقال حرارت در بخش های مختلف سیستم، میزان فلز به کار رفته و نحوه ساخت و چیدمان آرایه های فلزی و بسیاری از پارامترهای دیگر در تعیین مقدار این ثابت زمانی تعیین کننده هستند [۱۱]، اما در نهایت برای داشتن یک نگاه عمومی به کل مجموعه بویلر، می توان آن را به صورت یک سیستم درجه یک با ثابت زمانی ای از مرتبه دقیقه در نظر گرفت، هر چند که هر کدام از اجزاء آن، دینامیک و مشخصات خاص خود را دارا هستند.

در ادامه، به اجمال مروری بر کنترل کلاسیک بخش‌های مختلف این فرایند انجام می‌شود؛ با در نظر گرفتن این نکته که هر چند اصول حاکم کمابیش مشابه است، اما در راهبردهای کنترلی مختلف بر حسب نوع بویلر، کاربرد و میزان اهمیت فرآیند تفاوت‌هایی وجود دارد.

۲-۳-۲ - حلقه کنترل سوخت و هوا

این حلقه مهم‌ترین حلقه در تنظیم احتراق و در نهایت عملکرد کلی بویلر است. اصطلاح «بار»^۱ در بویلر به میزان دبی نهایی بخار خروجی گفته می‌شود که توسط واحد مصرف کننده بخار به فرآیند تشکیل بخار دیکته می‌شود. رابطه بین بار بویلر و میزان هوا و سوخت مصرفی تا حدود زیادی خطی است. علاوه بر این، فشار بخار خروجی عنصر اصلی کنترلی در رفتار ترمودینامیکی یک بویلر است^۲ و به عنوان نقطه تنظیم حلقه سوخت مورد استفاده قرار می‌گیرد.

در طراحی فرآیند احتراق یک بویلر، با در نظر گرفتن ارزش حرارتی ثابت برای سوخت، میزان هوای مورد نیاز برای احتراق از معادله شیمیایی سوخت قابل استخراج است، اما از آنجایی که خطر وقوع انفجار در هنگام برقراری نسبت معینی بین مقدار سوخت و هوا افزایش می‌یابد؛ به منظور افزایش ایمنی و فاصله گرفتن از نسبت ایجادکننده خطر انفجار، درصدی از مقدار هوا به مقدار استخراج شده از معادله شیمیایی سوخت اضافه می‌شود. میزان این هوا^۳ در زمان عملکرد بویلر در

¹ Load

² Boiler Master

³ Excess Air

بارهای کم که همراه با ایجاد شرایط ناپایداری در شعله و افزایش احتمال وقوع خطر است بیشتر از زمان عملکرد بویلر در بارهای زیاد در نظر گرفته می‌شود. علاوه بر این تدبیر و باز به منظور ایجاد ایمنی بیشتر، قاعده دیگری نیز بر حلقه تنظیم احتراق حاکم می‌شود که موسوم به هوا-افزونه^۱ است. با به کارگیری این قاعده نسبت سوخت و هوا همواره طوری تغییر می‌کند که میزان هوا بیشتر از میزان سوخت باشد یعنی در صورت افزایش تقاضا برای بار، ابتدای میزان هوا و سپس میزان سوخت اضافه می‌شود و در صورت کاهش بار، ابتدای میزان سوخت و سپس میزان هوا کم می‌شود.

۲-۳-۳ - حلقه کنترل آب و بخار

هدف این حلقه کنترلی، حفظ سطح مطلوب آب در درام بویلر با تغییرات بار است. به منظور افزایش سرعت در جبران تغییرات ارتفاع آب، این حلقه با اندازه‌گیری سه مقدار سطح فعلی درام، دبی آب ورودی و دبی بخار خروجی، برای شیر کنترل دبی آب ورودی فرمان صادر می‌کند. آرایش این حلقه معمولاً به صورت آبشاری^۲ است.

۲-۳-۴ - حلقه کنترل دمای بخار

در صورتیکه بخار خروجی از بویلر بیش از حد مورد نیاز گرم شده باشد، انرژی حرارتی اضافه می‌تواند باعث آسیب و سوختن بخش‌هایی از تجهیزات فرآیند تغذیه کننده از بخار شود. برای جلوگیری از این اتفاق و کنترل دقیق دمای بخار، بخار خارج شده از سوپرهیتر در دی‌سوپرهیتر

^۱ Air Rich

^۲ Cascade

خنک می‌شود. در دی‌سوپرهیتر بر حسب تفاوت دمای بخار با دمای مطلوب، در مورد میزان دبی آب خنک کننده تصمیم‌گیری شده و این مقدار آب به صورت اسپری به داخل مسیر حرکت بخار پاشیده می‌شود.

۲-۳-۵- سایر حلقه‌های کنترلی

سه حلقه یاد شده از مهمترین حلقه‌های کنترلی در فرآیندهای صنعتی تولید بخار هستند. اما در بسیاری از موارد با توجه به نوع فرآیند و ویژگی‌های آن، حلقه‌های کنترلی دیگری نیز مورد استفاده قرار می‌گیرند. به عنوان مثال برای کنترل سطح در مخازن سوخت یا مخازن آب از حلقه‌های ساده‌ای با یک عنصر کنترلی مثل شیر کنترل دبی و یک عنصر اندازه‌گیری مثل اندازه‌گیری دما یا سطح استفاده می‌شود.

در کنترل کلاسیک بویلر، مانند بسیاری از فرآیندهای صنعتی دیگر، کنترل کننده مورد استفاده از نوع PID است که تنظیم ضرایب آن معمولاً به صورت تجربی انجام می‌شود.

۲-۴- محاسبه راندمان بویلر

یکی از راه‌های تخمین میزان کارایی بویلر یا هر فرآیند تبادل انرژی، محاسبه تلفات گوناگون در آن فرآیند است [۱۲]. در بویلر گاز خروجی از فرآیند احتراق، میزان سوخت محترق نشده و تلفات ناشی از رطوبت و تشعشع از سطح کوره بخشی از این تلفات را نشان می‌دهند. بر این اساس، رابطه (۱-۲) یکی از پر استفاده ترین معیارها در محاسبه راندمان بویلر است.

$$BoilerEfficiency = 100 - [(L_G + L_{mf} + L_H + L_{mA} + L_{Co}) / H_u \times 100] - L_B \quad (1-2)$$

در این رابطه H_u حد پایین ارزش حرارتی سوخت را نشان می‌دهد. L_G نماینده تلف ناشی از خروج گاز حاصل از احتراق است که مهمترین عامل کاهنده از کارایی بویلر نیز می‌باشد. L_{mf} نشان دهنده تلف ناشی از رطوبت و ذرات آب موجود در ترکیبات سوخت است. L_H نماینده تلفات ناشی از ایجاد رطوبت بر اثر ترکیب هیدروژن سوخت با اکسیژن هوا و تشکیل آب در حین انجام فرآیند احتراق است. L_{mA} نماینده تلفات منبعت از رطوبت موجود در هواست، L_{Co} نشانگر تلفات ایجاد شده بر اثر ناقص سوزی و تشکیل مونوکسید کربن است و L_B به تلفات ناشی از تشعشع و انتقال حرارت از سطح کوره و سایر تاسیسات فرآیند اشاره می‌کند.

فصل ۳- بویلر نمونه و داده‌های آن

۳-۱- معرفی

بویلر نمونه مورد بحث در این پژوهش، یکی از شش بویلر ساخته شده توسط شرکت صنایع آذرآب است که در محل مجتمع پتروشیمی مارون واقع در منطقه ویژه اقتصادی پتروشیمی بندرماهشهر قرار دارد و با تولید ۹۰ تن بخار در ساعت، بخشی از نیاز واحد الفین این مجتمع به بخار را تأمین می‌کند.

این بویلر از نوع بویلرهای با لوله آب و مجهز به دو درام آب و بخار بوده که مشخصات آن به شرح زیر است:

مشخصات بخار: فشار ۴۲ Bar در دمای ۴۳۰ درجه سانتیگراد

سوخت: گاز طبیعی

نرخ تغییر بار پیش‌بینی شده: ۴/۵ تن در دقیقه

سیستم حلقه‌های کنترلی و منطق حفاظتی^۱ بویلر استفاده از DCS ساخت شرکت امرسون^۲ پیاده-

سازی شده است. نسخه نرم‌افزار مانیتورینگ و بایگانی اطلاعات فرآیند مورد استفاده در این DCS، از

^۱ Interlock

^۲ Emerson

نسخه‌های نسبتاً قدیمی این نرم‌افزار بوده که امکان انتقال و تبدیل اطلاعات از پایگاه داده بایگانی آن به صورت فایل‌های عمومی قابل استفاده مانند `.txt` , `.xls` وجود ندارد. بنابراین، ناگزیر اطلاعات ثبت شده به صورت دست‌نویس مورد استفاده قرار گرفتند. این اطلاعات شامل مقادیر پانزده پارامتر اصلی عملیاتی بویلر بودند که در بازه‌های زمانی دو ساعته، به تعداد ۱۲ نمونه در روز توسط اپراتورهای مسئول بهره‌برداری ثبت شده بودند. نمونه‌ای از صفحات این روند گزارش‌گیری در پیوست (۱) ارائه شده است.

متغیرهای ثبت شده در هر نمونه‌برداری و رنج تغییرات آنها عبارت‌اند از:

- دمای سوخت ($90^{\circ}-10$)، دبی سوخت ($7500 \frac{kg}{hr}-0$)، فشار سوخت ($4\text{ Bar}-0$)، دبی هوای

احتراق ($120 \frac{T}{hr}-0$)، دمای هوای احتراق ($90^{\circ}C-0$)، فشار کوه (300 mmHg_2O-0)، دمای گاز

خروجی ($500^{\circ}C-0$)، فشار آب تغذیه ($100\text{ Bar}-0$)، دبی آب تغذیه ($120 \frac{T}{h}-0$)، اکسیژن موجود

در گاز خروجی (%/.) سطح درام ($50 \sim -50\text{ mm}$)، فشار درام ($49\text{ Bar}-42$)، دبی بخار

خروجی ($120 \frac{T}{hr}-0$)، فشار بخار خروجی ($44\text{ Bar}-42$)، دمای بخار خروجی ($435-425^{\circ}C$)

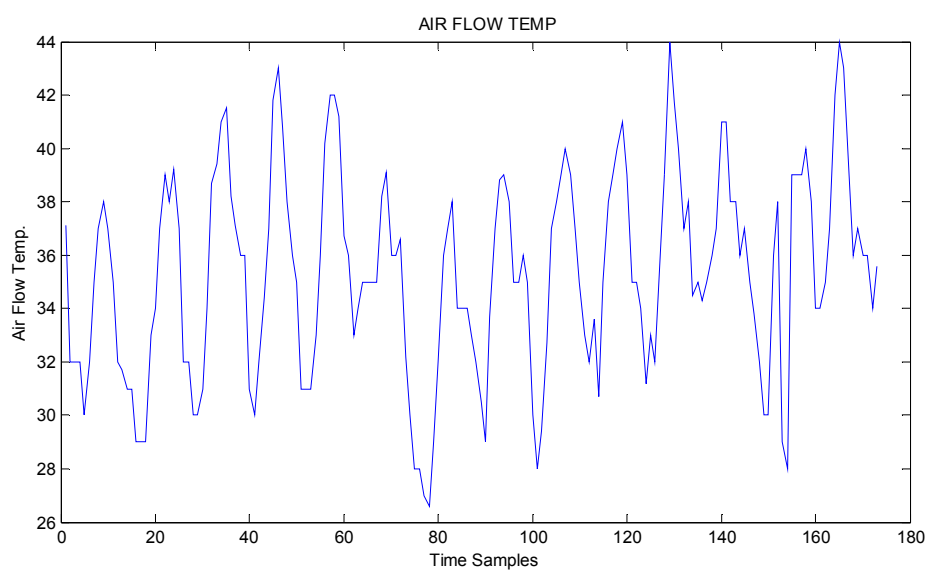
۳-۲- اعتبارسنجی و پیش پردازش داده‌ها

با توجه به اینکه مجموعه داده‌های در دسترس از دوره‌های مختلف اطلاعات ثبت شده به صورت دست‌نویس جمع‌آوری شده بودند، اولین گام اعتبارسنجی آنها بود. به منظور ارزیابی صحت اطلاعات، دو محک زیر مورد استفاده قرار گرفت:

۱- رسم رویه حاصل از مقادیر ثبت شده تعدادی از متغیرهای اندازه‌گیری شده در یک بازه زمانی پیوسته و بررسی تغییرات آن در طول زمان

۲- محاسبه ماتریس ضرایب همبستگی بین مقادیر ثبت شده و تحلیل همخوانی ماتریس حاصل با واقعیت‌های فیزیکی مورد انتظار

در اولین آزمون، نمودار تغییرات دمای هوای احتراق به عنوان معیاری از دمای هوای احتراق به عنوان معیاری از دمای هوای محیط رسم شد که این نمودار در شکل (۳-۱) مشاهده می‌شود. با توجه به بازه زمانی اختیار شده برای آزمایش اعتبار داده‌ها که یک بازه یک ماهه از اواخر اسفند تا اواخر فروردین بود؛ علاوه بر نوسانات به دامنه مقادیر در طول روز و شب، نمودار شیب ملایمی به سمت افزایش دامنه مقادیر داشت که حاکی از گرم شدن تدریجی هوا در گذشت روزهای اسفند و آغاز فرودین بود.



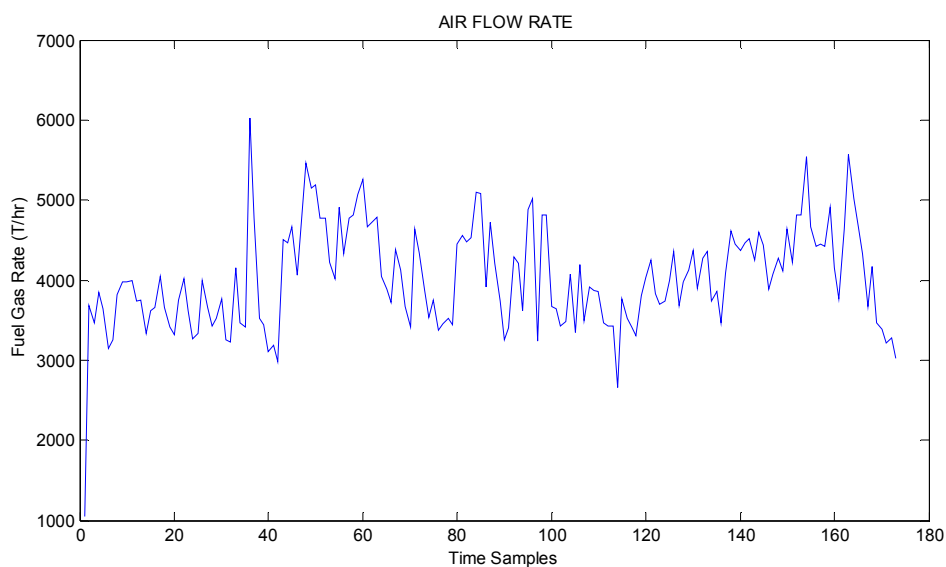
شکل (۳-۱): نمودار تغییرات دمای هوا

در آزمون بعدی، دو متغیر دبی هوای احتراق و دبی سوخت مورد بررسی قرار گرفتند. نمودار تغییرات

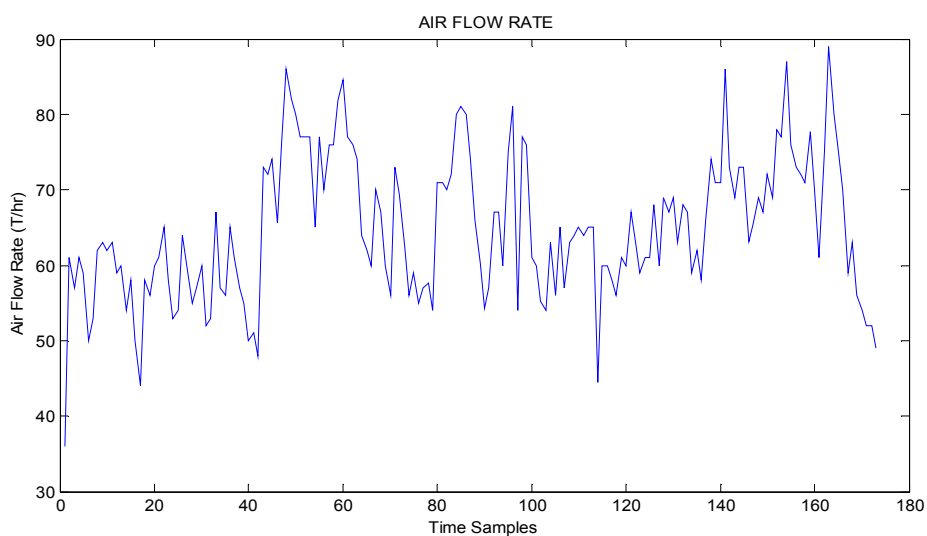
این دو متغیر در شکل (۳-۲) نمایش داده شده است. با توجه به اینکه رابطه این دو متغیر با هم تا حدود

بسیار زیادی خطی و مستقیم است، انتظاری رود روند تغییرات در هر شکل مشابهی داشته باشد که

این موضوع نیز با مقایسه دو بخش شکل (۳-۲) قابل مشاهده است.



شکل (۲-۳) الف: دبی سوخت



شکل (۲-۳) ب: دبی هوا

آزمون سوم محاسبه مقادیر ضرایب همبستگی ۱۵ متغیر در داده‌های مشاهده شده بود. این ماتریس در

پیوست (۲) نمایش داده شده است. مقادیر همبستگی بین متغیرها، اعم از میزان همبستگی کم یا

میزان همبستگی زیاد، با واقعیت‌های فیزیکی مورد انتظار همخوانی دارد. به عنوان مثال بین دو متغیر

فشار دردام و فشار آب تغذیه، ضریب همبستگی معنی‌دار و زیاد است.

پس از حصول اطمینان نسبی از صحت داده‌های ثبت شده نوبت به پیش پردازش داده‌ها می‌رسد. پیش پردازش گام آغازین بسیار مهمی در تمام مسائل مرتبط با آنالیز داده، از جمله روش‌های آمار و داده‌کاوی است که ضوابط و الزامات آن بسته به نوع مسأله متفاوت است. اما به طور عمومی شامل مراحل ماندن نویززدایی، حذف داده‌های نامعتبر و جایگزینی داده‌های از دست رفته می‌شود.

۳-۲-۱- داده‌های ناخوانا یا ثبت نشده

با توجه به اینکه داده‌ها به صورت دستی ثبت شده بودند، در موارد بسیار کمی به خاطر شتابزدگی در ثبت و یا دست‌خط اپراتور مقادیر قابل بازیابی نبودند. در این موارد یکی از دو راه حل زیر با توجه به شرایط اتخاذ شد:

- هر نمونه زمانی برداری شامل ۱۵ درایه است. به عنوان مثال در شرایطی که مقدار یکی از درایه‌ها ناخوانا بود، اگر مقدار سایر درایه‌های آن بردار بین دو مقدار درایه‌های متناظر بردار قبلی و بعدی قرار داشت، مقدار درایه ناخوانا با مقدار میانگین درایه متناظر در بردار قبلی و بعدی جایگزین می‌شد.

- در شرایطی که مقادیر سایر درایه‌ها در بردار حاوی درایه ناخوانا یا ثبت نشده بین مقادیر دو نمونه زمانی قبلی و بعدی قرار نداشت، کل بردار از مجموعه مورد مشاهده حذف شد.

۳-۲-۲ - نقاط پرتاب^۱

برای محاسبه نقاط پرتاب از تعریف رایج آماری نقاط پرتاب به عنوان معیار استفاده شد [۱۳]. طبق این تعریف در یک مجموعه داده، نقاطی که فاصله آنها با مقدار میانگین مجموعه بیش از ۱/۵ برابر فاصله بین ۲۵٪ تا ۷۵٪ رنج تغییرات^۲ باشد؛ نقطه پرتاب محسوب می‌شوند. بر اساس یک قاعده سرانگشتی برای حفظ اطلاعات موجود و پیوستگی مجموعه؛ در مجموعه داده شده در یک کاربرد صنعتی، مقادیر جایگزین و یا حذف شده نباید از ۱۰٪ تعداد کل مقادیر بیشتر باشند.

۳-۲-۳ - نرمالیزه کردن داده‌ها^۳

با توجه به اینکه ۱۵ متغیر مورد اندازه‌گیری کمیت‌هایی با ماهیت‌های متفاوت و رنج و واحدهای گوناگون هستند، به منظور استفاده شدن در کنار هم برای کاربردهایی مانند محاسبه و مقایسه فاصله‌ها و یا در ورودی الگوریتم تعلیم شبکه عصبی برای جلوگیری از اشباع شدن شبکه باید نرمالیزه شوند.

به منظور نرمالیزه کردن داده‌ها، در خصوص هر متغیر از مجموعه $x_1 - x_{15}$ از رابطه (۳-۱) استفاده می‌شود که در آن x_{\min} کمترین مقدار ثبت شده از متغیر و x_{span} فاصله بین کمترین و بیشترین مقدار ثبت شده از متغیر است.

¹ Outliers

² Interquartile

³ Normalization

$$x_{Normalized} = \frac{x - x_{\min}}{x_{span}} \quad (1-3)$$

فصل ۴ - مدل سازی و تخمین توابع

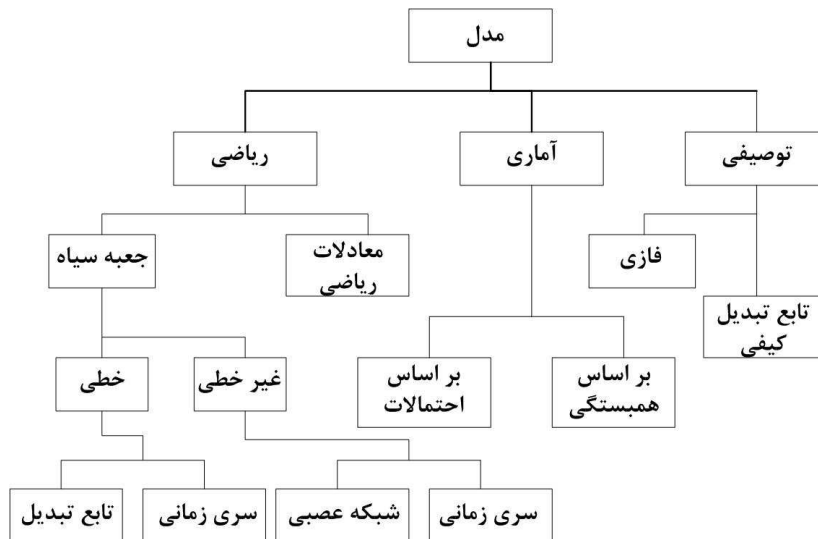
در کنار اهداف متفاوتی که از ساختن مدل یا تخمین زدن پاسخ در مواجهه با فرآیندها دنبال می‌شود، روش‌های متفاوتی نیز پایه گذاری شده و مورد توجه قرار گرفته اند. هر روش متناسب با هدف و شرایط خاص مسأله مورد استفاده قرار می‌گیرد. یک تقسیم بندی کلی از روشهای مطرح برای مدل سازی در تصویر (۴-۱) نمایش داده شده است. به طور عمومی برای شناسایی مدل یک سیستم پیمودن مراحل زیر در تمام روشها انجام می‌شود:

۱- تعیین ساختار مشخص و پارامتری شده از روابط ریاضی برای سیستمی که هدف تعیین مدل آن است.

۲- تعیین پارامترها به طوریکه ساختار حاصل به بهترین شکل با اعضای مجموعه آموزش همخوانی داشته باشد.

۳- راستی آزمایی مدل، با سنجش پاسخ مدل به داده‌ها

۴- تکرار مراحل فوق تا رسیدن به نتیجه مطلوب و در غیر اینصورت تغییر ساختار انتخاب شده در گام اول.



شکل (۴-۱): دسته بندی انواع مدلها

با توجه به قابلیت عمومی شبکه عصبی در تخمین توابع غیرخطی تا درجه مطلوبی از دقت، این روشها در کاربردهای بسیاری از شناسایی و مدل سازی سیستمها مورد استفاده قرار می گیرند. همچنین تکیه بر بایگانی اطلاعات فرآیند یا مدل سازی آماری آن، رهیافتی است که روز به روز بر اهمیت آن افزوده می شود. در این پژوهش مجموعه ای از ۱۱۳ نمونه ثبت شده از عملکرد بویلر مورد بحث برای مدل سازی مورد استفاده قرار گرفته است. این مجموعه، مجموعه ای از داده هاست که در آینده در فرآیند آزمایش الگوریتم کنترل هوشمند به عنوان مجموعه داده عملکرد جاری بویلر مورد استفاده قرار می گیرد. در فصلهای آتی دو رهیافت مدل سازی آماری و مدل سازی جعبه- سیاه با استفاده از شبکه های عصبی ضمن معرفی در مدل سازی فرآیند احتراق بویلر مورد استفاده قرار می گیرند.

فصل ۵- مدل سازی آماری و کاوش داده‌ها

۵-۱- معرفی

با توجه به ماهیت مسأله مورد بررسی که بر استفاده از داده‌های واقعی فرآیند برای بررسی رفتار و کنترل آن تأکید دارد، به موازات روش شبکه‌عصبی از روش آماری رگرسیون جزئی حداقل مربعات^۱ یا PLS نیز برای مدلسازی فرآیند بویلر استفاده شده است.

PLS یک روش مشهور برای مدلسازی نرم در کاربردهای مختلف است [۱۴]؛ به ویژه وقتی مسأله موردنظر با حجم زیادی از داده‌های اندازه‌گیری شده با رابطه‌های نامعلوم درگیر بوده و هدف ساختن یک مدل پیش‌بین است. در شرایطی که تعداد داده‌ها بسیار زیاد باشد؛ اتفاق رایج در فرآیند مدل‌سازی، بیش‌برازش^۲ شدن مدل حاصل است، به این معنی که این مدل تنها برای توصیف رفتار مجموعه داده‌هایی که با آن تعلیم داده شده است مناسب است. در چنین شرایطی، راه‌حل جستجو برای یافتن داده‌هایی است که هدایت‌کننده بخش عمده تغییرات در پاسخ سیستم هستند. در واقع هدف از استفاده PLS جستجو برای رابطه‌های پنهان^۳ بین داده‌ها و کاهش ابعاد

^۱ Partial Least Square Regression

^۲ Over fitting

^۳ Latent

مسئله بر پایه به کارگیری این رابطه‌ها است. از این رو از این روش با عنوان "Projection to Latent Structure" نیز یاد می‌شود.

روش PLS در دهه ۱۹۶۰ توسط هرمن ولد^۱ به عنوان یک روش برای انجام سنجش‌های محیطی^۲ پایه‌گذاری شد [۱۴]. پس از آن استفاده از این روش در شاخه‌های محاسباتی رشته شیمی و علوم اجتماعی نیز گسترش پیدا کرد و در حال حاضر نیز به عنوان یک روش مدل‌سازی در حال راه یافتن به کاربردهای صنعتی است.

برای بررسی روش PLS در ادامه ابتدا مروری بر رگرسیون خطی و آنالیز اجزاء اساسی^۳ انجام می‌شود و سپس روش PLS و الگوریتم مورد استفاده مورد بحث قرار می‌گیرد.

۵-۲- رگرسیون خطی چندگانه

در روش معروف آماری رگرسیون خطی، n نمونه از مقادیر مشاهده متغیر وابسته y به منظور مقادیر ثابت و معلوم x_1, \dots, x_n برای برآورد پارامترهای β_0 و β_1 در رابطه (۵-۱) مورد استفاده قرار می‌گیرند و به این ترتیب رفتار سیستم تولیدکننده مقادیر y با یک رابطه خطی تقریب زده می‌شود.

$$\hat{y}_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , \quad i = 1, 2, \dots, n \quad (1-5)$$

¹ Herman Wold

² Econometric

³ Principal Component Analysis

β_0 و β_1 از روش کمترین مربعات محاسبه می‌شوند. یعنی مقدار β_0 و β_1 به نحوی محاسبه می‌شود

که $\hat{\varepsilon}\hat{\varepsilon}^T$ یا جمع مربعات خطا کمینه باشد.

در رگرسیون چندگانه نیز هدف پیش‌بینی متغیر پاسخ Y بر مبنای یک رابطه خطی با چند متغیر

مستقل X_1, X_2, \dots, X_k از روی مجموعه داده‌ای به حجم n از مقادیر مشاهده شده از Y ها و

X هاست.

مشابه رگرسیون ساده، پارامترهای β_i در رابطه (۲-۵) باید طوری تعیین شوند که جمع مربعات

خطا کمینه باشد.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_n & & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (2-5)$$

ماتریس β ، ماتریس پارامترهای موسوم به ضرایب رگرسیون جزئی است. برای پیدا کردن بهترین

برازش بر حسب معیار کمترین مربعات خطا، مشتق $\sum_i \hat{\varepsilon}_i^2$ نسبت به هر $\hat{\beta}_j$ محاسبه می‌شود و

نتایج مساوی صفر قرار داده می‌شوند. از حل توأم $k+1$ معادله حاصل، رابطه (۳-۵) برای محاسبه

ضرایب مدل سیستم حاصل می‌شود. در صورتیکه X با مرتبه کامل باشد، X_X^T ناویژه بوده و $\hat{\beta}$ قابل

محاسبه است. در رابطه (۳-۵)، $\hat{\varepsilon}$ مقدار مانده مدل است و می‌توان از آن برای کنترل اعتبار مدل

استفاده کرد.

برای بررسی قاعده‌ای که روش PLS بر مبنای آن طراحی شده است، حالتی را در نظر می‌گیریم که تنها دو متغیر مستقل X_1, X_2 برای تخمین y در اختیار هستند. اگر X_1 و X_2 متعامد باشند، برآوردگر β_1 در مدل کامل بیان شده در رابطه (۳-۵) با برآوردگر β_1^* در مدل رابطه (۴-۵) برابر است [۱۳].

$$y = x_1 B_1 + x_2 B_2 + \varepsilon \quad (3-5)$$

$$y = x_1 B_1^* + \varepsilon^* \quad (4-5)$$

قضیه فوق اساس کار روش PLS است که در آن برحسب میزان خاصیت هم‌خطی^۱ موجود بین متغیرهای مستقل درمورد دخالت دادن آنها در مدل تصمیم‌گیری می‌شود. به این ترتیب که رگرسیون ساده y بر X_1 تنظیم شده و مقدار مانده آن محاسبه می‌شود. برای اضافه کردن متغیر X_2 به معادله پیش‌بینی که فقط شامل X_1 است، باید تعیین کنیم که چه میزان از تغییرات (واریانس) y بعد از محاسبه اثر X_1 مربوط به X_2 است و بر این اساس رابطه میان دو متغیر X_1 و X_2 را اصلاح کنیم. این کار با بررسی تغییرات مانده بعد از رگرسیون y بر X_1 تغییرات مانده بعد از رگرسیون X_2 بر X_1 انجام می‌شود.

¹ Co – linearity

۵-۳- آنالیز اجزاء اساسی

۵-۳-۱- معرفی

ایده اصلی استفاده از PCA، کاهش دادن ابعاد یک مجموعه داده‌ای بزرگ است که متغیرهای در هم تنیده دارد. هدف استفاده از این روش تحلیل، کاهش ابعاد مجموعه داده تا حد امکان و در ضمن حفظ اطلاعات موجود در تغییرات ثبت شده در مقدار متغیرهاست. بنا بر این، نتیجه استفاده از آنالیز اجزاء اساسی تنظیم مجموعه جدیدی از متغیرهاست که اولاً تا حد امکان "غیروابسته" بوده و ثانیاً "مرتب" شده هستند؛ به این معنی که بازتاب رفتار تمام متغیرهای قبلی در چند متغیر جدید اول نهفته است.

PCA، که استفاده از آن به دهه ۱۹۹۰ و دانشمند معروف حوزه آمار، پیرسون، بازمی‌گردد؛ بر محاسبه واریانس تمرکز دارد. قدم اول، جستجو برای یک تابع خطی به صورت $\alpha_1^T X$ است، در حالی که برداری از P متغیر تصادفی و α_1 برداری از P ثابت $\alpha_{11} \dots \alpha_{1p}$ می‌باشد (رابطه ۵-۶).

$$\sum_{j=1}^p \alpha_{1j} X_j = \alpha_{11} x_1 + \dots + \alpha_{1p} x_p = \alpha_1^T X \quad (۵-۶)$$

در گام بعدی هدف جستجوی تابع خطی دوم به صورت $\alpha_2^T X$ است که مشابه تابع اول تعریف شده، نسبت به تابع اول مستقل بوده و دارای بیشترین میزان واریانس ممکن است. این فرآیند تا گام K ادامه پیدامی‌کند که در آن تابع خطی K ام غیر وابسته با $K - 1$ تابع قبلی بوده و

حداکثر واریانس را داراست. تابع K ام K ، امین جزء اساسی است. تعداد گامها و توابع تا P قابل

افزایش است، به این معنی که مجموعه جدید می تواند تا هم بعد شدن با مجموعه قدیمی رشد پیدا

کند، هر چند که هدف مطلوب نمایش بیشترین تغییرات از طریق همان n جزء اساسی اول، به

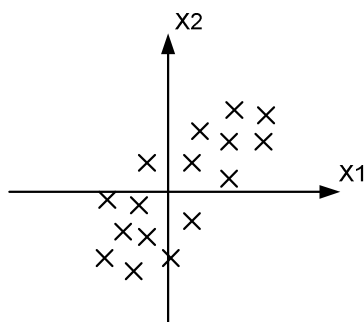
صورتی که $n \ll P$ باشد است.

به عنوان یک مثال برای ملموس شدن موضوع، در صورتی که توزیع شکل (۵-۱) برای دو متغیر

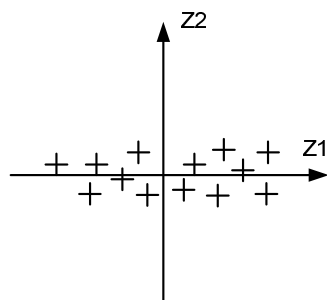
X_1 و X_2 وجود داشته باشد و دو PC متناظر محاسبه شده Z_1 و Z_2 در شکل (۵-۲) باشند،

مشاهده می شود که بیشترین تغییرات در امتداد تنها یک PC یعنی Z_1 اتفاق می افتد. بنابراین Z_1

می تواند نماینده رفتار مجموعه دو متغیر X_1 و X_2 باشد.



شکل (۵-۱): متغیرهای اصلی



شکل (۵-۲): اجزاء اساسی

۵-۳-۲ - استخراج اجزاء اساسی

روش استخراج اجزاء اساسی بر مبنای محاسبه ماتریس کوواریانس است. اگر ماتریس کوواریانس بردار X با نام Σ تعریف شود، درایه (i, j) آن کوواریانس بین درایه i ام و j ام بردار X بوده و وقتی $i = j$ ، این مقدار تبدیل به مقدار واریانس می‌شود. در صورتی که بردار ویژه ماتریس Σ و متناظر با بیشترین مقدار ویژه K ام، یعنی λ_K باشد، K امین جزء اساسی به صورت رابطه (۷-۵) تعریف می‌شود:

$$Z_K = \alpha_K^T X \quad (7-5)$$

و در رابطه فوق در صورتیکه بردار ویژه به صورتی انتخاب شده باشد که دارای اندازه واحد بوده، یعنی $\alpha_K^T \alpha_K = 1$ برابر یک باشد، خواهیم داشت:

$$\text{Var}(Z_K) = \lambda_K \quad (8-5)$$

۵-۴ - رگرسیون جزئی حداقل مربعات

روش رگرسیون جزئی حداقل مربعات یا PLS در واقع ترکیب کننده ویژگی‌های دو روش PCA و رگرسیون چندگانه است، در حالیکه از محدودیت‌های هر دو روش نیز عبور می‌کند.

در روش PCA، عوامل نهفته در متغیرهای X و Y ، از ماتریس‌های X و Y به طور جداگانه استخراج می‌شوند. یعنی ماتریس کوواریانس X و ماتریس کوواریانس Y هر کدام جداگانه برای استخراج عوامل پنهان به کار می‌روند و هیچگاه از اثر متقابل متغیرهای وابسته و مستقل برای انتخاب اجزاء

اساسی استفاده نمی‌شود، در حالیکه حاصل ضرب نقطه‌ای دو بردار X و Y در الگوریتم محاسبه

روابط پنهان روش PLS مورد استفاده قرار می‌گیرد.

در روش رگرسیون چندگانه نیز محاسبه پارامترهای β در شرایط رتبه کامل بودن ماتریس X

امکان پذیرند و در صورتیکه تعداد متغیرها در مقایسه با تعداد مشاهدات بیشتر باشد، ماتریس X به

سمت تکینی پیش می‌رود و امکان محاسبه ضرایب محدود می‌شود.

PLS ضمن غلبه بر دو محدودیت فوق، رگرسیون چندگانه را روی متغیرهایی از مجموعه X

انجام می‌دهد که بیشترین میزان کوواریانس با متغیرهای پاسخ را دارا هستند. بنابراین از جهت

داشتن قابلیت کاهش دادن ابعاد مجموعه‌های داده‌ای نیز به نوعی ابزاری از گروه ابزارهای دانش

داده‌کاوی محسوب می‌شود.

الگوریتم PLS متغیرهای متعدد موجود در هر نمونه مشاهده شده $X = (X_1, X_2, \dots, X_k)$ را به

روی برداری مانند بردار T با تعداد کمتری از عناصر تصویر می‌کند، یعنی $T = (T_1, T_2, \dots, T_n)$

به طوریکه $A < k$. این نگاهت به صورت $T = X\omega$ تعریف می‌شود که در آن ω_i ها بردارهای

بارگذار^۱ و T بردار امتیاز^۲ نامیده می‌شود. A تعداد پدیده‌ها، یا حداکثر رابطه‌های پنهان موجود

¹ Loading Vector

² Score

مورد انتظار است. A در واقع به تعداد بخش‌هایی از فضای متغیرهایی x که متناظر با وقوع واریانس زیاد در فضای پاسخ y هستند، اشاره می‌کند.

$$y = TQ + e \quad (14-5)$$

$$y = x\beta + e \quad (15-5)$$

$$\beta = \omega Q \quad (16-5)$$

پس از محاسبه T، بردار y به صورت رابطه (14-5) معرفی می‌شود که در مقایسه با رابطه اصلی رگرسیون، یعنی رابطه (15-5) مشاهده می‌شود که β یا ضرایب رگرسیون از رابطه (16-5) حاصل می‌شوند. در این رابطه ω بردار بارگذار در نگاشت معرفی شده T است.

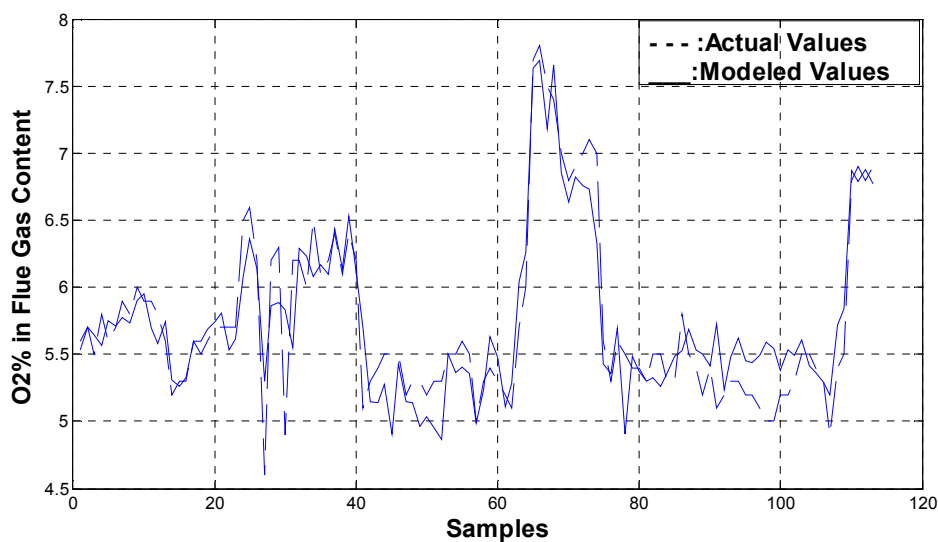
براساس مفاهیم فوق، خانواده‌ای از الگوریتم‌های PLS به وجود آمده‌اند که هر چند از دیدگاه تئوریک تا حدود زیادی یکسان هستند، اما با توجه به ویژگی‌های مجموعه داده یا تعداد فاکتورها ممکن است به نتایج متفاوتی منجر شوند. از میان اعضای این خانواده، دو الگوریتم نیپالس¹ و سیمپلس² شهرت زیادی دارند که در صورتیکه متغیر پاسخ y تک بعدی باشد استفاده از هر کدام از این دو الگوریتم تفاوت چندانی در نتیجه حاصل ایجاد نمی‌کند [15].

¹ NIPALS

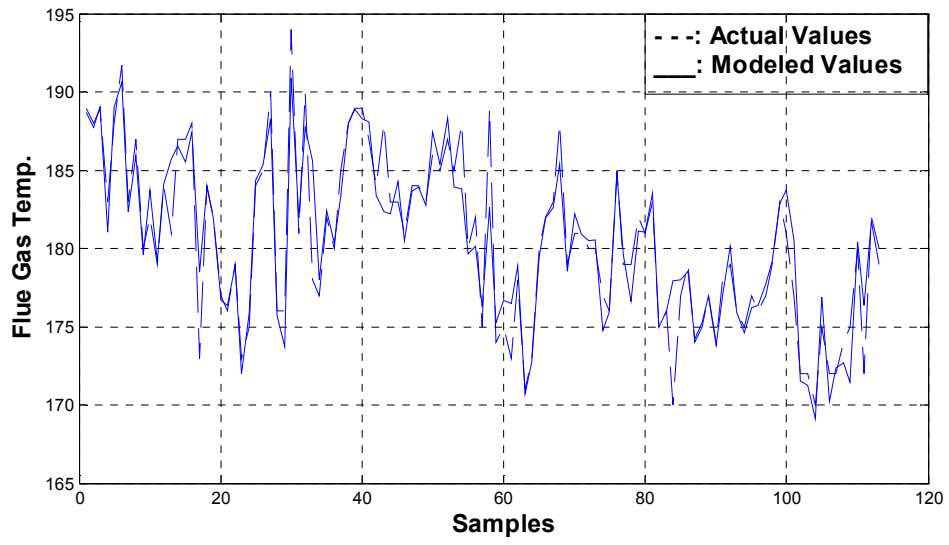
² SIMPLS

۵-۵- نتایج مدل‌سازی به روش PLS

نتایج مدل‌سازی دمای گاز خروجی از فرآیند احتراق و میزان اکسیژن موجود در آن بر اساس روش PLS در شکل (۳-۵) و (۴-۵) نمایش داده شده است. میانگین قدر مطلق خطا در تخمین دمای گاز خروجی ۱/۲۰۳ درجه سانتیگراد است، در حالیکه دامنه تغییرات دمای گاز خروجی بین مقدار کمینه و بیشینه خود ۲۴ درجه سانتیگراد بوده است. همچنین در مدل‌سازی میزان اکسیژن موجود در گاز خروجی میانگین قدر مطلق خطا ۰/۲٪ است؛ در حالیکه دامنه تغییرات ۳/۲٪ می‌باشد.



شکل (۳-۵): مدل اکسیژن در گاز خروجی با روش PLS



شکل (۴-۵): مدل دمای گاز خروجی با استفاده از روش PLS

فصل ۶- شبکه‌های عصبی

۶-۱- معرفی

شبکه‌های عصبی ساختاری است که براساس آن رابطه‌ی کلی بین ورودی‌ها و خروجی‌ها بر حسب مجموعه‌ای از پارامترهای قابل تغییر تعیین می‌شود [۱۶]. این ساختار شامل یک مجموعه از گره‌ها^۱ و اتصالات^۲ است که هر گره، یک تابع را روی مقدار ورودی خود اعمال کرده و خروجی را تولید می‌کند. در شبکه‌های عصبی ایستا هیچ دینامیک و یا حالت داخلی تأثیرگذاری وجود ندارد و خروجی هر گره تنها به مقدار فعلی ورودی‌اش وابسته است.

شبکه‌های عصبی از دیدگاه‌های مختلف مورد تقسیم‌بندی قرار می‌گیرند، از نظر الگوریتم یادگیری به دو نوع باسرپرست^۳ و بدون سرپرست^۴ تقسیم می‌شوند و از دیدگاه معماری شامل دو دسته پیش‌خور^۵ و بازگشتی^۶ هستند.

^۱ Node

^۲ Link

^۳ Supervised

^۴ Unsupervised

^۵ Feedforward

^۶ Recurrent

استفاده از شبکه‌های عصبی در مقوله مدل‌سازی، شناسایی سیستم‌ها و تخمین توابع با به‌کارگیری الگوریتم‌های باسرپرست امکان‌پذیر می‌شود، زیرا در این روشها هدف نهایی تعیین یک نگاشت غیرخطی بین ورودی‌ها و خروجی‌های موردنظر است.

از دیدگاه معماری شبکه، در شبکه‌های پیش‌خور خروجی هر گره از سمت ورودی به سمت خروجی شبکه حرکت می‌کند. در صورتیکه در مسیر گره‌ها اتصالات پس‌خوردی وجود داشته باشد شبکه به یک شبکه بازگشتی تبدیل می‌شود. از دیدگاه نظریه گراف در یک شبکه پیش‌خور هیچ‌گاه حلقه‌ای تشکیل نمی‌شود در حالی که یک شبکه بازگشتی حداقل یک حلقه دارد.

نگاشت غیرخطی حاصل از به‌کارگیری شبکه عصبی، مجموعه داده تشکیل شده از زوج‌های ورودی - خروجی را به سیستم مدل‌سازی شده هدف تبدیل می‌کند. این مجموعه داده موسوم به مجموعه داده^۱ تعلیم است. برای تنظیم پارامترها یا به عبارتی آموزش دادن شبکه برای عملکرد بهتر، از رویه‌های متفاوتی موسوم به الگوریتم یادگیری استفاده می‌شود. همچنین عملکرد شبکه معمولاً به صورت تفاوت بین خروجی موجود و خروجی هدف سنجیده می‌شود که این تفاوت موسوم به مقدار خطای شبکه است.

¹ Training Data set

۶-۲- شبکه پرسپترون

اولین بار در اواخر دهه ۱۹۵۰، روزنبلات^۱ از پرسپترون تک لایه استفاده کرد. بعد از مشخص شدن محدودیت‌های شبکه‌عصبی تک لایه در دهه ۱۹۷۰ اقبال علمی نسبت به شبکه‌های عصبی بسیار کمرنگ شد، تا زمانی که با معرفی الگوریتم‌های جدیدتر و برطرف شدن محدودیت‌های مطرح شده، این روش‌ها دوباره مورد توجه قرار گرفتند. در بحث شبکه‌های عصبی، شبکه پرسپترون ابتدایی‌ترین ساختاری بود که با الهام از عملکرد نورون‌های مغز استخوان پایه‌گذاری شد. الگوی کلی یادگیری این شبکه در دو گام زیر خلاصه می‌شود [۱۷]:

۱- یک داده از اعضای مجموعه داده‌های تعلیم به شبکه ارائه می‌شود.

۲- در صورتیکه پاسخ شبکه به داده ارائه شده نادرست بود، وزن هر مسیر ارتباطی براساس رابطه (۱-۶) تازه‌سازی می‌شود.

$$\Delta \omega_i = \mu t_i x_i \quad (1-6)$$

در رابطه فوق x_i ورودی مسیر، t_i مقدار هدف و μ گام یادگیری است.

در یک الگوریتم پس انتشار خطا، گره‌ها یا سلول‌های عصبی تابع یکسانی را بر روی ورودی‌های خود اعمال می‌کنند. این تابع ترکیبی است از یک جمع وزن‌دهی نشده و یک تابع غیرخطی مشتق‌پذیر به عنوان تابع فعال‌ساز که به تابع انتقال نیز معروف است.

¹ Rouzenblot

الگوریتم پس انتشار خطا برای هر دو ساختار پیش خور و بازگشتی قابل استفاده است. در واقع روش پس انتشار خطا، نحوه استفاده از جهت گرادیان برای تنظیم وزن‌ها را نشان می‌دهد.

۶-۳- روش‌های بر پایه گرادیان^۱

هر الگوریتم یادگیری شبکه در واقع یک الگو برای تازه‌سازی وزن‌هاست و هر الگوی تازه‌سازی براساس یک روش بهینه‌سازی تدوین می‌شود. از مشهورترین روشهای بهینه‌سازی، روش‌های بر پایه گرادیان هستند. در تعلیم شبکه‌های عصبی، روشهای بر پایه گرادیان در ترکیب با فرآیند پس انتشار خطا از الگوریتم‌هایی هستند که کاربرد بسیاری دارند.

روشهای بر پایه گرادیان امکان تعیین جهت تغییرات براساس اطلاعات مشتق تابع را فراهم می‌کنند. مشهورترین اعضای این خانواده از روش‌ها، روش شدیدترین فرود و روش نیوتن هستند که پایه سایر روشها محسوب می‌شوند. از روش‌های مشابه دیگری توان گوس نیوتن، گرادیان مزدوج و روش لوبنبرگ-مارکواردت^۲ را نام برد.

برای بررسی روشهای بر پایه گرادیان، مرور سه تعریف زیر ضروری است:

۱. اگر $f(X)$ یک تابع اسکالر از X باشد و $X = [x_1, x_2, \dots, x_n]$ بردار گرادیان $f(X)$ به

صورت رابطه (۶-۲) تعریف می‌شود.

¹ Gradient-Based

² Levenberg-Marquardt

$$\Delta f(X) = \begin{bmatrix} \frac{\partial f(X)}{\partial x_1} \\ \vdots \\ \frac{\partial f(X)}{\partial x_n} \end{bmatrix} \quad (2-6)$$

۲. اگر $f(X)$ یک تابع برداری از X باشد و به صورت

$$f(X) = [f_1(X), f_2(X), \dots, f_n(X)]^T$$

$m \times n$ است که به صورت رابطه (۳-۶) تعریف می‌شود:

$$J = \begin{bmatrix} \frac{\partial f_1(X)}{\partial x_1} & \dots & \frac{\partial f_1(X)}{\partial x_n} \\ \vdots & \dots & \vdots \\ \frac{\partial f_m(X)}{\partial x_1} & \dots & \frac{\partial f_m(X)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \Delta^T f_1(X) \\ \vdots \\ \Delta^T f_m(X) \end{bmatrix} \quad (3-6)$$

۳. ماتریس هرمیشن از یک تابع عددی $f(X)$ روی X به صورت رابطه (۴-۶) تعریف می‌شود.

در واقع ماتریس هرمیشن از یک تابع عددی ژاکوبین گرادیان آن تابع است.

$$H_f = \begin{bmatrix} \frac{\partial^2 f(X)}{\partial x_1^2} & \frac{\partial^2 f(X)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(X)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(X)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(X)}{\partial x_2^2} & \dots & \frac{\partial^2 f(X)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial^2 f(X)}{\partial x_n \partial x_1} & \frac{\partial^2 f(X)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(X)}{\partial x_n^2} \end{bmatrix} \quad (4-6)$$

در روش‌های بر پایه گرادیان با فرض اینکه بردار پارامتر موردنظر به صورت رابطه (۵-۶) تعریف

شده باشد در جستجوی مقدار $\omega = \omega^*$ هستیم که تابع $E(\omega)$ یامیزان خطا را کمینه کند.

$$\omega = [\omega_1, \omega_2, \dots, \omega_N] \quad (5-6)$$

E می‌تواند تابعی غیرخطی و پیچیده باشد، بنابراین از یک الگوریتم گام به گام برای نزدیک شدن

به ω^* استفاده می‌شود:

$$\omega_{\text{next}} = \omega_{\text{now}} + yd \quad (6-6)$$

$$\omega_{k+1} = \omega_k + y_k d_k \quad (7-6)$$

در این راستا، ابتدا در مورد d_k یا جهت تصمیم‌گیری می‌شود و سپس اندازه گام یا y تعیین می‌شود

برای همگرا شدن ω_k به ω_k^* رابطه (6-8) باید برآورده شود:

$$E(\omega_{\text{next}}) = E(\omega_{\text{now}} + yd) < E(\omega_{\text{now}}) \quad (8-6)$$

در یک روش بر پایه گرادیان، گرادیان خطا به صورت رابطه (6-9) تعریف می‌شود.

$$g(t) = \nabla E(\omega) = \left[\frac{\partial E(\omega)}{\partial \omega_1}, \dots, \frac{\partial E(\omega)}{\partial \omega_n} \right]^T \quad (9-6)$$

وقتی $d = -g$ باشد، d جهت شدیدترین ورود در نقطه محلی $\omega_{(\text{new})}$ را نشان می‌دهد.

بر این اساس رابطه زیر، به صورت فرم مشهور رابطه شدیدترین ورود مطرح می‌شود که در آن y

اندازه گام الگوریتم است.

$$\omega_{\text{next}} = \omega_{\text{now}} - yg \quad (10-6)$$

جهت حرکت d می‌تواند براساس مشتق مرتبه دوم تابع خطا انتخاب شود. در این صورت، روش

موسوم به روش نیوتن - رافسن خواهد بود و در رابطه تازه‌سازی پارامتر به این روش از ماتریس

همیشه استفاده می‌شود.

$$\omega_{new} = \omega_{now} - H^{-1}g \quad (11-6)$$

در روش نیوتن، اگر ماتریس هرمیشن مثبت معین نباشد، جهتی که حاصل می‌شود ممکن است به یک ماکزیمم محلی یا یک نقطه زین اسبی¹ منجر شود. برای برای جلوگیری از این پدیده، ماتریس هرمیشن طبقه رابطه (12-6) با ماتریس دیگری جایگزین می‌شود که در آن I ماتریس یک، λ یک مقدار غیرمنفی و H همان ماتریس هرمیشن است.

$$\omega_{next} = \omega_{now} - (H + \lambda I)^{-1}g \quad (12-6)$$

این روش، لونبرگ-مارکواردت نام دارد و بر حسب اینکه مقدار λ چقدر باشد، این روش بین دو روش قبلی نیوتن و شدیدترین فرود حرکت می‌کند، به این ترتیب که با نزدیک شدن λ به صفر روش نیوتن و با حرکت λ به سمت بی‌نهایت روش شدیدترین فرود حاصل می‌شود. با جایگزین کردن تقریب ماتریس هرمیشن به جای H، رابطه (13-6) حاصل می‌شود:

$$\omega_{k+1} = \omega_k - [J^T J + \lambda I]^{-1}g \quad (13-6)$$

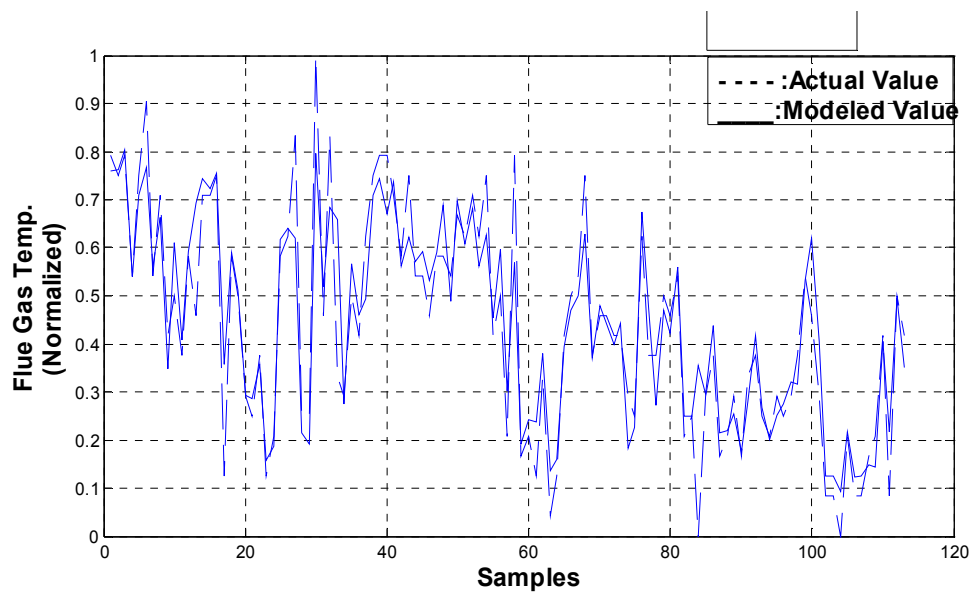
ماتریس هرمیشن در شرایطی به صورت استفاده شده در رابطه (13-6) قابل تقریب زدن است. در این رابطه J ماتریس ژاکوبین است. پیچیدگی محاسبه ماتریس ژاکوبین به مراتب کمتر از محاسبه ماتریس هرمیشن است و این تقریب یک مزیت محاسباتی محسوب می‌شود. رابطه نهایی این روش

¹ Saddle point

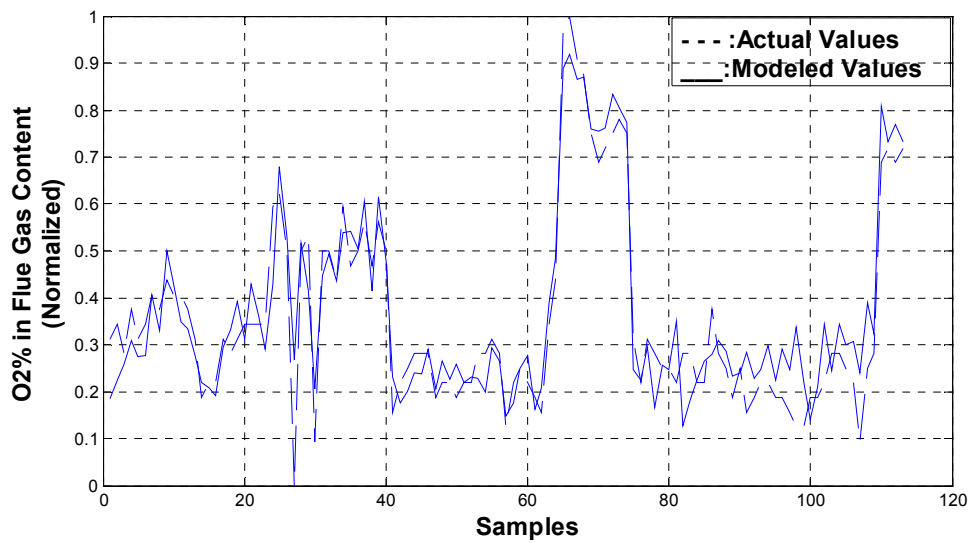
به صورت (۶-۱۳) خواهد بود که در آن با کم شدن مقدار تابع عملکرد، λ شروع به کم شدن کرده و با افزایش مقطعی آن افزایش می‌یابد.

۶-۴- نتایج مدل‌سازی با استفاده از شبکه‌های عصبی

برای مدل‌سازی فرآیند احتراق در بویلر به کمک شبکه‌های عصبی، از یک شبکه پرسپترون ۳ لایه با ۱۳ نرون در لایه مخفی استفاده شده است. برای تعلیم شبکه از الگوریتم پس انتشار خطا با روش بهینه‌سازی لونیبرگ-مارکوارت استفاده شد و در نهایت تخمین‌های حاصل برای دو پاسخ دما و اکسیژن موجود در گاز خروجی از فرآیند احتراق به شکل‌های نمایش داده شده در (۶-۱) و (۶-۲) به دست آمدند. میانگین قدر مطلق خطای مدل برای گاز خروجی در تخمین اکسیژن ۵/۷٪ و در تخمین دما ۵/۵٪ است.



شکل (۶-۱): مدل دمای گاز خروجی با شبکه عصبی



شکل (۶-۲): مدل اکسیژن گاز خروجی با شبکه عصبی

فصل ۷- آشنایی با داده‌کاوی

۷-۱- معرفی

ایده مدیریت بهینه منابع اطلاعات و استفاده از اطلاعات ذخیره شده، به موازات رشد امکانات سخت افزاری و نرم افزاری برای ثبت و ذخیره سازی داده‌ها در حوزه‌های مختلف علم و فن شکل گرفت. گسترش یافتن و نظام مند شدن مسائل مطرح در حوزه این ایده منجر به تولد دانش جدید داده‌کاوی شد. داده‌کاوی دانش استخراج اطلاعات مفید از مجموعه‌های داده^۱ یا پایگاه‌های بزرگ داده^۲ است. این دانش با حوزه‌های مختلف آمار، مدیریت داده‌ها، آموزش ماشین، بازشناسی الگوها، هوش مصنوعی و سایر زمینه‌های آنالیز داده در ارتباط است و در حوزه هوش مصنوعی از آن با عنوان KDD^۳ نیز یاد می‌شود. این دانش نو ظهور مسیر رشد خود را از حوزه‌های تجارت پی‌گرفت و نزدیک به دو دهه از استفاده از آن در حوزه‌های کسب و کاری گذرد. راه داده‌کاوی بعد از حوزه تجارت و کار با پایگاه‌های داده، به بهینه سازی و کنترل پروسه‌های صنعتی نیز باز شد. در حوزه فرآیندهای صنعتی نیز وقتی با پایگاه‌های بسیار بزرگی از داده مواجه هستیم باید مهارت استخراج

^۱ Data Sets

^۲ Data Bases

^۳ Knowledge Discovery in Data Base

داده‌های با ارزش از میان حجم بسیار بزرگ داده‌های مشاهده شده و اندازه‌گیری شده را داشته باشیم و بتوانیم " جریان داده " را به اطلاعات قابل استفاده تبدیل کنیم.

۷-۲- تفاوت دانش داده‌کاوی و آمار

با وجود ارتباط تنگاتنگ دانش داده‌کاوی با علم آمار، تفاوت‌هایی نیز بین رهیافت‌ها و اهداف در این دو حوزه وجود دارد. اولین جنبه تفاوت داده‌کاوی و آمار این است که در آمار، داده‌ها به منظور پاسخگویی به سوالاتی مشخص و از پیش تعیین شده و براساس راهبردهای مشخصی که برای هر نوع سوال تنظیم شده است گردآوری می‌شوند. اما الگوریتم‌های داده‌کاوی، معمولاً بر مجموعه‌های داده‌ای اعمال می‌شوند که از پیش و به منظور تأمین خواسته دیگری مثل ثبت اطلاعات و بایگانی مقادیر اندازه‌گیری شده جمع آوری شده‌اند.

دومین جنبه تفاوت این دو دانش این است که از داده‌کاوی معمولاً در مواقع روبرو بودن با پایگاه‌های بسیار بزرگ داده و حجم زیاد پردازش استفاده می‌شود؛ اما آمار چنین توانایی ندارد. گاهی اوقات در داده‌کاوی داده مورد بررسی یک نمونه از یک جمعیت آماری است و هدف تعمیم‌دادن یک ویژگی از داده مورد بررسی به کل جامعه مورد نظر است. در بسیاری از موارد این تعمیم دادن با روشهای کلاسیک آماری امکان پذیر نیست اما به کمک الگوریتم‌های داده‌کاوی ممکن می‌شود.

در کل می‌توان تفاوت دو دانش داده‌کاوی و آمار را به تفاوت " نگاه " این دو علم به مساله نسبت داد؛ به این ترتیب که داده‌کاوی به مساله نگاهی به صورت " پردازش داده‌ها در طول یک الگوریتم و رسیدن به نتیجه مطلوب " دارد، در حالی که آمار بیشتر به دنبال " مدل‌سازی " و " تخمین " جوامع آماری است.

۷-۳- قابلیت‌های داده‌کاوی

با توجه به آنچه در مورد ظهور و ماهیت دانش داده‌کاوی مطرح شد مجموعه وظایف و قابلیت‌های مورد انتظار از به‌کارگیری دانش داده‌کاوی را می‌توان در موارد زیر خلاصه کرد:

۱. آنالیز اکتشافی داده^۱

در این مجموعه از قابلیت‌های داده‌کاوی، بدون اینکه هدف جستجوی موضوع یا الگوی خاصی باشد، داده‌ها دسته‌بندی می‌شوند. این قابلیت معمولاً برای مقاصد مشابه نمایش و خلاصه‌سازی داده‌ها و یا ارائه ویژگی‌های مجموعه‌های داده‌های از طریق گرافها یا نمودارها مورد استفاده قرار می‌گیرد.

¹ Exploratory Data Analysis

۲. مدل‌سازی توصیفی^۱

با استفاده از یک مدل، رفتار تمام مجموعه داده‌ای توصیف می‌شود، یا یک الگوی توصیف رفتار بلند مدت از آنچه مجموعه داده‌ای را تولید کرده استخراج و تعیین می‌شود. این قابلیت داده‌کاوی، یکی از پرکاربردترین قابلیت‌های این دانش است.

۳. مدل‌سازی پیش‌بین^۲

مشابه حالت قبل، رفتار یک مجموعه داده‌ای یا یک زیرمجموعه از داده‌ها مورد تحلیل و تعیین قرار می‌گیرد. از طریق آنچه برای توصیف رفتار عمومی داده‌ها حاصل شده است، می‌توان مقادیر ناشناخته یا اندازه‌گیری نشده تعدادی از متغیرها را براساس مقادیر معلوم سایر متغیرها تعیین کرد.

۴. اکتشاف قوانین و الگوها^۳

با استفاده از این قابلیت داده‌کاوی، بخشهایی از فضای داده که نقاط داده‌ای در آنها به طور مشخصی متمایزاند تعیین می‌شوند. به این معنی که الگوهای تکراری و معنی‌دار در یک

¹ Descriptive Modeling

² Predictive Modeling

³ Discovery Patterns and Rules

فضای داده‌ای شناسایی شده و استخراج می‌گردند. این قابلیت داده‌کاوی نیز به همراه دو قابلیت فوق از کاربردهای بسیار رایج داده‌کاوی است.

۵. بازیابی از طریق مضمون^۱

استفاده کننده از این روش، با در دست داشتن یک الگوی مشخص، در یک مجموعه داده‌ای به دنبال الگوهای تکرار شونده مشابه با الگوی خود می‌گردد. موارد یافت شده بر حسب مشابهت با مضمون مورد جستجو طبقه بندی و اولویت بندی می‌شوند. موتورهای جستجوگر اینترنت، مثل گوگل از روشی مشابه این روش برای جستجوی صفحات وب استفاده می‌کنند.

۷-۴ - ساختارهای داده‌ای

داده‌های اندازه گیری شده رامی‌توان در قالب n شیء^۲ داده‌ای با P مشخصه اندازه گیری شده^۳ تجسم کرد که هر مشخصه اندازه گیری شده خود می‌تواند شامل چند زیر مجموعه اندازه گیری باشد.

عبارت مدل یک توصیف مختصر و عام از یک مجموعه داده است و به طور عمومی در مورد هر نقطه از فضای اندازه گیری صحبت می‌کند. مثلاً اگر فضای داده‌ای مورد بحث یک ستون

¹ Retrieval by Content

² Data Object

³ Attribute

از یک ماتریس $n \times p$ باشد که تعدادی از درایه‌های آن نیز از بین رفته باشد، فضای داده هنوز با یک عبارت عمومی قابل توصیف است؛ مثلاً می‌توان گفت که "مدل" فضا به فرم $y=ax+b$ است و مقادیر پارامترهای a و b موضوعی است که تعیین آن محدود کننده تعریف مدل نیست.

در مقابل کلمه "مدل"، از عبارت "الگو"¹ زمانی استفاده می‌شود که در مورد یک ناحیه محدود از فضا اظهار نظری کنیم. معمولاً قیدی برای مقادیر متغیرها وجود دارد که موضوع بحث را به جزء نسبتاً کوچکی از ناحیه داده‌ها محدود می‌کند. الگوها ساختارهای غیر عادی در میان داده‌ها هستند که انحرافات ثابت از رویه اصلی داده‌ها را نشان می‌دهند. الگوها هم برای منظورهای توصیفی و هم برای منظورهای پیش بینی قابل استفاده هستند.

برای ساختن مدل‌های توصیفی عام، روشهای مختلفی مورد استفاده قرار می‌گیرد؛ از جمله روشهایی که بر پایه توزیع‌های احتمال عمل کرده و داده‌ها را بر حسب توزیع نهفته در آنها توصیف می‌کنند. روشهای آنالیز خوشه‌ای و روشهای آماری مثل بررسی ضرایب همبستگی نیز به این منظور قابل استفاده هستند.

برای ساختن الگوهای محلی نیز رهیافت‌های مختلفی وجود دارد. از جمله استخراج قوانین متحد برای پیدا کردن همه قوانین از یک نوع مشخص که دارای بسامد یا دقت خاص از

¹ Pattern

پیش تعیین شده هستند. همچنین با تعیین آستانه‌های مشخص می‌توان داده‌هایی که از روند اصلی و طبیعی داده‌ها "بیرون زدگی" دارند را مشخص کرد.

رهیافت‌های ساخت مدل‌های توصیفی براساس روشهای فوق نیز عبارتند از استفاده از شبکه‌های عصبی مثل شبکه‌های خود سازمانده کوهونن^۱، آنالیز فاکتورهای آماری^۲ و منطق فازی برای کشف قوانین متحد و خوشه‌سازی.

در مدل‌سازی پیش بین برای ساختن مدل مورد استفاده و بعد پیش بینی مقدار یا مشخصه تعدادی از متغیرها، از یک مجموعه داده‌ای در دسترس استفاده می‌شود. این مجموعه شامل ورودی‌ها و خروجی‌های مدل است. مدل تخمین زده شده براساس نگاشت حاصل از این مجموعه ورودی و خروجی تعیین شده و مورد استفاده قرار می‌گیرد.

شبکه‌های عصبی قابلیت دو گانه‌ای دارند که استفاده از آنها را هم به منظور دسته بندی و هم به منظور تخمین مدل ممکن می‌سازد. شبکه‌های عصبی انعطاف پذیرند، یعنی تعداد لایه‌ها و عصب‌های آنها با توجه به ملاحظات طراحی قابل تغییر است و با افزایش حجم نمونه‌های مورد بررسی، پارامترهای شبکه‌های عصبی به‌طور نمایی و چشمگیر رشد نمی‌کند. به خاطر داشتن این مجموعه ویژگی مناسب، شبکه‌های عصبی به طور گسترده در

¹ Kohonen

² Statistical Factor Analysis

کاربردهای مختلف داده‌کاوی مورد استفاده قرار می‌گیرند. اولین قدم در استفاده از شبکه‌های عصبی تعیین ساختار یا معماری شبکه است. تعداد لایه‌ها، عصب‌ها، ساختار و ابعاد شبکه متناسب با طبیعت و میزان پیچیدگی مسأله تعیین می‌شود. بنابراین، این تعیین ساختار معمولاً با آزمون و خطا همراه است. گام دوم تعلیم دادن شبکه بر حسب ورودیها و خروجیهای در دسترس و گام سوم راستی آزمایی مدل پیش بین حاصل از شبکه تعلیم یافته است.

۷-۵- گامهای لازم برای داده‌کاوی

برای تحلیل یک مسأله از طریق دانش داده‌کاوی، مراحل زیر باید پیموده شوند [۱۸]:

۱. تحلیل صورت مساله : نخستین گام در حل مساله مهم موضوع مساله و تعیین نتیجه‌ای است که هدف از حل مساله با داده‌کاوی رسیدن به آن است.
۲. جمع آوری داده و انتخاب پایگاه داده: بعد از تعریف دقیق صورت مسأله، متناسب با هدفی که مساله دنبال می‌کند، باید به جستجوی مناسب‌ترین مجموعه داده‌ای یا پایگاه داده پرداخت.
۳. آماده سازی داده‌ها: در مجموعه‌های بزرگ داده، اتفاقاتی نظیر وقوع داده‌های متناقض، وجود نقاط پرتاب و داده‌های نامتعادل و یا از دست رفتن داده‌ها بسیار متحمل است.

برای مقابله با چنین مشکلاتی روش‌های مختلفی وجود دارد که باید روش مناسب با توجه به نوع مسأله انتخاب و اختیار شود. علاوه بر این برای پیش پردازش و نویز زدایی از داده‌ها نیز از روش‌هایی مثل میانگین متحرک^۱، یعنی استفاده از تعداد n داده مشاهده شده قبلی برای نویز زدایی از داده جدید و یا خانواده الگوریتم‌های موجک^۲ استفاده می‌شود. فرآیند داده‌کاوی بعد از انتخاب پایگاه داده و آماده سازی آن، قابل اجراست.

۷-۶- کاربردهای داده‌کاوی

با رشد روز افزون امکانات ذخیره سازی و پردازش داده، جایگاه داده‌کاوی نیز هر روز بیش از پیش مورد توجه قرار می‌گیرد. به موازات کاربردهای صنعتی داده‌کاوی، کاربردهای جالب توجه دیگری از این دانش نیز وجود دارد. هواشناسی، خدمات حوزه وب مثل موتورهای جستجو، تجارت الکترونیکی، و مطالعات حوزه بهداشت و سلامت عمومی، جرم شناسی و مطالعات جمعیتی از این جمله‌اند.

کاربرد دیگر داده‌کاوی در بانکداری و تجارت است. رقابت فراینده بانک‌ها، شرکت‌های تجاری، شرکت‌های بیمه و ایجاب می‌کند که داده‌های مربوط به کسب سود، محاسبه

¹ Moving Average

² Wavelet

میزان خطر پذیری یا عوامل جلب مشتری و ارائه خدمات دائماً مورد تحلیل قرار گرفته و در

تصمیم سازی دخالت داده می‌شوند.

اما کاربرد داده‌کاوی در تحلیل و کنترل فرآیندهای صنعتی؛ آنچنان که در این پژوهش مورد

استفاده قرار گرفته است، از کاربردهای نوین داده‌کاوی بوده که روز به روز بر دامنه استفاده

از آن افزوده می‌شود.

فصل ۸ - الگوریتم‌های داده‌کاوی

۸-۱- آشنایی با الگوریتم‌های داده‌کاوی و اجزای آنها

برآوردن آنچه از کاربردها و قابلیت‌های دانش داده‌کاوی بر عهدهٔ بدنه دانش داده‌کاوی، یعنی الگوریتم‌های داده‌کاوی است. این الگوریتم‌ها از تنوع قابل توجهی برخوردارند که بنابر نوع مسأله مورد بحث باید خانواده الگوریتم مناسب مسأله انتخاب شود [۱۹]. با وجود تنوع این الگوریتم‌ها، می‌توان سازمان مشابهی برای تمام آنها تعریف کرد. تمام این الگوریتم‌ها از این اجزا تشکیل شده‌اند:

- I. ساختار مدل یا الگو، که تعیین کننده شکل الگوریتم است.
- II. تابع امتیاز دهی، که معیاری برای سنجش عملکرد الگوریتم است. براساس مقدار حاصل از این تابع می‌توان قضاوت کرد که مدل یا ساختار انتخاب شده تا چه اندازه توصیف کننده مجموعه داده‌ای است. معروف ترین و رایج ترین توابع امتیاز دهی، توابع بیشترین شباهت^۱ و مجموع مربعات هستند.
- III. روش بهینه سازی، که الگوریتمی است که پارامترهای مدل اختیار شده را در جهت کاهش مقدار خطا و بهبود و مقدار تابع امتیازدهی تغییر می‌دهد.

¹ Maximum Likelihood

IV. روش مدیریت داده، هر الگوریتم بنابر ویژگیهای خود و روش مشخصی برای

مدیریت مجموعه داده‌ای خود انتخاب می‌کند.

تعیین ساختار و تابع امتیازدهی به صورت خارج از خط و توسط طراح الگوریتم

انجام می‌شود. پس از آن پیموده شدن الگوریتم و بهینه سازی مراحل محاسباتی هستند که

پی در پی انجام می‌شوند. به عنوان مثالی برای اجزای الگوریتم، الگوریتم نمونه پس انتشار

خطا در نظر گرفته می‌شود. ساختار الگوریتم از نوع شبکه‌های عصبی، تابع امتیازدهی از نوع

مربع خطا و روش بهینه سازی جواب گرادیان نزولی است.

در ادامه، الگوریتم خوشه‌سازی که در این پژوهش مورد استفاده قرار گرفته است به تفصیل

بررسی می‌شود. همچنین دو الگوریتم پر استفاده دیگر نیز مورد اشاره قرار خواهند گرفت.

۸-۲- الگوریتم‌های خوشه‌سازی

الگوریتم‌های خوشه‌سازی به منظور مدیریت کردن داده‌های مورد مشاهده در قالب

گروه‌های معنی دار مورد استفاده قرار می‌گیرند. یک مسأله خوشه‌سازی باید قابلیت مقیاس

پذیری داشته باشد، یعنی به‌طور همزمان روی مسائل کوچک و روی مسائل بزرگ قابل

پیاده سازی باشد [۲۰].

علاوه بر این باید در مقابل نویز و خطای داده‌ای نیز مقاوم بوده و به ترتیب ورود داده‌ها وابسته نباشد.

گام ابتدایی در استفاده از این الگوریتم تعیین ویژگی مورد استفاده برای تقسیم بندی داده‌هاست. ویژگی مورد نظر باید هم به خوبی "توصیف کننده" داده‌ها بوده و با اهمیت باشد و هم به خوبی "متمایز کننده" باشد. همچنین در صورتیکه چند ویژگی همزمان مورد نظر باشد، می‌توان با توجه به اهمیت هر ویژگی برای آن وزن مناسب تعیین کرد. تفکیک شدن داده‌ها، بر مبنای "فاصله" ویژگی‌های مورد نظر در آنها انجام می‌پذیرد. فاصله مفهومی است که درک آن به سادگی امکان پذیر است و علاوه بر این ویژگی‌های زیر را نیز دارد:

I. همیشه مقداری مثبت است.

II. مقدار فاصله x تا y با مقدار فاصله y تا x برابر است.

III. فاصله x با خودش برابر صفر است.

IV. فاصله از قضیه حمار تبعیت می‌کند، یعنی $xz + zy > xy$

اگر فاصله دو نقطه x و y را با $D(x,y)$ نمایش بدهیم؛ تعاریف متعددی برای نحوه

محاسبه D وجود دارد. معروف‌ترین این تعاریف عبارتند از :

I. فاصله اقلیدسی یا نرم L_2 مرسوم‌ترین نوع فاصله است که به صورت

$$\sum_i (x_i - y_i)^2)^{1/2}$$
 تعریف می‌شود.

II. فاصله منهن با نرم L_1 ، که به صورت $\sum_i |x_i - y_i|$ تعریف می‌شود و در بسیاری

از موارد به نتایجی مسابه با نرم افزار اقلیدسی منجر می‌شود.

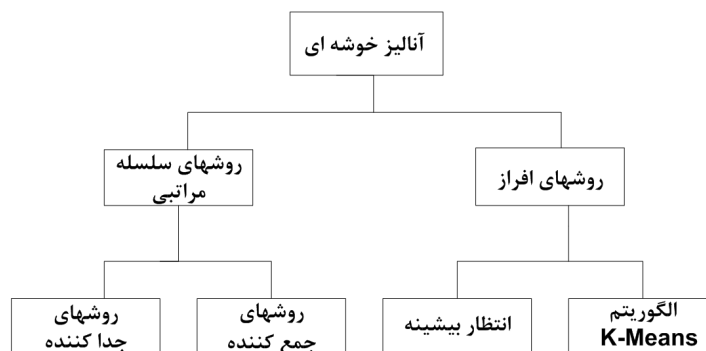
III. فاصله چیبی شف یا نرم L_∞ ؛ ماکزیمم فاصله بین ویژگی‌هاست و به صورت

$$\max |x_i - y_i|$$
 تعریف می‌شود

به منظور دسته بندی داده‌ها بر حسب فاصله بین ویژگی‌هایشان روشهای متفاوتی وجود

دارد. دو دسته اصلی از این روشها که در شکل (۸-۱) نمایش داده شده‌اند در ادامه مورد

بررسی قرار می‌گیرند.



شکل (۸-۱): تقسیم بندی روشهای خوشه‌سازی

۸-۲-۱- روشهای افراز^۱

در این روشها یک دیوارکشی تک سطحی بین داده‌ها انجام می‌گیرد و داده‌ها در خوشه‌های مختلف تقسیم بندی می‌شوند. هیچ شیء داده‌ای متعلق به بیش از یک خوشه نیست و تخصیص پیدا کردن داده‌ها به خوشه‌ها ممکن است در گامهای مختلف تغییر کند.

مهمترین روش در مجموعه روشهای افراز، استفاده از الگوریتم K-Means است. الگوریتم K-Means با هدف قرار دادن N شیء، هرکدام با P پارامتر اندازه گیری شده در K گروه (که K عدد معلومی است) طراحی شده است. ابتدا K هسته ابتدایی برای K خوشه انتخاب می‌شود که هر هسته یک بردار P بعدی است. هر بردار از N بردار مورد بحث، در گروهی قرار می‌گیرد که فاصله اش با مرکز آن گروه از فاصله اش با مرکز سایر خوشه‌ها کمتر است.

بعد از عضو گیری اولیه خوشه‌ها، میانگین متغیرهای اعضا گرفته می‌شود. براساس این میانگین، بردار مرکزی جدید تعیین شده و با بردار جدید گروه بندی دوباره انجام می‌شود. این مرحله آنقدر تکرار می‌شود که با تعیین مرکز جدید و اجرای مجدد الگوریتم هیچ داده‌ای بین دو خوشه جابه‌جا نشود.

¹ Partitional Methods

برای استفاده از الگوریتم K-Means، می‌توان داده‌ها را در بازه مشخصی مثل (۰ و ۱) مقیاس بندی^۱ کرد. مثلاً در شرایطی که هدف وزن دهی به داده‌ها نیز مورد نظر باشد، استفاده از فرم نرمالیزه داده‌ها ضروری است. مقیاس‌بندی بایاس ویژگیها را که ناشی از داشتن مقیاس‌های مختلف است حذف می‌کند و به این ترتیب همه داده‌ها را هم وزن می‌سازد. برای تبدیل مقیاس بندی از روشهای زیر استفاده می‌شود:

۱. مقدار هر ویژگی بر مقدار میانگین ویژگی تقسیم می‌شود. به این ترتیب مقدار میانگین هر ویژگی به عدد ۱ تبدیل می‌شود.

۲. ویژگی بر تفاضل کمترین و بیشترین مقدار ویژگی تقسیم می‌شود. به این ترتیب میانگین ویژگی‌هایی که طیف زیادی از مقادیر را به خود اختصاص داده‌اند کم و میانگین ویژگی‌هایی که طیف کمی از مقادیر را دربردارند زیاد می‌شود.

۳. در این روش برخلاف دو روش قبل هم مقدار میانگین و هم مقدار انحراف از معیار استاندارد می‌شود؛ به این ترتیب که میانگین مقدار هر ویژگی از مقدار آن کم و به مقدار انحراف معیارش تقسیم می‌شود.

مسئله دیگری که در مورد استفاده از الگوریتم K-Means مطرح است مقادیر ابتدایی الگوریتم است. معمولاً در اولین گام استفاده از این الگوریتم، اطلاعات کمی در مورد چگونگی

¹Scaling

پخش مقادیر اولیه و تعداد خوشه‌های مناسب وجود دارد و دید کلی مورد نیاز، معمولاً پس از چندین مرحله آزمون و خطا حاصل می‌شود. علاوه بر آزمون و خطا رهیافت معمول دیگری نیز به این منظور وجود دارد که در آن یک مرکز برای تمام پایگاه داده انتخاب شده و بعد مرکز به نقاط مختلفی شکسته می‌شود که آن نقاط به عنوان هسته‌های مرکزی خوشه‌ها اختیار می‌شوند. رهیافت دیگری نیز وجود دارد و آن انجام خوشه بندی برای یک نمونه تصادفی از مجموعه داده و استفاده از ایده حاصل برای کل مجموعه داده است.

در مورد استفاده از الگوریتم K-Means ملاحظات دیگری نیز وجود دارد:

۱. وقتی با حجم زیادی از داده مواجه هستیم، چندین بار پیمودن این حجم از داده حافظه زیادی را به کار گرفته و به خود مشغول می‌کند.

۲. الگوریتم K-Means چون براساس فاصله‌های اقلیدسی عمل می‌کند، تنها قادر به تولید خوشه‌های کروی است.

۳. نتایج الگوریتم بسیار به نحوه انتخاب مقادیر اولیه حساس است.

۴. به اندازه خوشه‌های ایجاد شده حساسیتی ندارد و ممکن است خوشه‌هایی بسیار بزرگ یا بسیار کوچک در کنار هم ایجاد کند.

۵. به مقادیر نقاط پرتاب بسیار حساس است. اگر نقطه پرتاب به عنوان نقطه مرکزی یک خوشه انتخاب شود، خوشه تک عضو تولید می‌شود و اگر نقطه پرتاب در طول عملیات بین خوشه‌ها جا به جا شود، مقادیر میانگین به شدت تحت تاثیر قرار می‌گیرند.

۸-۲-۲- روش‌های سلسله مراتبی^۱

در این روشها یک تقسیم بندی تو در تو از اشیاء انجام می‌پذیرد که به یک ساختار درختی از خوشه‌ها منجر می‌شود. این روشها در دو حالت اجرایی شوند. یکی اینکه همه داده‌ها در مرحله اول به تنها یک خوشه متعلق دانسته و بعد آن خوشه به خوشه‌های کوچکتر شکسته می‌شود؛ و دیگری اینکه هر داده به یک خوشه تعلق داشته باشد و در مراحل بعد خوشه‌ها در هم ادغام شده و به خوشه‌های بزرگتر تبدیل شوند.

در روشهای افراز یک مجموعه هموار از خوشه‌ها تولید می‌شود. در حالی که روشهای سلسله مراتبی یک مجموعه تو در تو یا یک درخت از خوشه‌ها تولید می‌کنند. روشهای سلسله مراتبی به دو گروه عمده روشهای جمع کننده و روشهای تقسیم کننده تقسیم می‌شوند.

۸-۲-۲-۱- روشهای جمع کننده^۲

در این دسته از روشها، ابتدا هر شیء به تنهایی یک خوشه را به خود تخصیص می‌دهد. بعد با محاسبه فاصله بین خوشه‌ها، خوشه‌ها در هم ادغام می‌شوند تا تعداد مناسبی از

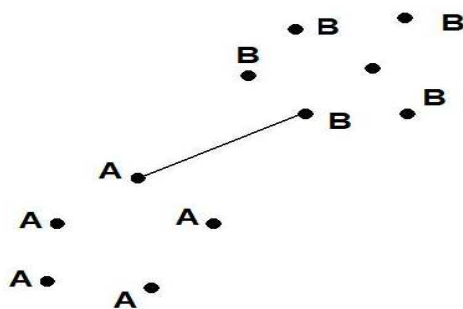
¹ Hierarchical

² Agglomerative

خوشه‌ها حاصل شود. بنابراین مساله خوشه‌سازی به روش سلسله مراتبی، تبدیل به مساله سنجش فاصله بین خوشه‌ها می‌شود. با توجه به اینکه در شرایطی تعداد زیاد اعضای خوشه‌ها منجر به حجم زیادی از پردازش مورد نیاز می‌شود، معیارها و روشهای مختلفی برای سنجش فاصله خوشه‌ها تعریف می‌شود. تعدادی از این معیارها به شرح زیرند:

I. معیار تک خط^۱: فاصله بین دو خوشه به صورت کمترین فاصله بین دو عضو

خوشه (زوج دوتایی) تعریف می‌شود.

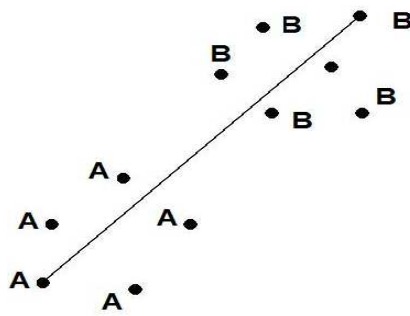


شکل (۸-۲): معیار تک خط

II. معیار خط کامل^۲: فاصله به صورت بیشترین فاصله بین دو عضو تعریف می‌شود.

¹ Single Link

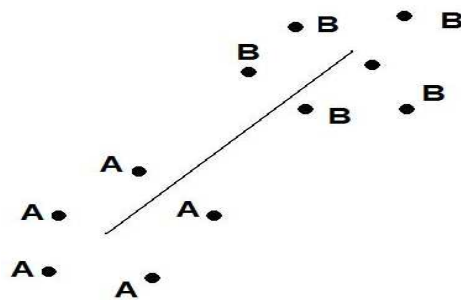
² Complete Link



شکل (۳-۸): معیار خط کامل

III. معیار خط مرکز^۱: فاصله بین مراکز دو خوشه که معمولاً بر حسب اقلیدسی

سنجیده شده به عنوان فاصله دو خوشه در نظر گرفته می‌شود.



شکل (۴-۸): معیار خط مرکز

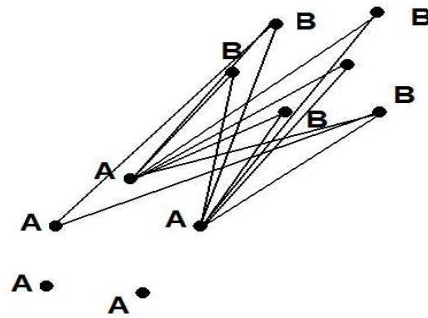
IV. معیار خط میانگین^۲: فاصله بین تک تک اعضای دو خوشه با هم محاسبه شده

و میانگین این مقدار به عنوان فاصله دو خوشه معرفی می‌شود. این معیار در مقایسه

با ۳ معیار قبلی به نتایج بهتری منجر می‌شود.

¹ Centroid Link

² Average Link



شکل (۸-۵): معیار خط میانگین

در هر بار پیمایش الگوریتم بعد از محاسبه فاصله بر اساس یکی از چهار معیار فوق، دو خوشه که کمترین فاصله را با هم دارند، درهم ادغام شده و یک خوشه واحد تشکیل می‌دهند و الگوریتم تا رسیدن به تعداد مشخصی از خوشه‌ها ادامه پیدامی‌کند.

گامهای اجرای این الگوریتم رامی‌توان به صورت زیر خلاصه کرد:

مرحله ۱ : ساختن n خوشه از شی داده‌ای

مرحله ۲ : ساختن یک ماتریس فاصله که هر درایه اش فاصله بین دو عضو بر حسب یکی از

۴ معیار فوق است.

مرحله ۳ : انتخاب کمترین فاصله از بین درایه‌های ماتریس

مرحله ۴ : ادغام خوشه‌های متناظر با کمترین فاصله و بازگشت به مرحله ۲ و تازه سازی

ماتریس فاصله‌ها.

۸-۲-۲-۲- روش‌های جدا کننده^۱

در این روش که مسیری کاملاً بر عکس روش قبل دارد الگوریتم با قرار دادن تمام اشیاء داده‌ای در یک خوشه آغاز می‌شود و تا رسیدن به قید مشخصی ادامه پیدامی‌کند. مشابه حالت قبل ابتدا یک معیار برای اندازه‌گیری فاصله انتخاب شده و یک مقدار آستانه نیز برای این فاصله تعریف می‌شود. خلاصه مراحل اجرای الگوریتم به شرح زیر است:

مرحله ۱: ماتریس فاصله بر حسب فاصله بین هر دو عضو در هر خوشه تنظیم می‌شود. دو شی‌داده‌ای که بیشترین فاصله از یکدیگر را دارند انتخاب می‌شوند. فاصله دو شی با مقدار فاصله آستانه مقایسه می‌شود؛ اگر کمتر از مقدار آستانه بود، الگوریتم همین جا متوقف می‌شود و اگر بیشتر بود به مرحله بعد می‌رود.

مرحله ۲: دو شی‌ای که بیشترین فاصله را داشتند به عنوان دو مرکز برای خوشه جدید انتخاب می‌شوند و خوشه بزرگ از طریق الگوریتم K-Means به دو خوشه جدید شکسته می‌شود.

مرحله ۳: برای انتخاب خوشه بعدی برای شکسته شدن و بازگشت به گام ۱، معیارهای قضاوت مختلفی وجود دارد. خوشه بعدی می‌تواند خوشه با بیشترین تعداد اعضا یا خوشه با بیشترین واریانس باشد و یا اینکه تمام خوشه‌ها به نوبت در معرض این الگوریتم قرار بگیرند.

^۱ Divisive

روش‌های سلسله مراتبی خلاصه شده به شرح قبل، از ویژگی‌هایی نیز برخوردارند، اولاً روش‌های سلسله مراتبی در مقایسه با روشهای افراز و دسته بندی دید بهتری نسبت به مسأله فراهم می‌کند و ثانیاً از قابلیت فهم بهتری برخوردارند. اما در مقابل هیچ ساز و کاری برای اصلاح جایگاه داده‌ای که در خوشه نامناسبی قرار گرفته وجود ندارد و داده‌ها در طول الگوریتم بین خوشه‌ها جابه جا نمی‌شوند. علاوه بر این، روشهای سلسله مراتبی بسیار حساس به انتخاب روشهای مختلف و مقیاس بندی و محاسبه فاصله هستند. حجم بالای پردازش مورد نیاز برای محاسبه ماتریس‌های فاصله در پایگاه‌های بزرگ داده‌ای نیز از نقاط ضعف این روشها است.

۸-۳- کلاس بندی و الگوریتم‌های درخت تصمیم‌گیری

کلاس بندی^۱ جداسازی یا مرتب کردن اشیاء در گروه‌های از پیش تعیین شده است. اگر کلاسها بدون در نظر گرفتن داده‌ها تعریف شده باشند؛ مسأله از نوع کلاس بندی از پیش تعیین شده^۲ است و اگر کلاسها با مشاهده داده‌ها تعیین شده باشند مسأله از نوع کلاس بندی پسین^۳ است [۲۱].

^۱ Classification

^۲ Apriori classification

^۳ Postteriori Classification

معمولاً دو هدف با کلاس بندی دنبال می‌شود؛ پیش بینی و فهم بهتر داده‌ها. اگر ویژگی‌های مورد بحث در مساله کلاس بندی عددی باشند، مساله کلاس بندی در واقع یک مساله رگرسیون است. در صورتی که ویژگی‌ها کیفی باشند مساله کلاس بندی را می‌توان از طریق الگوریتم‌های درخت تصمیم‌گیری حل کرد.

درخت تصمیم‌گیری مدلی است که هم پیش بینی کننده و هم توصیف کننده است. تعداد داده‌هایی که برای تعلیم الگوریتم‌های کلاس بندی مورد نیاز است، متناسب با تعداد ویژگی‌های مستقل هر شی داده‌ای مطرح در مساله کلاس بندی است. معمولاً تعداد محدودی از ویژگی‌ها هستند که در تعیین تعلق هر شی به هر کلاس ایفای نقش می‌کنند و بسیاری از ویژگی‌ها تأثیر چندانی در دسته بندی اشیاء ندارند.

هرگره داخلی در درخت، نماینده یک آزمایش دو حالت است که از یک حد آستانه برای متغیرهای عددی و یا عضویت‌های زیر مجموعه‌ای برای متغیرهای گروهی استفاده می‌کند. در مورد ساختار درخت و تعداد گره‌ها قبلاً تصمیم‌گیری نمی‌شود، بلکه داده‌های تصمیم به تدریج و در طول الگوریتم ساختار درخت را شکل می‌دهند.

مراحل اجرای الگوریتم درخت تصمیم‌گیری را می‌توان به صورت زیر خلاصه کرد:

مرحله ۱: مجموعه داده‌های S انتخاب می‌شود.

مرحله ۲: یک ویژگی «به خوبی متمایز کننده» استخراج شده و یک ایستگاه تصمیم گیری ایجاد می شود.

مرحله ۳: در صورت وقوع یکی از این موارد الگوریتم متوقف می شود:

هیچ ویژگی دیگری برای تصمیم گیری موجود نباشد؛

تصمیم گیری منجر به ایجاد دو کلاس متمایز نشده باشد و هیچ ایستگاه تصمیم گیری دیگری نیز وجود نداشته باشد.

۸-۴- الگوریتم شناسایی قوانین متحد

الگوریتم شناسایی قوانین متحد^۱ از روشهای بسیار رایج در کاوش داده هاست. از طریق این الگوریتم، داده هایی که بسامد و الگوی تکرار معنی داری در طول پایگاه داده دارند شناسایی شده و بر حسب اهمیت و احتمال وقوع دسته بندی می شوند و در جهت هدف مساله مورد استفاده قرار می گیرند.

قوانین متحد معمولاً به صورت $X \rightarrow Y$ نمایش داده می شوند. این یک رابطه علی و معلولی نیست و تنها نشان دهنده احتمال وقوع x در حضور y است. دو معیار برای کشف قوانین به هم پیوسته در نظر گرفته می شود:

X در چه تعدادی از کل وقایع وجود دارد؟ (Support)

¹ Association Rules

در شرایط وجود X احتمال رویت شدن Y چقدر است؟ (Confidence)

با توجه به شرایط مساله، در یک پایگاه داده مشخص به دنبال قوانین متحدی می‌گردیم که حداقل

$P\%$ مقدار Support و حداقل $Q\%$ مقدار Confidence داشته باشند.

فصل ۹- الگوریتم کنترل هوشمند مبتنی بر داده‌کاوی

۹-۱- معرفی

با مرور آنچه در فصول گذشته گفته شد، به نظرمی‌رسد که داده‌کاوی می‌تواند به عنوان ابزاری در مسیر کنترل هوشمند و به منظور هدایت فرآیند احتراق بویلر به سمت نقاط کاری با راندمان بیشتر به کار گرفته شود.

روشهای بسیا متفاوتی تا کنون برای کنترل این فرآیند به کار گرفته شده است. علاوه بر روشهای کنترل کلاسیک که در فصل ۱ مورد اشاره قرار گرفتند، روشهای فازی، روشهای خطی سازی فیدبکی و روشهای کنترل پیش بین مدل^۱ یا MPC از جمله راهبردهای کنترل پیشرفته‌ای هستند که در حوزه کنترل احتراق بویلر به خدمت گرفته شده‌اند[۶].

با وجود همه پیشرفتها، به دلایل مختلفی از قبیل خاصیت غیرخطی و تغییرپذیر با زمان و مجموعه عدم قطعیت‌هایی که فرآیند تولید بخار با آن درگیر است، عدم تطابق خوب کنترل کننده‌ها با شرایط واقعی در طول زمان منجر به انباشت خطا و کاهش کارایی کنترل کننده‌می‌شود. علاوه براین معمولاً در تنظیم این کنترل کننده‌ها معیارهایی از قبیل پایداری و افزایش سرعت پاسخ حلقه‌های کنترلی در نظر گرفته می‌شود و این تنظیمات معمولاً بدون داشتن نگاهی عمومی

^۱. Model Predictive Control

به افزایش کارایی انجام می‌شوند و معیارهای سنجش عملکرد فرآیند در طراحی کنترل کننده‌ها دخالت داده نمی‌شود. به موازات این محدودیت‌ها، روشهای مبتنی بر داده‌کاوی بر تحلیل شرایط واقعی حاکم بر فرآیند تکیه دارند و از انعطاف پذیری خوبی به منظور اعمال و سنجش معیارهای متفاوت برخوردارند.

در راهبرد کنترل هوشمند مورد استفاده در این پژوهش، با استفاده از الگوریتم خوشه‌سازی K-Means بهبود یافته در فضای بایگانی اطلاعات عملکرد بویلر به جستجوی الگوهای خوب عملکرد، یعنی نقاط کاری که بویلر در آنها راندمان بهتری دارد می‌پردازیم. سپس با اعمال تغییر در متغیرهای کنترل پذیر نقطه کاری فعلی بویلر، عملکرد آنرا به سمت نزدیک شدن به الگوهای عملکرد با راندمان بیشتر تغییر می‌دهیم.

۹-۲- بررسی داده‌ها و معیار راندمان

پایگاه داده موجود از تاریخچه عملکرد بویلر، پس از مراحل پیش پردازش به دو دسته داده‌های آموزش و داده‌های آزمایش تقسیم بندی شدند. داده‌های آموزش مجموعه‌ای از ۳۰۰ نمونه داده‌ای و مجموعه آموزش شامل ۱۱۳ نمونه داده‌ای بودند که هر نمونه داده‌ای شامل مقادیر پانزده متغیر اندازه‌گیری شده بود. در مجموعه این متغیرها تعدادی کنترل پذیر و تعدادی کنترل ناپذیر هستند. متغیرهای کنترل ناپذیر متغیرهایی هستند که امکان تغییر مقدار آنها از طریق ورودی وجود

ندارد؛ مثل دمای هوا که خارج از حیطه کنترل بوده یا فشار کوره که برای کنترل آن مکانیزم مستقیمی پیش بینی نشده است. متغیرهای کنترل پذیر متغیرهایی هستند که به ازای آنها عنصرنهایی کنترل وجود دارد و مقادیر مطلوب انتخاب شده برای آنها قابل اعمال هستند؛ مثل مقدار دبی آب تغذیه، که با تغییر مقدار بازشدگی شیرکنترلی موجود در مسیر قابل تغییر است. در جدول شکل (۹-۱) پانزده متغیر به تفکیک کنترل پذیر، کنترل ناپذیر و پاسخ نمایش داده شده‌اند.

جدول (۹-۱): تفکیک متغیرها

متغیرهای اندازه‌گیری شده		
متغیرهای کنترل پذیر	متغیرهای کنترل ناپذیر	متغیرهای پاسخ (معیار راندمان)
۱. دبی سوخت	۱. دمای سوخت	۱. دمای گاز خروجی
۲. دبی هوای احتراق	۲. دمای هوای احتراق	۲. اکسیژن موجود در گاز
۳. دمای بخار خروجی	۳. فشار سوخت	خروجی
۴. فشار بخار	۴. فشار درام	
۵. دبی آب تغذیه	۵. فشار کوره	
۶. سطح درام	۶. دبی بخار خروجی	
	۷. فشار آب تغذیه	

مبنای اندازه‌گیری راندمان احتراق در فرآیند تولید بخار در فصل معرفی شده است. به منظور محاسبه راندمان در چهار چوب روابط معمول پایش و ثبت متغیرهایی مانند رطوبت هوا و رطوبت گاز موردنیاز است. در مواجهه با پایگاه داده‌ای که بخشی از پارامترهای موثر در سنجش راندمان در آن ثبت نشده‌اند، باید از اطلاعات موجود و ثبت شده برای نزدیک شدن به معیار گویایی از میزان راندمان بهره جست. از میان پانزده متغیر ثبت شده، دمای گاز خروجی از فرآیند به عنوان معیاری از میزان انرژی خارج شده از فرآیند و اکسیژن موجود در گاز خروجی به عنوان معیاری

از هوای اضافه‌ای که بدون استفاده در فرآیند احتراق گرم شده و به عنوان عامل انتقال انرژی به خارج عمل می‌کند، دو محک مهم در ارزیابی راندمان فرآیند هستند.

همانگونه که در فصل اول مطرح شد، میزان انتخاب شده برای هوای ورودی به کوره از هوای احتراق مورد نیاز که از معادله شیمیایی سوخت حاصل می‌شود بیشتر است. افزودن این مقدار هوای اضافه به هوای احتراق گریزناپذیر بوده و در بارهای مختلف متفاوت است. در بویلر مورد بحث، میزان هوای اضافه در بارهای زیر ۴۰٪ بارنامی معادل ۱۴٪ و در بارهای بالای ۴۰٪ معادل ۸٪ در نظر گرفته شده است. بخشی از میزان اکسیژن موجود در گاز خروجی از فرآیند از میزان هوای اضافه تبعیت می‌کند؛ اما در بازه بار مورد بحث که بین ۵۰-۸۰٪ بار نامی بویلر اختیار شده است، میزان هوای اضافه ثابت است. بنابراین اکسیژن موجود در گاز خروجی در این بازه به عنوان معیاری از کارایی فرآیند احتراق قابل استفاده است.

بر این اساس، معیار مطلوب برای سنجش راندمان به صورت حاصل جمع نرمالیزه شده دمای گاز خروجی و اکسیژن موجود در گاز خروجی مورد توجه قرار می‌گیرد. کمتر بودن این حاصل جمع نشانگر بهتر بودن کارایی سیستم است.

۹-۳- بدنه الگوریتم کنترل هوشمند

این الگوریتم به دفعات تعداد اعضای مجموعه آموزش تکرار می‌شود، یعنی مقدار i از ۱ تا ۱۱۳ تغییر می‌کند:

۱. عضو A_i از مجموعه داده آزمایش اختیاری می‌شود.
۲. کارایی بویلر در نقطه کاری A_i با استفاده از معیار معرفی شده اندازه گیری می‌شود و A_{indi} نام می‌گیرد.
۳. مجموعه داده آموزش برای یافتن نقاط کاری که مشخصه راندمان آنها بهتر از مشخصه راندمان A_i یعنی A_{indi} است مورد جستجو قرار می‌گیرد. این نقاط کاری در مجموعه S_i گردآوری می‌شوند.
۴. الگوریتم K-Means بهبود یافته روی مجموعه S_i به شرح زیر اجرا می‌گردد:
 - I. ۱۰٪ از تعداد اعضای S_i به تصادف انتخاب و به عنوان مراکز خوشه‌ها یعنی C_j معرفی می‌شوند.
 - II. فاصله تک تک اعضای S_i با به کارگیری فرمول نرم اقلیدسی با تک تک مراکز خوشه‌ها سنجیده می‌شود. هر عضو به خوشه‌ای تعلق پیدا می‌کند که در کمترین فاصله با مرکز آن خوشه قرار دارد.

III. میانگین اعضای یک خوشه محاسبه شده و به عنوان مرکز جدید خوشه

معرفی می‌شود.

IV. گام II و III تکرار می‌شوند. در صورتی که با تکرار مجدد هیچ عضوی

بین دو خوشه جا به جا نشود و هر عضو در خوشه‌ای که به آن تعلق دارد

باقی بماند، الگوریتم متوقف می‌شود.

V. حاصل اجرای الگوریتم K خوشه با K مرکز $C_j, j=1, \dots, k$ بوده و هر

خوشه نمایش دهنده یک وضعیت کاری مطلوب برای بویلر است.

۵. فاصله A_i با هر C_j با معیار نرم اقلیدسی سنجیده می‌شود. نزدیک‌ترین C_j به

عنوان مرکز خوشه هدف یا TC_i انتخاب می‌شود.

۶. مقدار هر متغیر کنترل پذیر در A_i ، با مقدار متناظر در TC_i جایگزین می‌شود.

مقادیر متغیرهای کنترل ناپذیر A_i بدون تغییر باقی می‌مانند.

حاصل پیمودن الگوریتم فوق برای تمام داده‌های مجموعه آزمایش، یعنی وضعیت‌های فعلی

عملکرد بویلر منجر به مجموعه جدیدی از وضعیت‌های عملکرد بویلر خواهد شد.

۹-۴ - محاسبه فاصله

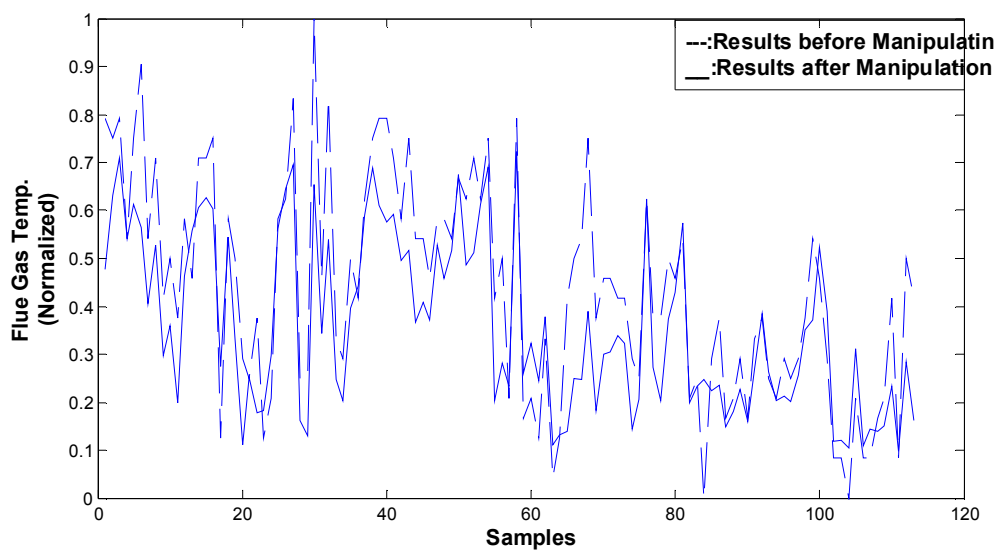
در گامهای II و III از الگوریتم کنترل هوشمند، کمیت "فاصله" با استفاده از مفهوم نرم اقلیدسی مورد محاسبه قرار می‌گیرد. با توجه به اینکه بنا بر اقتضای شرایط و خواسته‌ها در هر فرآیند ممکن است اهمیت متغیرها متفاوت باشد، امکان تعیین وزن برای متغیرها در ساختن مقدار فاصله وجود دارد. علاوه بر این در سنجش فاصله دو وضعیت کاری مطلوب و فعلی ممکن است تنها فاصله بین متغیرهای کنترل پذیر یا فاصله بین متغیرهای کنترل ناپذیر مورد توجه باشد. در الگوریتم اجرا شده در این پژوهش بردار فاصله بین تمام متغیرها اعم از کنترل پذیر یا کنترل ناپذیر و با اهمیت یکسان بر اساس رابطه نرم اقلیدسی محاسبه می‌شود، زیرا فاصله بین متغیرهای کنترل پذیر از جهت امکان نزدیک شدن وضعیت کاری فعلی به وضعیت مطلوب در عمل؛ و فاصله بین متغیرهای کنترل پذیر از جهت اهمیت میزان تلاش کنترلی مورد نیاز و دامنه تغییراتی که سیستم متحمل آن می‌شود اهمیت دارد.

۹-۵ - نتیجه به کارگیری الگوریتم کنترل هوشمند

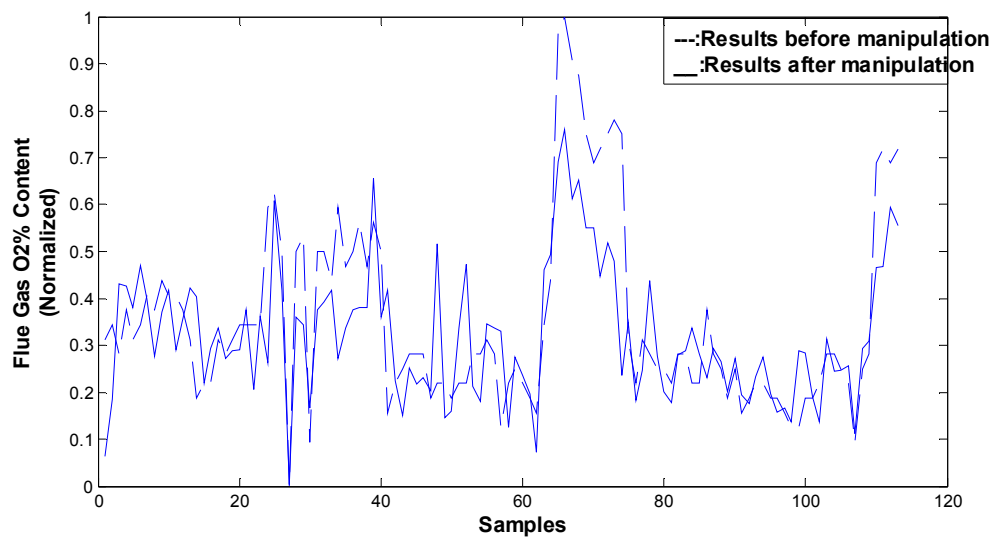
کارایی مجموعه جدید حاصل از اجرای الگوریتم با استفاده از مدل‌های به دست آمده در بخش قبل بر اساس شبکه‌های عصبی مورد ارزیابی قرار می‌گیرد.

مجموعه جدید به هر دو مدل وارد می‌شود. پاسخ مدل اکسیژن به مجموعه جدید در شکل (۹-۱) و پاسخ مدل دما به مجموعه جدید در شکل (۹-۲) نمایش داده شده است. با وجود خطای موجود

مدل بهبود حاصل در کارایی با مشاهده کاهش ملموس دمای گاز خروجی و اکسیژن موجود در گاز خروجی قابل ارزیابی است. معیار تعریف شده به صورت حاصل جمع نرمالیزه شده دما و اکسیژن، در مجموعه داده‌های اصلی بویلر برابر ۰/۸ و در مجموعه پاسخ مدل به داده‌های بهبود یافته ۰/۶۸ است.



شکل (۹-۱): پاسخ سیستم به ورودیهای تغییر یافته- دما



شکل (۲-۹) پاسخ سیستم به ورودی‌های تغییر یافته- اکسیژن

فصل ۱۰- نتیجه گیری و پیشنهادات

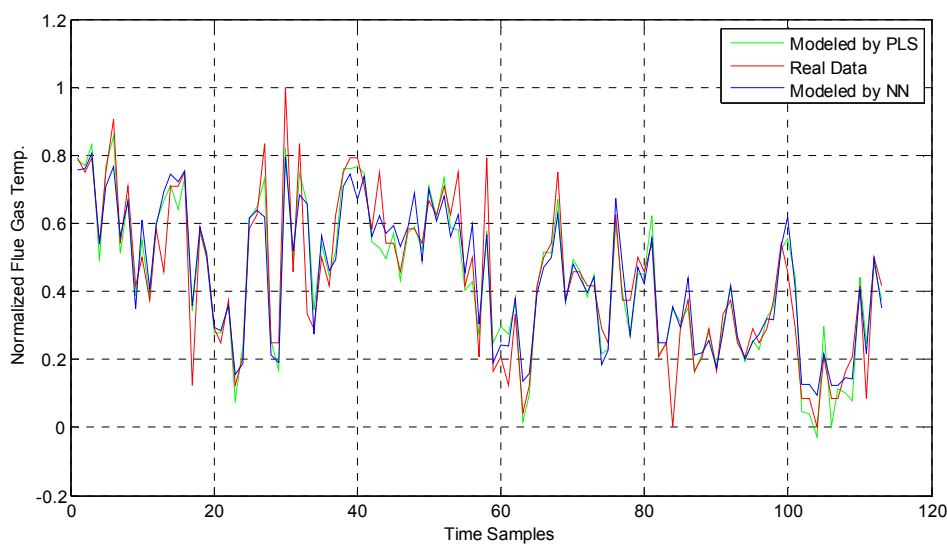
در این پژوهش بهبود کارایی فرآیند احتراق در یک بویلر نمونه با تکیه بر داده‌های بایگانی شده از عملکرد بویلر مورد بررسی قرار گرفت. به این منظور ابتدا مدلی از فرآیند مورد نظر در بازه مشخصی از بار نامی با استفاده از دو روش PLS و شبکه عصبی استخراج شد و سپس با به کار گیری الگوریتم کنترل هوشمند مبتنی بر داده کاوی و با تکیه بر مدل‌های استخراج شده، نشان داده شد که استفاده از بایگانی داده‌های فرآیند و جستجو برای بهترین و نزدیک‌ترین نقطه کار به نقطه کار فعلی و هدایت فرآیند به سمت قرار گرفتن در این نقاط کار، منجر به افزایش کارایی فرآیند خواهد شد. این پژوهش در دو بخش اصلی انجام پذیرفته است که نتایج هر بخش به تفکیک مورد بررسی قرار می‌گیرد:

۱۰-۱- نتایج بخش مدل سازی

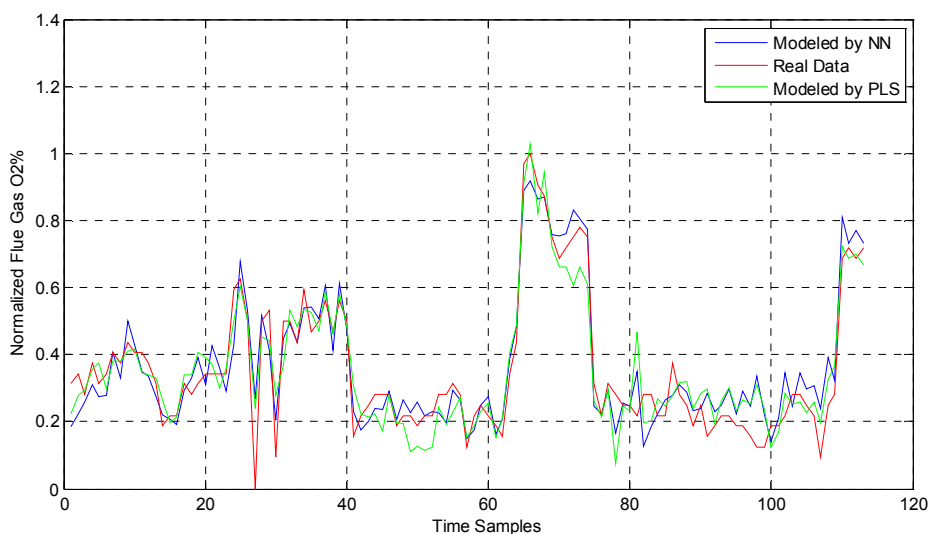
ابتدا روش PLS برای مدل کردن پاسخ سیستم برگزیده شد. این انتخاب به دلیل ماهیت روش PLS بود که با تکیه بر خاصیت "هم خطی" بودن‌های موجود بین متغیرها، ابعاد مجموعه داده‌ای را کاهش می‌دهد و به نوعی یک روش مدل سازی از خانواده روش‌های داده کاوی تلقی می‌شود. سپس روش شبکه عصبی به عنوان یک روش عمومی برای مدل سازی فرآیندهای غیر خطی مورد استفاده قرار گرفت. دو روش در تخمین دمای گاز خروجی از فرآیند عملکرد مشابهی داشتند؛ در حالی که

روش شبکه‌عصبی در تخمین اکسیژن موجود در گاز خروجی عملکرد بهتری در مقایسه با روش PLS داشت. با بررسی همبستگی خطی دو متغیر پاسخ دما و اکسیژن گاز خروجی با سایر متغیرها مشاهده می‌شود که همبستگی خطی دمای گاز خروجی با متغیرهای تخمین زنده خود بیشتر از میزان همبستگی خطی اکسیژن است.

استفاده از شبکه‌عصبی در تخمین پاسخ در شرایطی که خاصیت غیر خطی موجود در سیستم پررنگ تر بود، قابلیت بهتری در مقایسه با روش رگرسیون PLS از خود نشان داد. شکل‌های (۱-۱۰) و (۲-۱۰) نتایج تخمین دو پاسخ به دو روش را نمایش می‌دهند.



شکل (۱-۱۰): مقایسه PLS و شبکه‌عصبی در مدل‌سازی دما



شکل (۱۰-۲): مقایسه PLS و شبکه عصبی در مدل‌سازی اکسیژن

۱۰-۲- نتایج بخش کنترل

با به‌کارگیری الگوریتم خوشه‌سازی در بایگانی داده‌های فرآیند، الگوهای عملیاتی متناظر با راندمان بیشتر با جستجو در بخشی از تاریخچه عملکرد بویلر استخراج شدند. فاصله هر نقطه کاری جدید با الگوهای عملیاتی بهتر از خود سنجیده شد و نزدیکترین الگو به عنوان الگوی هدف انتخاب گردید. طی الگوریتمی مقادیر متغیرهای کنترل‌پذیر نقطه کاری جدید با مقادیر متغیرهای کنترل-پذیر الگوهای هدف جایگزین شد. نتیجه به دست آمده، بهبود معیار سنجش راندمان بود. این بهبود از کاهش دما و محتوی اکسیژن موجود در گاز خروجی حاصل شده بود که در تصاویر (۹-۱) و (۹-۲) نشان داده شده‌اند.

در بهره‌برداری واقعی از یک فرآیند صنعتی، می‌توان با محاسبه الگوهای هدف، اطلاعاتی در اختیار اپراتورهای مجموعه گذاشت که با به‌کارگیری آن فرآیند را به سمت وضعیت‌های کاری بهتر هدایت کنند. علاوه براین، با به‌روزشدن پایگاه داده الگوهای هدف، امکان پوشش شرایط جدید در عملکرد فرآیند، مثل آثار گذشت زمان و استهلاک و تغییر شرایط محیطی نیز وجود دارد.

۱۰-۳- پیشنهادات

- I. با توجه به اینکه پایگاه داده در اختیار در انجام این پژوهش به طور دستی ثبت و تنظیم شده بود، پیشنهاد می‌شود که در صورت در اختیار بودن پایگاه داده‌ای دقیق و با ثبت کامپیوتری از حجم زیادی از متغیرها، تمام مراحل مدل‌سازی و کنترل مجدداً تکرار شده و نتایج مورد بررسی قرار بگیرند.
- II. در کنار افزایش راندمان، اهداف کنترلی دیگری نیز مانند کاهش آلاینده‌ها در گاز خروجی یا کاهش تغییرات در مقدار فشار بخار خروجی که ثابت بودن آن مطلوب است دنبال شود.
- III. علاوه بر اهداف کنترلی، در یک مجموعه صنعتی پیگیری اهداف مدیریتی و برنامه ریزی نیز در صورت ثبت اطلاعات مورد نیاز با روش‌های داده‌کاوی امکان پذیر است؛ به عنوان مثال با بررسی تاریخچه عملکرد تجهیزات می‌توان زمان تعمیر آنها را پیش بینی کرد.

فهرست منابع

- 1 Wang, Xue Z., *Data Mining and Knowledge Discovery for Process Monitoring and Control*, Springer, 1999.
- 2 J. Q. Li, J.Z. Liu, Y.G. Niu, *Online Self-Optimization Control of Coal Fired Boiler Combustion System*, in Proc. IEEE TENCON, 2004, pp.589-592.
- 3 V. Havlena and J. Findejs, Application of Model Predictive Control to Advanced Combustion Control, *Control Eng. Practice*, Vol. 13, No. 6, pp.671-680, 2005.
- 4 R. Cass and B. Radl, Adaptive Process optimization using Functional Link Networks and Evolutionary Optimization, *Control Eng. Practice*, Vol. 4, No. 11, pp. 1579-1584, 1997.
- 5 Andrew Kusiak and Zhe Song, *Clustering-Based Performance Optimization of the Boiler-Turbine System*, *IEEE Transaction on Industrial Informatics*, Vol.2, No. 3, Aug. 2006.
- 6 Andrew Kusiak and Zhe Song, *Combustion Efficiency Optimization and Virtual Testing: A Data Mining Approach*, *IEEE Transaction on Energy Conversion*, Vol.23, No. 2, JUN.2008.
- 7 Zhe Song and Andrew Kusiak, *Constraint-Based Control of Boiler Efficiency: A Data Mining Approach*, *IEEE Transactions on Industrial Informatics*, Vol. 3, No.1, Feb. 2007.
- 8 R. Cass and B. Radl, Adaptive Process optimization using Functional Link Networks and Evolutionary Optimization, *Control Eng. Practice*, Vol. 4, No. 11, pp. 1579-1584, 1997.
- 9 Jian-qiang Li, Jun-jie Gu and Cheng-lin Niu, *The Operation Optimization Based on Correlation Analysis of Operation Parameters in Power Plant*, *Industrial Symposium on Computational Intelligence and Design*, 2008.
- 10 Tao Jin and Zhongguang Fu., *Data Mining for Complex Thermal System Modeling*, Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2008.
- 11 Dukelow, Sam G., *The Control of Boilers*, ISA (Instrument Society of America), 2nd Edition, 1991.
- 12 Ganapathy, V., *Industrial Boilers and Heat Recovery Steam Generators*, MARCEL DEKKER Inc., 2003.
- 13 رنچر، الوین. مدل‌های خطی برای آمار، ترجمه حسنعلی آذرنوش و ابولقاسم بزرگ نیا، انتشارات دانشگاه فردوسی مشهد، ۱۳۸۶.
- 14 Harald Martens and Tormod Næs, *Multivariate Calibration*, John Wiley & Sons Ltd., 1994.
- 15 Martin Andersson, *A Comparison of nine PLSI Algorithms*, *Journal of Chemometrics*, 23:518-529, 2009.
- 16 Madan M. Gupta, Liang Jin, and Noriyasu Homma, *Static and Dynamic Neural Networks*, John Wiley & Sons Ltd., 2003.
- 17 Jang, J.S Roger, Chuen-Tsai sun and Eiji Mizutani, *Neuro-Fuzzy and Soft Computing*, Prentice Hall, 1997.
- 18 Ian H. Witten, *Data Mining, Practical Machine Learning Tools and Techniques*, Elsevier, 2005
- 19 Gupta, G.K. , *Introduction to Data Mining with Case Studies*, Prentice Hall of India, 2008

20 Hand David, *Principles of Data Mining*, Prentice Hall of India, 2005

21 Larose Daniel T., *Data Mining, Methods and Models*, WILEY-Interscience, 2005.

پیوست ۱

نمونه‌ای از صفحات ثبت مقادیر گزارش‌گیری شده از عملکرد بویلر

پیوست ۲

ماتریس محاسبه شده از ضرایب همبستگی خطی موجود در میان متغیرها

