

دانشگاه صنعتی امیرکبیر

(پلی‌تکنیک تهران)

دانشکده مهندسی کامپیوتر

گزارش سمینار کارشناسی ارشد

مهندسی کامپیوتر-معماری سیستم‌های کامپیوتری

شتا بدھی سخت افزاری پیش‌بینی عمر باقی‌مانده مفید
دستگاه‌های دوار با استفاده از شبکه عصبی ترانسفورمر بر بستر
FPGA

نگارش

رضا آدینه پور

استاد راهنما

جناب آقای دکتر مرتضی صاحب‌الزمانی

۱۴۰۳ مرداد



اول

سپاس

از استاد بزرگوارم، جناب آقای دکتر مرتضی صاحب‌الزمانی، که به بنده اعتماد کردند و با کمک‌ها و راهنمایی‌های بی‌دریغ‌شان، مرا در به سرانجام رساندن این پایان‌نامه یاری دادند، سپاسگزاری و قدردانی می‌کنم. حمایت‌های مستمر و بی‌وقفه ایشان، نه تنها انگیزه و توان مضاعفی به من بخشید، بلکه مسیر پژوهشی این پایان‌نامه را با روشنایی دانش و تجربه‌شان هموار ساخت. بدون همراهی، مشورت‌ها و تشویق‌های استاد محترم، تحقق این هدف امکان‌پذیر نمی‌بود. از صمیم قلب برای ایشان آرزوی سلامتی، موفقیت و برکت دارم.

چنان مجنون شوم گویی که شهرآشوب دورانم
نه پندی و نه اندرزی به گوشم پنبه چسباندم
بماند یادگار این شعر
که من از دار این دنیا فقط این شعر می‌دانم.

چکیده

در محیط‌ها و کارخانه‌های صنعتی، همواره یکی از مهم‌ترین دغدغه‌ها، نگهداری و تعمیر ابزارآلات و دستگاه‌های صنعتی است. دستگاه‌هایی که هرکدام وظیفه‌ای مهم را در خط تولید هر شرکتی بر عهده دارند و کوچک‌ترین آسیب و خرابی می‌تواند خط تولید شرکت را مختل کند. بنابراین، نیاز است که همواره بتوانیم وضعیت فعلی سلامت دستگاه‌ها را مورد پایش قرار دهیم و بتوانیم زمان خرابی دستگاه و عمر مفید باقی‌مانده آن را پیش‌بینی نماییم.

پیش‌بینی عمر مفید باقی‌مانده مفید (Remaining Useful Life) یا به اختصار RUL وظیفه‌ای بسیار مهم در زمینه پیش‌بینی و مدیریت سلامت (PHM) ابزار و تجهیزات صنعتی است. پیش‌بینی دقیق RUL امری حیاتی و مهم است چرا که امکان تعمیر و نگهداری به موقع را فراهم می‌کند، مدت زمان خرابی دستگاه را کاهش می‌دهد و می‌توان بر اساس عمر مفید باقی‌مانده دستگاه، آن را تحت بار قرار داده و کارایی عملیاتی را بهبود بخشد.

کلیدواژه‌ها: عمر باقی‌مانده مفید، RUL، ترانسفرمر، FPGA

فهرست مطالب

۱	۱	مقدمه
۱	۱-۱	تعريف مسئله
۲	۲-۱	اهمیت موضوع
۲	۳-۱	اهداف پژوهش
۲	۴-۱	ساختار پایان نامه
۳	۲	مفاهیم اولیه
۳	۱-۲	عمر باقی مانده مفید
۴	۱-۱-۲	اولین زمان خرابی
۴	۲-۱-۲	عمر پایانی دستگاه
۵	۲-۲	داده ها
۶	۱-۲-۲	XJTU-SY مجموعه داده
۸	۲-۲-۲	PRONOSTIA مجموعه داده
۱۲	۳-۲-۲	C-MAPSS مجموعه داده
۱۳	۳-۲	شبکه عصبی ترانسفرمر
۱۴	۱-۳-۲	ساختار کلی شبکه
۱۴	۲-۳-۲	ورودی تعبیه شده
۱۵	۳-۳-۲	تعبیه موقعیتی

۱۷	توجه چند سر	۴-۳-۲
۲۰	لایه روبه جلو	۵-۳-۲
۲۱	لایه خطی رمزگشا	۶-۳-۲
۲۱	لایه خطی رمزگذار	۷-۳-۲
۲۲	کارهای پیشین	۳
۲۳	روش‌های مبتنی بر مدل	۱-۳
۲۴	روش‌های مبتنی بر داده	۲-۳
۲۵	روش‌های مبتنی بر یادگیری ماشین سنتی	۱-۲-۳
۲۶	روش‌های مبتنی بر یادگیری عمیق	۲-۲-۳
۳۵	چالش‌ها و نوآوری‌ها	۴
۳۵	تعریف دقیق مسئله	۱-۴
۳۵	راه حل پیشنهادی و ایده‌ی اولیه	۲-۴
۳۷	نتیجه‌گیری	۵
۳۹	مراجع	
۴۴	واژه‌نامه	

فهرست جداول

۱-۲	پارامترهای بلبرینگ‌های آزمایش شده	۷
۲-۲	اطلاعات مجموعه‌داده XJTU-SY	۸
۳-۲	تقسیم بندی داده‌های جمع‌آوری شده در مجموعه داده PRONOSTIA	۱۱
۴-۲	مشخصات داده‌های ضبط شده در فایل‌های csv	۱۱
۵-۲	اطلاعات زیرمجموعه‌ها در مجموعه‌داده C-MAPSS	۱۳

فهرست تصاویر

۱-۲	تابع تقریب زننده عمر باقی مانده یک دستگاه	۳
۲-۲	سیگنال ارتعاش بلبرینگ از ابتدای فعالیت تا تخریب کامل [۱]	۴
۳-۲	سیگنال RUL شکل [۱]، ۲-۲	۵
۴-۲	بستر تهیه مجموعه داده XJTU-SY [۲]	۶
۵-۲	تنظیمات نمونه برداری برای سیگنال های ارتعاشی	۷
۶-۲	سیگنال های ارتعاش افقی و عمودی و RUL بلبرینگ ۲-۱	۸
۷-۲	سیگنال های ارتعاش افقی و عمودی و RUL بلبرینگ ۲-۲	۹
۸-۲	بستر تهیه مجموعه داده PRONOSTIA [۳]	۹
۹-۲	تنظیمات نمونه برداری سیگنال ها در این آزمایش	۱۰
۱۰-۲	سیگنال خام تهیه شده در این آزمایش و بلبرینگ تخریب شده، [۳]	۱۰
۱۱-۲	چند سیگنال تصادفی خام جمع آوری شده توسط حسگرهای در موتور از مجموعه داده آموزشی	
۱۲	FD002 [۴]	
۱۲-۲	ساختار کلی شبکه ترانسفرمر	۱۴
۱۳-۲	تبديل کلمات به بردار	۱۵
۱۴-۲	خروجی گرافیکی بلوک تعییه موقعیتی به ازای $i = ۹۶$ و $d = ۴۸$	۱۶
۱۵-۲	جمع بردارهای ورودی با بردارهای موقعیتی تعییه شده	۱۶
۱۶-۲	عملیات ورودی تعییه شده و تعییه موقعیتی	۱۷
۱۷-۲	لایه توجه به خود	۱۸

۱۹	۱۸-۲ لایه خطی
۱۹	۱۹-۲ فرایند اعمال فیلتر توجه به تصویر
۱۹	۲۰-۲ فیلترهای توجه مختلف در تصویر
۲۰	۲۱-۲ لایه توجه چند سر
۲۱	۲۲-۲ لایه روبه جلو
۲۳	۱-۳ نتایج ارائه شده [۵]
۲۴	۲-۳ نمودار پیش‌بینی شده [۵]
۲۵	۳-۳ مدل ارائه شده مبتنی بر بردارهای پشتیبان در [۶]
۲۶	۴-۳ نتایج ارائه شده [۶]
۲۶	۵-۳ نتایج ارائه شده [۷]
۲۷	۶-۳ ویژگی‌های استفاده شده در [۸]
۲۷	۷-۳ نتایج ارائه شده در [۸]
۲۸	۸-۳ ساختار ارائه شده در [۹]
۲۸	۹-۳ نتایج ارائه شده در [۹]
۲۹	۱۰-۳ مدل ارائه شده در [۱۰]
۲۹	۱۱-۳ نتایج ارائه شده در [۱۰]
۳۰	۱۲-۳ ساختار ارائه شده در [۱۱]
۳۰	۱۳-۳ نتایج ارائه شده در [۱۱]
۳۱	۱۴-۳ ساختار ارائه شده در [۱۲]
۳۱	۱۵-۳ نتایج ارائه شده در [۱۲]
۳۲	۱۶-۳ معماری ارائه شده در [۱۳]
۳۳	۱۷-۳ خروجی پیش‌بینی ارائه شده در [۱۳]
۳۳	۱۸-۳ معماری ارائه شده در [۱۴]
۳۳	۱۹-۳ نتیجه ارائه شده در [۱۴]

۲۰-۳ ساختار ارائه شده در [۱۵]

۳۴

فصل ۱

مقدمه

پیش‌بینی عمر مفید باقی‌مانده^۱ (یا به اختصار RUL) یکی از مباحث کلیدی در حوزه مدیریت سلامت و پیش‌بینی^۲ (PHM) ابزارها و تجهیزات صنعتی است. RUL به مدت زمانی اشاره دارد که یک دستگاه یا ابزار قبل از رسیدن به نقطه خرابی و از کار افتادن نهایی، می‌تواند به طور مؤثر کار کند. این حوزه پژوهشی با استفاده از تکنیک‌های مختلف و پیشرفته در تلاش است تا به صنایع کمک کند تا بهره‌وری و کارایی خود را افزایش دهند و هزینه‌های ناشی از تعمیر و نگهداری غیرضروری را کاهش دهند.

۱-۱ تعریف مسئله

در صنایع مختلف، از جمله خودروسازی، هواپما، نفت و گاز و تولیدات صنعتی، تجهیزات و ماشین‌آلات به طور مداوم تحت شرایط کاری سخت و پیچیده قرار دارند. هرگونه خرابی ناگهانی این تجهیزات می‌تواند به وقفه‌های غیرمنتظره در تولید منجر شود که علاوه بر خسارات مالی، ممکن است اثرات زیان‌باری بر کیفیت محصول نهایی و رضایت مشتریان داشته باشد. بنابراین، نیاز است که وضعیت فعلی سلامت دستگاه‌ها به طور مستمر پایش شود و زمان خرابی دستگاه و عمر مفید باقی‌مانده آن با دقت بالایی پیش‌بینی شود. این پیش‌بینی نیازمند استفاده از تحلیل داده‌های حسگرها، مدل‌سازی ریاضی و الگوریتم‌های یادگیری ماشین^۳ و عمیق^۴ است.

¹Remaining Useful Life

²Prognostics and Health Management

³Machine Learning

⁴Deep Learning

۲-۱ اهمیت موضوع

اهمیت پیش‌بینی دقیق RUL در صنعت به دلیل تاثیر مستقیم آن بر بهره‌وری، کارایی و کاهش هزینه‌های تعمیر و نگهداری غیرضروری، به خوبی شناخته شده است. تکنیک‌های پیشرفته در حوزه هوش مصنوعی^۵ و تحلیل داده‌های بزرگ^۶ این امکان را فراهم کرده‌اند که داده‌های جمع‌آوری شده از تجهیزات به صورت بلادرنگ^۷ تحلیل شوند و مدل‌های پیش‌بینی RUL با دقت بالاتری ارائه شوند. این پیشرفتهای، به شرکت‌ها این امکان را می‌دهد که فرآیندهای نگهداری و تعمیرات خود را بهینه‌سازی کنند و تصمیم‌گیری‌های بهتری در زمینه مدیریت دارایی‌های خود انجام دهند.

۳-۱ اهداف پژوهش

این پژوهش بر توسعه یک راه حل شتابدهی سخت‌افزاری بر بستر FPGA پیش‌بینی RUL با استفاده از شبکه عصبی مصنوعی که وظیفه آن آموزش و یادگیری توالی و درنهایت پیش‌بینی آن است تمرکز دارد. با استفاده از قابلیت پردازش موازی FPGA و همچنین توان مصرفی بسیار پایین آن، هدف ما افزایش کارایی و مقیاس‌پذیری سیستم‌های پیش‌بینی RUL به ویژه برای دستگاه‌های دور است.

۴-۱ ساختار پایان‌نامه

این پایان‌نامه در پنج فصل به شرح زیر ارائه می‌شود. مفاهیم اولیه‌ی و نحوه‌ی کار شبکه عصبی ترانسفرمر در فصل «۲»^۸ اشاره شده است. فصل «۳»^۹ به مطالعه و بررسی کارهای پیشین مرتبط با پیش‌بینی عمر باقی‌مانده می‌پردازد. در فصل «۴»^{۱۰}، چالش‌های موجود در این پژوهش و راه حل‌های ما برای رفع این چالش‌ها ارائه شده است. فصل «۵»^{۱۱} به جمع‌بندی کارهای انجام شده در این پژوهش و ارائه‌ی پیشنهادهایی برای انجام کارهای آتی خواهد پرداخت.

⁵ Artificial Intelligence

⁶ Big Data Analytics

⁷ Real Time

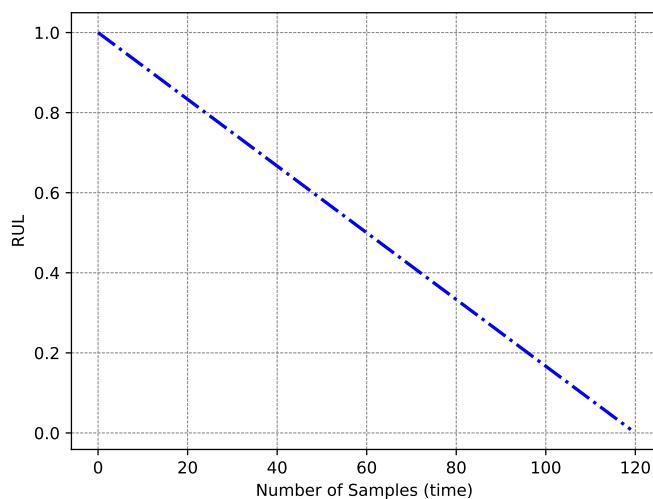
فصل ۲

مفاهیم اولیه

۱-۲ عمر باقی‌مانده مفید

مانند انسان‌ها، همه دستگاه‌ها و قطعات نیز عمری دارند و برای پایش سلامت دستگاه نیاز است که بتوانیم از عمر باقی‌مانده قطعه مطلع باشیم.

عمر یک قطعه را می‌توان به‌وسیله یک تابع خطی که آن را تابع RUL می‌نامیم «شکل ۳-۲» تقریب بزنیم.



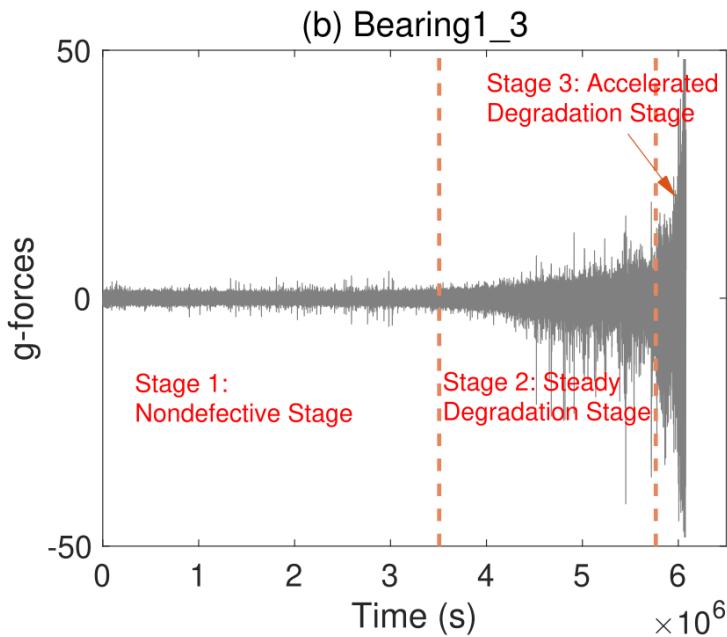
شکل ۱-۲: تابع تقریب زننده عمر باقی‌مانده یک دستگاه

محور عمودی در شکل «۳-۲» نشان‌دهنده میزان سلامت دستگاه و محور افقی نشان‌دهنده زمان است که معمولاً بر حسب دقیقه بیان می‌شود.

همه دستگاه‌ها زمانی که در آستانه بروز خطا و خرابی قرار می‌گیرند رفتار غیرعادی از خودشان نشان می‌دهند.

از جمله این رفتارها می‌توان به نوسانات غیرطبیعی، افزایش دمای دستگاه، افزایش سرعت^۱ در دستگاه اشاره نمود.

برای مثال در [۱] سیگنال ارتعاشات یک بلبرینگ به عنوان یکی از اصلی‌ترین قطعات صنعتی از ابتدای شروع به کار تا زمان بروز اولین تخریب^۲ و تخریب کامل^۳ در مدت زمان ۶۹ روز جمع‌آوری شده است. «شکل ۲-۲»



شکل ۲-۲: سیگنال ارتعاش بلبرینگ از ابتدای فعالیت تا تخریب کامل [۱]

۱-۱-۲ اولین زمان خرابی

اولین زمانی را که دستگاه دچار نوسانات شدید می‌شود را به عنوان اولین زمان شروع فرایند تخریب در نظر می‌گیریم و آن را «FPT» می‌نامیم.

۲-۱-۲ عمر پایانی دستگاه

با افزایش دامنه نوسانات ثبت شده از دستگاه، تخریب دستگاه بیشتر شده و دستگاه گرم‌تر می‌شود. از این فرایند به عنوان یک بازخورد^۴ مثبت یاد می‌شود که افزایش گرما، نوسانات را بیشتر کرده و نوسانات بیشتر نیز گرمای دستگاه را افزایش می‌دهد. با تشدید هرچه بیشتر نوسانات، دستگاه به پایان عمر خود نزدیک‌تر شده و درنهایت

¹Noise

²First Prediction Time

³End of Life

⁴Feedback

از کار می‌افتد. زمان از کار فتادن نهایی دستگاه را به عنوان زمان پایان زندگی «EOF» تعریف می‌کنیم.

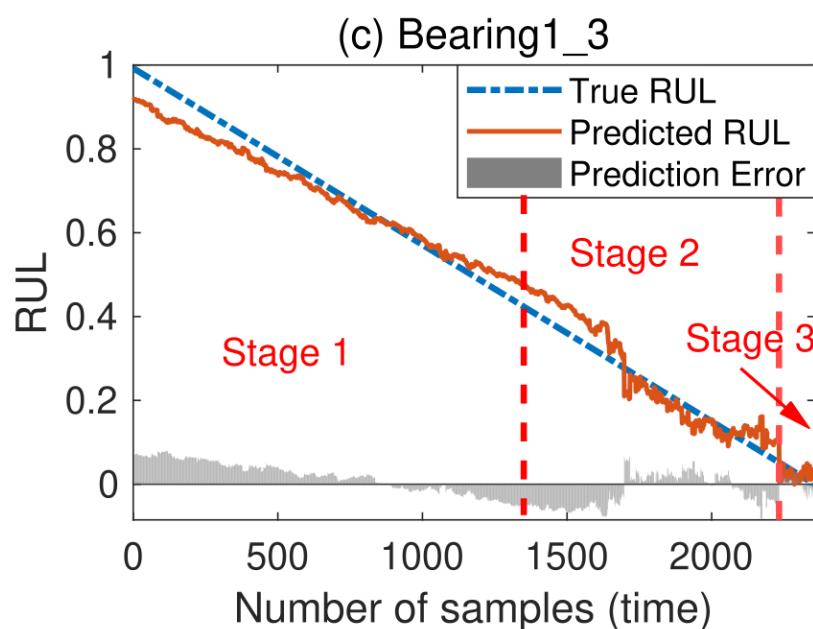
و درنهایت سیگнал RUL به صورت تفاضل این دوزمان تعریف می‌شود:

$$T_{RUL} = T_{EOF} - T_{FPT} \quad (1-2)$$

این سیگнал از جنس زمان است و مقدار RUL در این باز زمانی از رابطه زیر پیروی می‌کند:

$$RUL(t) = -t \quad (2-2)$$

برای مثال سیگнал RUL برای شکل «۲-۲» به صورت زیر محاسبه می‌شود:



شکل ۲-۳: سیگнал RUL شکل ۲-۲ [۱]

۲-۲ داده‌ها

برای پیش‌بینی عمر باقی‌مانده مفید، چندین مجموعه داده^۵ وجود دارد که در ادامه آنها را معرفی و بررسی می‌کنیم.

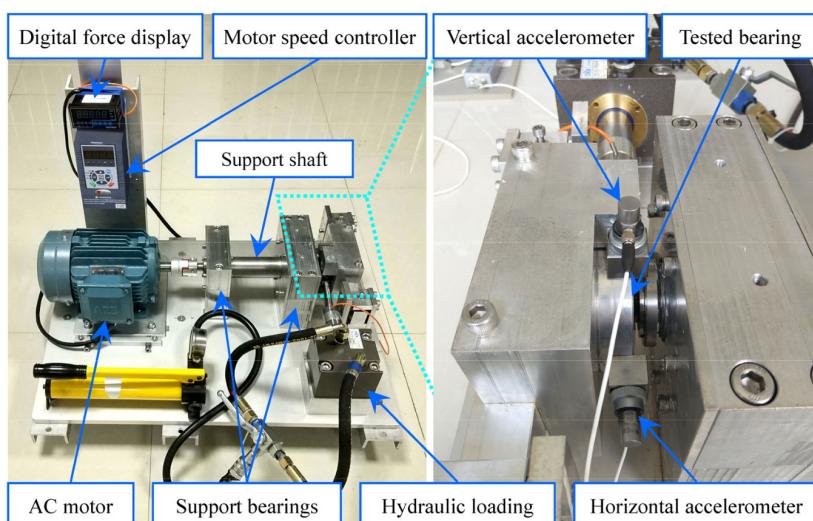
⁵Dataset

۱-۲-۲ مجموعه داده XJTU-SY

این مجموعه داده^۶ شامل داده های ثبت شده از ۱۵ بلبرینگ است که با انجام آزمایش های تخریب سریع، دچار تخریب شده اند. [۲]

این مجموعه داده توسط سیستمی که در شکل «۴-۲» نشان داده شده است، مشکل از یک موتور القایی جریان متناوب^۷، یک کنترل کننده سرعت موتور، یک محور^۸ پشتیبان، دو بلبرینگ پشتیبان (بلبرینگ های سنگین) و یک سیستم بارگذاری هیدرولیک تشکیل شده است.

این بستر آزمون برای انجام آزمایش های تخریب تسربی شده بلبرینگ، تحت شرایط مختلف عملیاتی (نیروی شعاعی و سرعت چرخشی مختلف) طراحی شده است. نیروی شعاعی توسط سیستم بارگذاری هیدرولیک تولید شده و به محفظه بلبرینگ های آزمایش شده اعمال می شود و سرعت چرخش نیز توسط کنترلر سرعت موتور القایی AC تنظیم و نگه داشته می شود.



شکل ۴-۲: بستر تهیه مجموعه داده XJTU-SY [۲]

بلبرینگ های مورد استفاده در این آزمایش از نوع LDK UER204 هستند که پارامترهای دقیق آنها در جدول «۱-۲» آورده شده است.

این آزمایش، تحت ۳ شرط عملیاتی مختلف انجام شده است و هر ۵ بلبرینگ موجود در این آزمایش تحت این سه شرط عملیاتی قرار گرفته اند. این شرایط عملیاتی شامل موارد زیر هستند:

• ۲۱۰۰ دور در دقیقه^۹ (۳۵ هرتز) و بار دینامیکی ۱۲ کیلونیوتن

^۶ می توانید این مجموعه داده را از اینجا دریافت کنید: biaowang.tech/xjtu-sy-bearing-datasets/

⁷ Alternating Current

⁸ Shaft

⁹ RPM

جدول ۱-۲: پارامترهای بلبرینگ‌های آزمایش شده

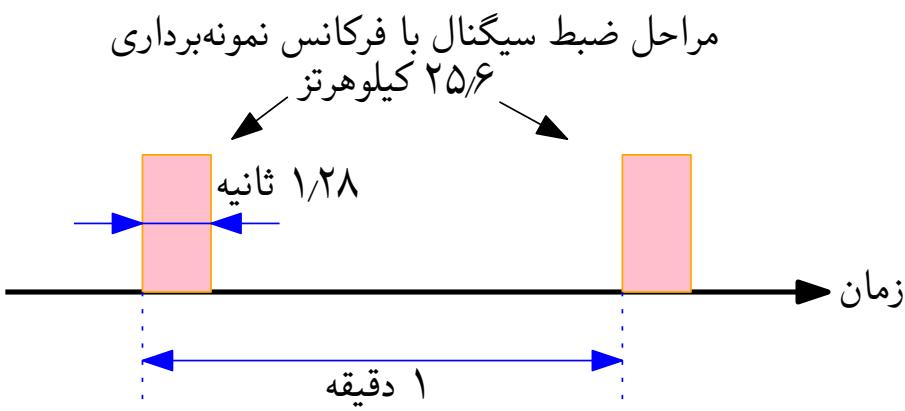
پارامتر	مقدار	پارامتر	مقدار	مقدار
قطر مسیر بیرونی	۳۹/۸۰ mm	قطر مسیر داخلی	۲۹/۳۰ mm	۷۹۲ mm
قطر متوسط بلبرینگ	۳۴/۵۵ mm	قطر توب	۷۹۲ mm	۰°
تعداد توب‌ها	۸	زاویه تماس	۸	۱۲۸۲ kN
بار استاتیک	۶,۶۵ kN	بار دینامیک	۱۲۸۲ kN	

• ۲۲۵° دور در دقیقه (۳۷/۵ هرتز) و بار دینامیکی ۱۱ کیلونیوتون

• ۲۴۰° دور در دقیقه (۴۰ هرتز) و بار دینامیکی ۱۰ کیلونیوتون

برای جمع‌آوری سیگنال‌های ارتعاشی بلبرینگ‌های آزمایش شده، همان‌طور که در شکل «۴-۲» نشان داده شده است، دو شتاب‌سنج از نوع PCB۳۵۲۰۳۳ در زاویه ۹۰ درجه بر روی محفظه بلبرینگ‌های آزمایش شده قرار داده شده است، یعنی یکی بر روی محور افقی و دیگری بر روی محور عمودی نصب شده است.

همچنین فرکانس نمونه‌برداری بر روی ۲۵/۶ کیلوهرتز تنظیم شده است. همان‌طور که در شکل «۵-۲» نشان داده شده است، در مجموع ۳۲۷۶۸ نقطه داده (به مدت ۱/۲۸ ثانیه) برای هر نمونه‌برداری ثبت می‌شوند و دوره نمونه‌برداری برابر با ۱ دقیقه است.



شکل ۲-۵: تنظیمات نمونه‌برداری برای سیگنال‌های ارتعاشی

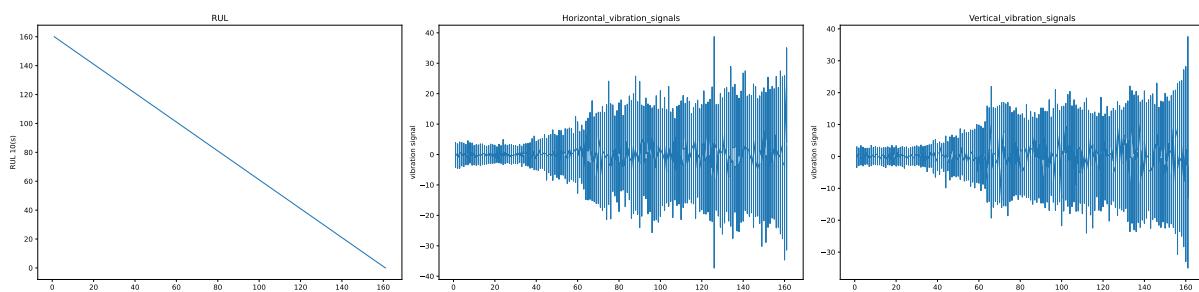
برای هر نمونه‌برداری، داده‌های به‌دست‌آمده در یک فایل csv ذخیره شده است که در آن ستون اول، سیگنال‌های ارتعاشی افقی و ستون دوم سیگنال‌های ارتعاشی عمودی را شامل می‌شود. جدول «۲-۲» اطلاعات دقیق هر بلبرینگ آزمایش شده، شامل تعداد فایل‌های csv، عمر بلبرینگ و عنصر خواری را فهرست می‌کند.

نقشه EOF در این مجموعه داده $A_H \times 10^\circ$ در نظر گرفته شده است که A_H بیشینه دامنه سیگنال ارتعاش عمودی و افقی در حالت کاری طبیعی است.

جدول ۲-۲: اطلاعات مجموعه داده XJTU-SY

عامل خطا	مجموعه داده های بلبرینگ	طول عمر بلبرینگ	تعداد فایل ها	شرایط عملکرد
بیرونی	بلبرینگ ۱	۲ ساعت و ۳ دقیقه	۱۲۳	شایط ۱ ۳۵ هرتز، ۱۲ کیلونیوتون)
بیرونی	بلبرینگ ۲	۲ ساعت و ۴۱ دقیقه	۱۶۱	
بیرونی	بلبرینگ ۳	۲ ساعت و ۳۸ دقیقه	۱۵۸	
قفسه	بلبرینگ ۴	۲ ساعت و ۲ دقیقه	۱۲۲	
داخلی و بیرونی	بلبرینگ ۵	۵ دقیقه	۵۲	۱
داخلی	بلبرینگ ۱_۲	۸ ساعت و ۱۱ دقیقه	۴۹۱	شایط ۲ ۳۷/۵ هرتز، ۱۱ کیلونیوتون)
بیرونی	بلبرینگ ۲_۲	۲ ساعت و ۴۱ دقیقه	۱۶۱	
قفسه	بلبرینگ ۳_۲	۸ ساعت و ۵۳ دقیقه	۵۳۳	
بیرونی	بلبرینگ ۴_۲	۴۲ دقیقه	۴۲	
بیرونی	بلبرینگ ۵_۲	۵ ساعت و ۳۹ دقیقه	۳۳۹	
بیرونی	بلبرینگ ۱_۳	۴۲ ساعت و ۱۸ دقیقه	۲۵۳۸	شایط ۳ ۴۰ هرتز، ۱۰ کیلونیوتون)
داخلی، ساچمه، قفسه و بیرونی	بلبرینگ ۲_۳	۴۱ ساعت و ۳۶ دقیقه	۲۴۹۶	
داخلی	بلبرینگ ۳_۳	۶ ساعت و ۱۱ دقیقه	۳۷۱	
داخلی	بلبرینگ ۴_۳	۲۵ ساعت و ۱۵ دقیقه	۱۵۱۵	
بیرونی	بلبرینگ ۵_۳	۱ ساعت و ۵۴ دقیقه	۱۱۴	

در شکل های «۶-۲» و «۷-۲» نمونه ای از داده های این مجموعه داده آورده شده است.



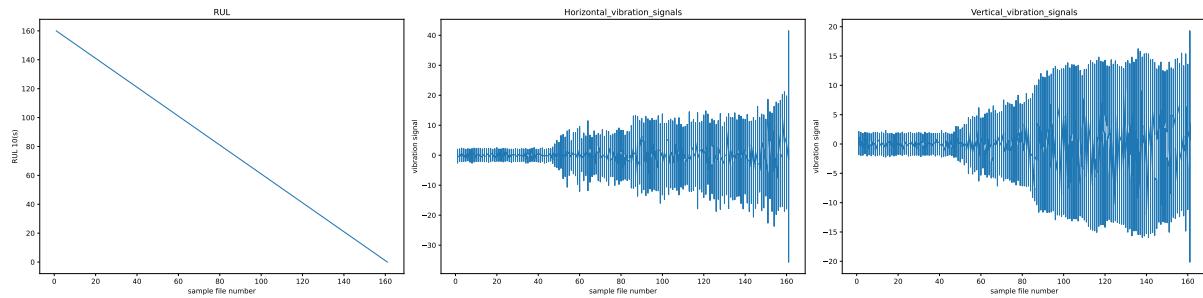
شکل ۲-۶: سیگنال های ارتعاش افقی و عمودی و RUL بلبرینگ ۱

۲-۲-۲ مجموعه داده PRONOSTIA

همانند مجموعه داده XJTY-SY این مجموعه داده^{۱۰} نیز، داده های ثبت شده از بلبرینگ توسط بستر آزمایشی نشان داده شده در شکل «۸-۲» است و امکان انجام آزمایش از ابتدای کار بلبرینگ تا زمان خرابی و شکست

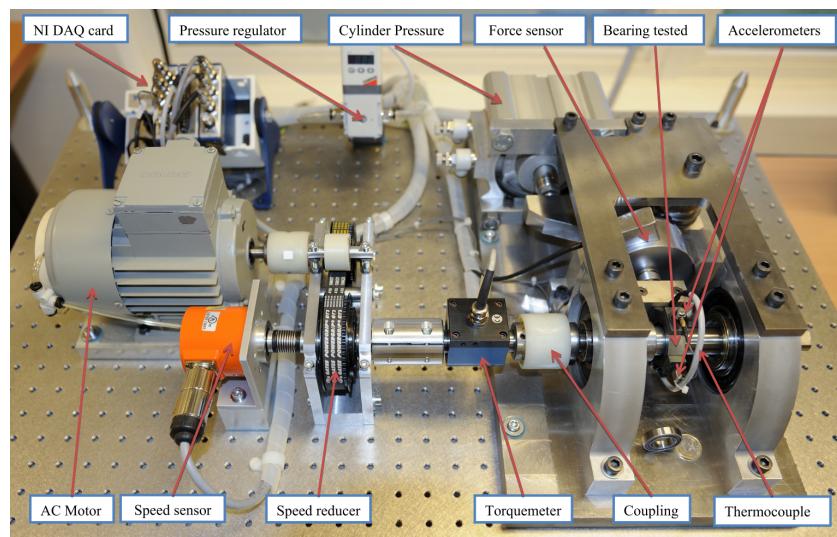
^{۱۰} می توانید این مجموعه داده را از اینجا دریافت کنید:

github.com/Lucky-Loek/ieee-phm-2012-data-challenge-dataset.git



شکل ۷-۲: سیگنال‌های ارتعاش افقی و عمودی و RUL ببلبرینگ ۲-۲

کامل را فراهم می‌آورد. [۳]



شکل ۸-۲: بستر تهیه مجموعه داده [۳]، PRONOSTIA

در این آزمایش، شرایط عملیاتی با اندازه‌گیری‌های لحظه‌ای ۱) نیروی شعاعی^{۱۱} اعمال شده بر ببلبرینگ، ۲) سرعت چرخش شفتی که ببلبرینگ را جایه‌جا می‌کند و ۳) گشتاور اعمال شده به ببلبرینگ تعیین می‌شوند. هر یک از این سه اندازه‌گیری که به صورت آنالوگ انجام شده است، با فرکانسی برابر با ۱۰۰ هرتز به دست آمده است.

در این مجموعه داده، از دو نوع حسگر^{۱۲} برای تشخیص و توصیف عملیات تخریب استفاده شده است. حسگرهای:

- ارتعاش

- دما

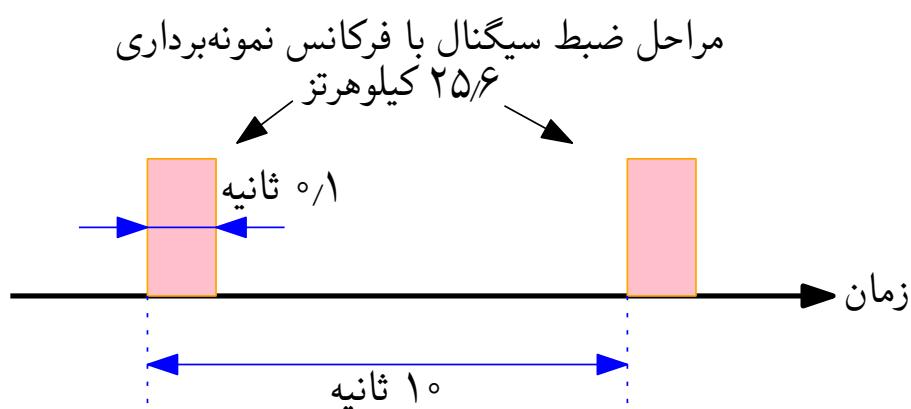
¹¹Radial Force

¹²Sensor

همانند مجموعه‌داده XJTU-SY، حسگرهای ارتعاش در این مجموعه آزمایشی نیز شامل دو شتاب‌سنج هستند که با زاویه 90° درجه نسبت به هم قرار دارند (اولی به صورتی عمودی و دومی به صورت افقی). این دو شتاب‌سنج به صورت شعاعی روی مسیر خارجی بلبرینگ قرار گرفته‌اند.

در این مجموعه آزمایشی علاوه بر حسگرهای ارتعاش، حسگرهای دما نیز وجود دارد. حسگر دما یک پروب از جنس پلاتینیم است که در داخل سوراخی نزدیک به حلقه خارجی بلبرینگ قرار می‌گیرد. RTD

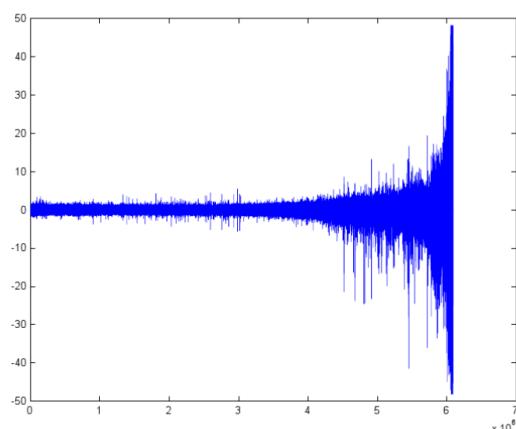
حسگرهای شتاب، نمونه‌ها را با فرکانس $25/6$ کیلوهرتز و حسگر دما با فرکانس 1° هرتز ثبت می‌کنند. یعنی سیگنال‌های ارتعاشی 256° نمونه ($\frac{1}{10}$ ثانیه) در هر 1° ثانیه ضبط می‌شوند و سیگنال‌های دما، 600° نمونه هر دقیقه ضبط می‌شوند.



شکل ۹-۲: تنظیمات نمونه‌برداری سیگنال‌ها در این آزمایش

در تهیه این مجموعه داده، برای جلوگیری از انتشار آسیب به کل بستر آزمایش، آزمایش‌ها زمانی متوقف شدند که دامنه سیگنال ارتعاش از 20° فراتر رفت.

شکل «۱۰-۲» مثالی است از آنچه می‌توان بر روی اجزای بلبرینگ قبل و بعد از یک آزمایش مشاهده کرد و همچنین یک سیگنال خام ارتعاش جمع‌آوری شده در طول یک آزمایش کامل را نشان می‌دهد.



شکل ۱۰-۲: سیگنال خام تهیه شده در این آزمایش و بلبرینگ تخریب شده، [۳]

ذکر این نکته الزامی است که بلبرینگ‌ها رفتارهای بسیار متفاوتی را در زمان تخریب نشان می‌دهد که منجر به تفاوت در زمان آزمایش و تخریب می‌شود.

این آزمایش نیز تحت سه شرط عملیاتی مختلف انجام شده است:

- ۱۸۰° دور در دقیقه و بار دینامیکی ۴ کیلونیوتن

- ۱۶۵° دور در دقیقه و بار دینامیکی ۴/۲ کیلونیوتن

- ۱۵۰° دور در دقیقه و بار ۵ کیلونیوتن

داده‌های جمع‌آوری شده از این آزمایش به صورت جدول «۳-۲» تقسیم‌بندی می‌شود
جدول ۳-۲: تقسیم‌بندی داده‌های جمع‌آوری شده در مجموعه داده PRONOSTIA

مجموعه داده‌ها	شرایط عملیاتی ۱	شرایط عملیاتی ۲	مجموعه داده‌ها
مجموعه آموزشی	بلبرینگ ۱_۳	بلبرینگ ۱_۲	بلبرینگ ۱_۱
	بلبرینگ ۲_۳	بلبرینگ ۲_۲	بلبرینگ ۲_۱
مجموعه آزمون	بلبرینگ ۳_۳	بلبرینگ ۳_۲	بلبرینگ ۳_۱
		بلبرینگ ۴_۲	بلبرینگ ۴_۱
		بلبرینگ ۵_۲	بلبرینگ ۵_۱
		بلبرینگ ۶_۲	بلبرینگ ۶_۱
		بلبرینگ ۷_۲	بلبرینگ ۷_۱

هر دو مجموعه داده‌های آموزشی و آزمایشی حاوی فایل‌هایی csv ارتعاش به نام acc_xxxxxx.csv و فایل‌های دما به نام temp_xxxxxx.csv هستند.

همه فایل‌های csv ضبط شده از قالب زیر تبعیت می‌کنند:
جدول ۴-۲: مشخصات داده‌های ضبط شده در فایل‌های csv

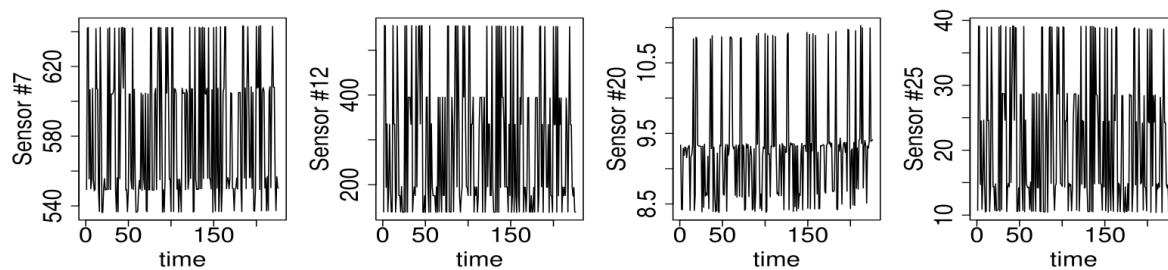
ستون	۱	۲	۳	۴	۵	۶
سیگنال ارتعاش	ساعت	دقیقه	ثانیه	میکروثانیه	حسگر افقی	حسگر عمودی
سیگنال دما	ساعت	دقیقه	ثانیه	x/۰	حسگر RTD	

۳-۲-۲ مجموعه داده C-MAPSS

برخلاف دو مجموعه داده قبلی، این مجموعه داده^{۱۳}، داده‌های تهیه شده از موتور هوایی جت از ابتدای کارکرد تا زمان تخریب است.^[۱۶]

این مجموعه داده، شامل چندین سری زمانی چندمتغیره است. هر مجموعه داده به زیرمجموعه‌های آموزشی و آزمایشی تقسیم می‌شود. داده‌ها از موتورهای متفاوتی تهیه شده است. هر موتور با درجات مختلفی از سایش اولیه و تغییرات تولیدی که برای کاربر ناشناخته است، شروع به کار می‌کند. این سایش و تغییرات به عنوان وضعیت طبیعی در نظر گرفته می‌شوند و به عنوان شرایط خطا تلقی نمی‌شوند.

در این مجموعه داده، سه تنظیمات عملیاتی وجود دارد که تأثیر قابل توجهی بر عملکرد موتور دارند و جمع‌آوری داده‌ها بر اساس این تنظیمات انجام شده است. برخلاف دو مجموعه داده قبلی، داده‌های این مجموعه به نویز حسگرهای آلوود شده است.



شکل ۱۱-۲: چند سیگنال تصادفی خام جمع‌آوری شده توسط حسگرهای در موتور از مجموعه داده آموزشی FD002^[۱۴]

موتور در ابتدای هر سری زمانی به طور عادی عمل می‌کند و در یک نقطه از سری دچار خطا می‌شود. در مجموعه آموزشی، خطا بهشدت افزایش می‌یابد تا زمانی که سیستم به طور کامل تخریب شود. اما در مجموعه آزمایشی، سری زمانی کمی قبل از خرابی سیستم خاتمه می‌یابد.

داده‌های جمع‌آوری شده توسط ۲۶ حسگر، به صورت یک فایل متنه فشرده شده با فرمت txt ارائه شده است؛ بنابراین هر فایل دارای ۲۶ ستون است که هر کدام از این ستون‌ها به صورت زیر معرفی می‌شود:

۱. شماره واحد

۲. زمان (برحسب تعداد چرخه^{۱۵})

۳. تنظیمات عملیاتی

^{۱۳} می‌توانید این مجموعه داده را از اینجا دریافت کنید:

data.nasa.gov/Aerospace/CMAPSS-Jet-Engine-Simulated-Data/ff5v-kuh6/about_data

^{۱۴}Cycle

۴. تنظیمات عملیاتی ۲

۵. تنظیمات عملیاتی ۳

۶. مقدار اندازه‌گیری شده توسط حسگر شماره ۱

۷. مقدار اندازه‌گیری شده توسط حسگر شماره ۲

۸. مقدار اندازه‌گیری شده توسط حسگر شماره ۳

... ۹

۱۰. مقدار اندازه‌گیری شده توسط حسگر شماره ۲۶

در جدول «۵-۲» اطلاعات دقیقی از این مجموعه‌داده آورده شده است.

جدول ۵-۲: اطلاعات زیرمجموعه‌ها در مجموعه‌داده C-MAPSS

FD004	FD003	FD002	FD001	
۲۴۹	۱۰۰	۲۶۰	۱۰۰	تعداد موتورها در مجموعه آموزش
۲۴۸	۱۰۰	۲۵۹	۱۰۰	تعداد موتورها در مجموعه آزمون
۶	۱	۶	۱	تعداد شرایط عملیاتی
۲	۲	۱	۱	تعداد حالت‌های خطا

۳-۲ شبکه عصبی ترانسفرمر

برای اولین بار در سال ۲۰۱۷ مفهومی به نام مکانیزم توجه^{۱۵} در مقاله‌ای تحت عنوان «Attention is All» در مکانیزم توجه^{۱۶} در مقاله‌ای تحت عنوان «You Need» توسط شرکت گوگل با کاربرد در زمینه پردازش زبان طبیعی^{۱۷} معرفی شد [۱۷].

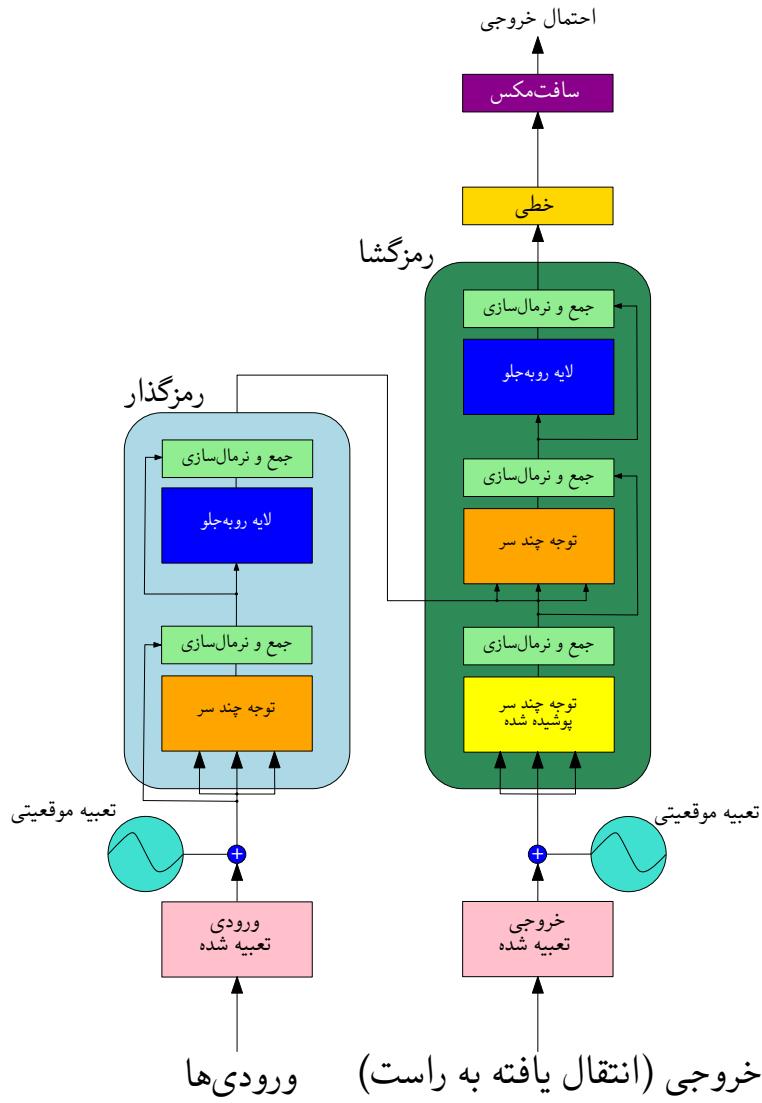
در این مقاله، ساختار جدیدی از یک شبکه عصبی نیز معرفی شد که بر پایه مکانیزم توجه کار می‌کند و توانست خیلی زود جایگزین مناسبی برای مدل‌های سنتی قدیمی و ساختارهای RNN شود. در ادامه به بررسی جزئی معماری این ساختار می‌پردازیم.

¹⁵Attention

¹⁶Natural Language Processing

۱-۳-۲ ساختار کلی شبکه

شبکه ترانسفرمر از یک ساختار رمزگذار^{۱۷} - رمزگشا^{۱۸} پیروی می‌کند. شکل ۱۲-۲



شکل ۱۲-۲: ساختار کلی شبکه ترانسفرمر

۲-۳-۲ ورودی تعابیه شده

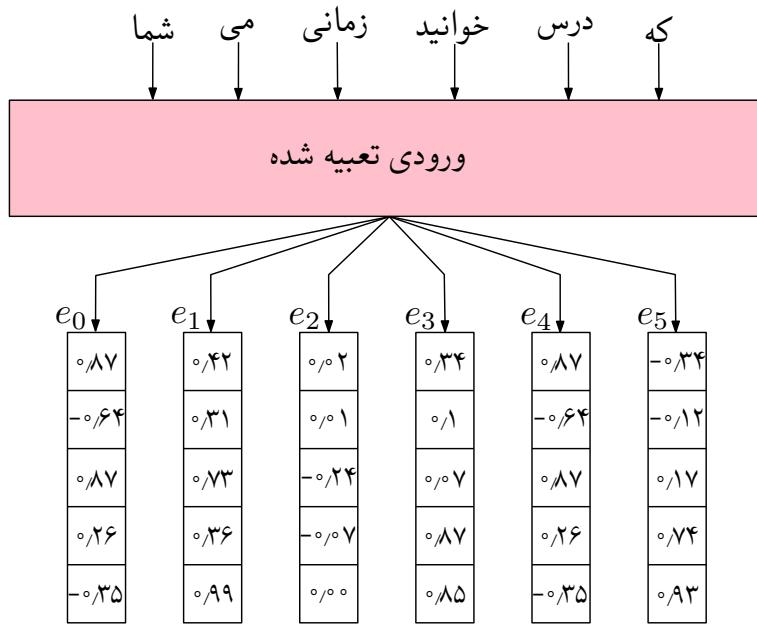
برای توضیح این قسمت از شبکه، فرض می‌شود که می‌خواهیم یک جمله ناقص را به عنوان ورودی به شبکه بدهیم و شبکه باید آن را تکمیل کند. برای مثال فرض شود که جمله ورودی ما به صورت زیر است:

”زمانی که شما درس می‌خوانید...“

¹⁷Encoder

¹⁸Decoder

اولین مرحله در تمامی الگوریتم‌های یادگیری ماشین که ورودی آن‌ها کلمه است، تبدیل کلمات به بردارهای عددی است. وظیفه این مرحله با بلوک ورودی تعییه شده است. «شکل ۱۳-۲»



شکل ۱۳-۲: تبدیل کلمات به بردار

۳-۳-۲ تعییه موقعیتی

برخلاف سایر شبکه‌ها مانند RNN و LSTM که ورودی‌ها به صورتی ترتیبی و پشت‌سرهم وارد می‌شوند، در شبکه ترنسفرمر، ورودی‌ها به صورت موازی باهم و همزمان وارد لایه تعییه شده می‌شوند؛ بنابراین نیاز است که ترتیب و جایگاه ورودی‌ها نسبت به یکدیگر مشخص باشد. ترتیب ورودی‌ها بر حسب دوتابع سینوسی و کسینوسی بر حسب رابطه‌های زیر تعیین می‌شود.

$$PE_{(pos, 2i)} = \sin\left(\frac{pos}{1000 \frac{2i}{d}}\right) \quad (3-2)$$

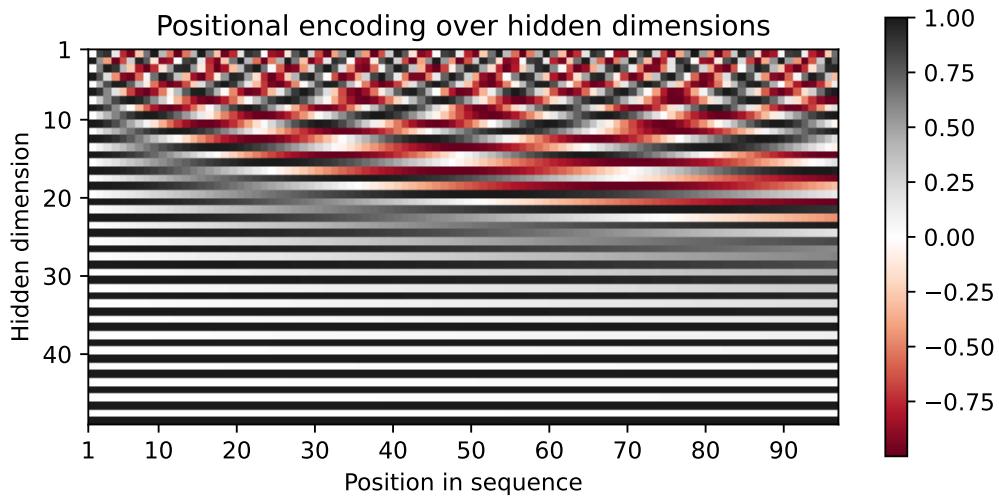
$$PE_{(pos, 2i+1)} = \cos\left(\frac{pos}{1000 \frac{2i+1}{d}}\right) \quad (4-2)$$

ورودی‌هایی با نمایه^{۱۹} «زوج، از رابطه ۳-۲» و نمایه فرد از رابطه «۴-۲» استفاده می‌کنند.

در روابط تعییه موقعیتی، مقدار pos برابر است با جایگاه کلمه، مقدار d برابر است با طول بردار تعییه شده (در مثال ما $d = 5$) و مقدار i برابر است با جایگاه‌های هر بردار.

¹⁹Index

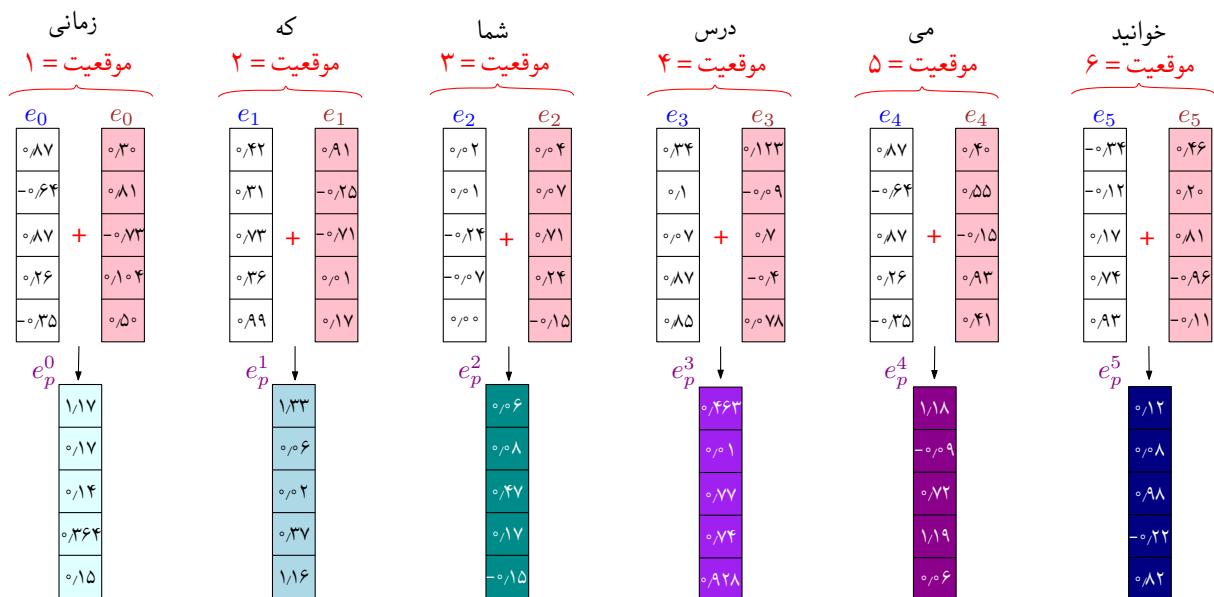
با رسم روابط تعییه موقعیتی، خروجی ای مانند شکل «۱۴-۲» حاصل می‌شود.



شکل ۱۴-۲: خروجی گرافیکی بلوک تعییه موقعیتی به ازای $d = 48$ و $i = 96$

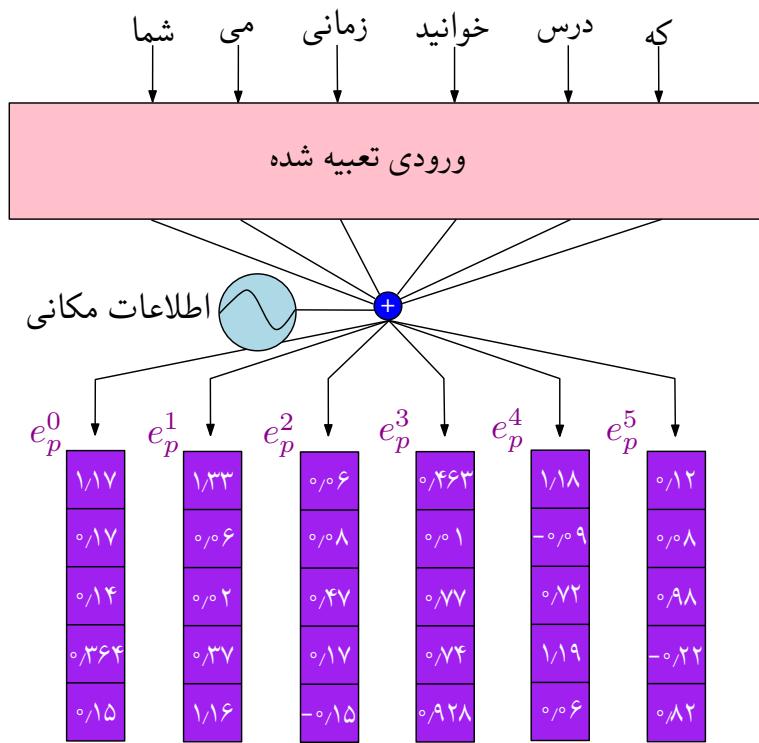
سپس مقادیر ساخته شده توسط روابط لایه تعییه موقعیتی، با بردارهای ورودی جمع می‌شوند. «شکل ۱۵-۲»

«۱۵-۲



شکل ۱۵-۲: جمع بردارهای ورودی با بردارهای موقعیتی تعییه شده

و درنهایت می‌توان عملکرد دو بلوک (۱) ورودی تعییه شده و (۲) تعییه موقعیتی را به صورت شکل «۱۶-۲» نمایش داد.



شکل ۱۶-۲: عملیات ورودی تعبیه شده و تعبیه موقعیتی

۴-۳-۲ توجه چند سر

mekanizm توجه، به مدل کمک می‌کند تا به ورودی‌های مهم توجه بیشتری شود. بلوک توجه چند سر^{۲۰} متشکل است از چندین لایه توجه به خود^{۲۱}. وظیفه این لایه به دست آوردن و درک وابستگی‌های مهم در ورودی است. برای مثال فرض شود ورودی شبکه جمله زیر است:

”شب‌هنگام که ماه را در آسمان دیدم، به یاد ماه رویت افتادم.“

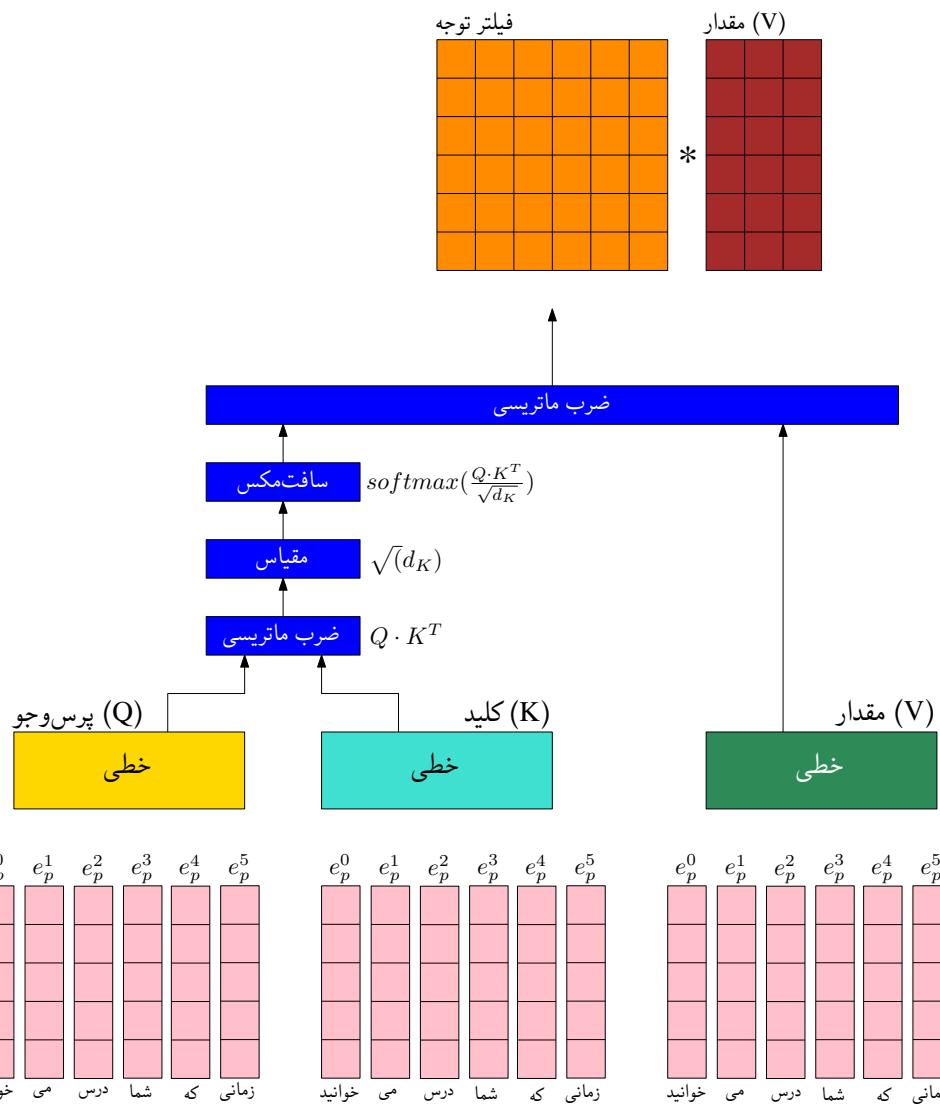
در این جمله دو بار کلمه ”ماه“ به‌کاررفته است که اولین آن به معنی قمر آسمانی و دومین آن به معنی زیبایی است.

مدل چگونه باید تشخیص دهد که کدام ”ماه“ به معنی قمر و کدام به معنی زیبایی است؟ لایه توجه به خود در اینجا به کمک می‌آید و وابستگی کلمه ”ماه“ را با سایر کلمات جمله بررسی می‌کند و بر اساس نقشه توجهی که به دست می‌آورد می‌تواند تشخیص دهد که هر کدام از ”ماه“‌ها به چه معناست.

عملیاتی که برای به‌دست آوردن نقشه‌های توجه انجام می‌شود، در شکل ۱۷-۲ خلاصه می‌شود.

²⁰Multi-Head Attention

²¹Self-Attention



شکل ۱۷-۲: لایه توجه به خود

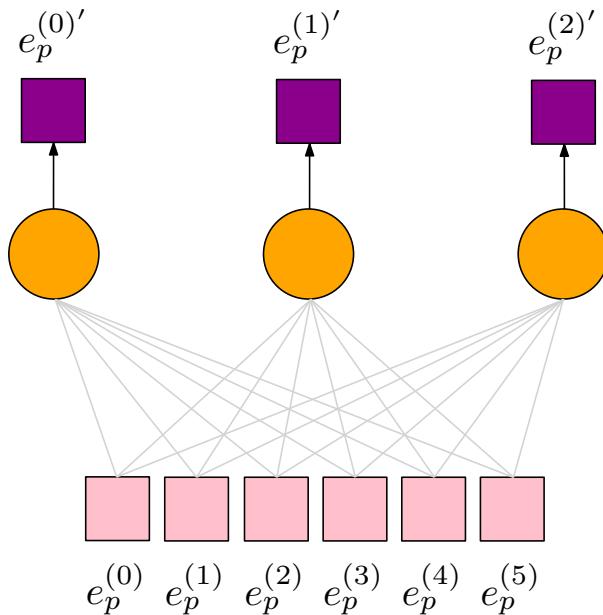
در ابتدا از بردارهای به دست آمده از لایه‌های تعبیه شده،^۳ نمونه می‌سازیم. بردارهای به دست آمده برای هر کلمه، به یک لایه شبکه عصبی خطی با تعداد نورون‌های دلخواه وارد می‌شود. «شکل ۱۸-۲» پس از آموزش شبکه خطی و به دست آمدت بهترین وزن‌ها برای شبکه، خروجی سه شبکه خطی را به ترتیب ماتریس‌های پرس‌وجو^{۲۲} (Q)، کلید^{۲۳} (K) و مقدار^{۲۴} (V) می‌نامیم. معیار توقف آموزش شبکه خطی، میزان شباهت بین ماتریس‌ها است که از رابطه «۵-۲» می‌توان آن را به دست آورد:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_K}}\right) \cdot V \quad (5-2)$$

²²Query

²³Key

²⁴Value



شکل ۱۸-۲: لایه خطی

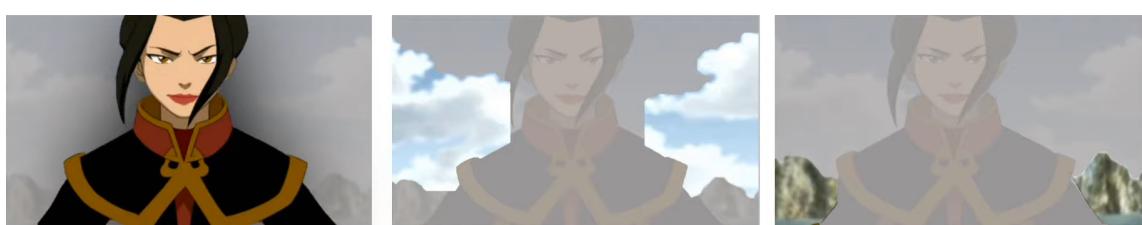
و در نهایت فیلتر توجه به دست آمده را در ماتریس V ضرب می‌کنیم و مقدار فیلترها به دست می‌آید.

شکل «۱۹-۲» مثالی از یک فیلتر توجه و در نهایت ضرب فیلتر در ورودی و ایجاد تصویر نهایی را نمایش می‌دهد.

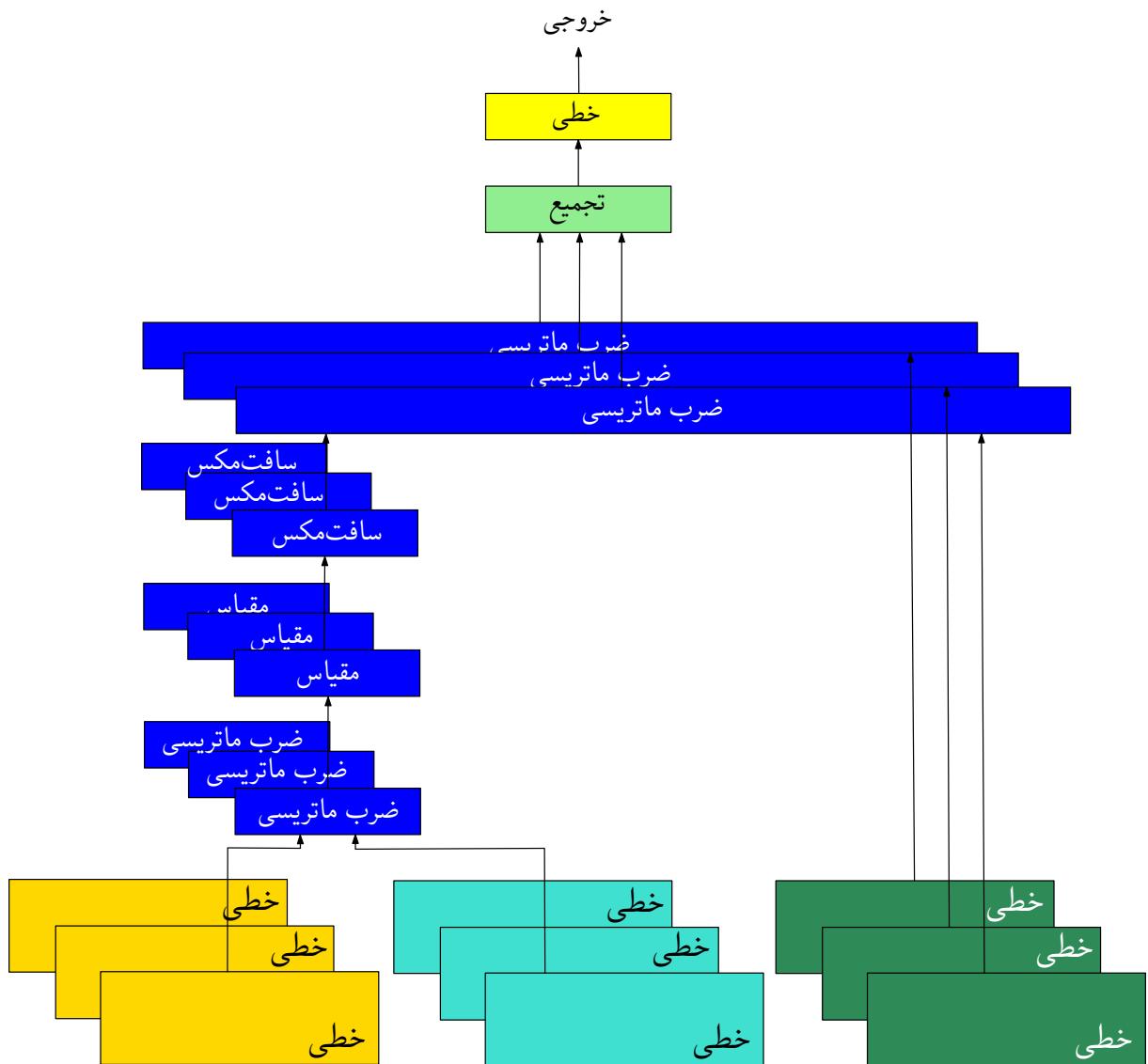


شکل ۱۹-۲: فرایند اعمال فیلتر توجه به تصویر

مراحل توضیح داده شده برای به دست آوردن صرفاً یک فیلتر توجه است. برای به دست آوردن بهترین نتیجه در خروجی، نیاز است که فیلترهای توجه مختلفی داشته باشیم. یا به عبارتی دیگر توجه‌های چند سر داشته باشیم و توجه‌ها را در همه جای جمله (یا تصویر) به دست آوریم. با قراردادن چند لایه از لایه توجه به خود در کنار هم «شکل ۲۰-۲» می‌توان لایه توجه چند سر را ایجاد نمود و فیلترهای توجه مختلفی را تولید کرد. «شکل ۲۱-۲»



شکل ۲۰-۲: فیلترهای توجه مختلف در تصویر



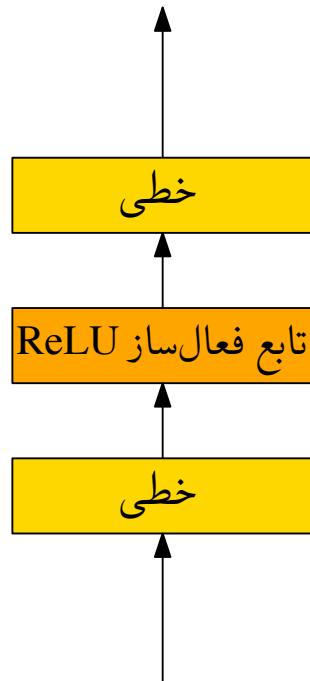
شکل ۲-۲۱: لایه توجه چند سر

۵-۳-۲ لایه رویه جلو

لایه رویه جلو متشکل است از دو لایه خطی که در قسمت های قبل توضیح داده شد و یک لایه تابع فعال ساز^{۲۵} رلو^{۲۶} «شکل ۲۲-۲».

²⁵ Activation Function

²⁶ ReLU



شکل ۲۲-۲: لایه روبه‌جلو

۶-۳-۲ لایه خطی رمزگشا

پس از تشکیل فیلترهای توجه در بلوک رمزگذار، ۲ نمونه مشابه از آن ساخته می‌شود. این دو نمونه ماتریس‌های Q و K برای بلوک رمزگشا خواهند بود و ماتریس سوم که ماتریس V است، همانند توضیحات قسمت‌های قبل به صورت مستقیم وارد بلوک رمزگذار شده و درنهایت این سه ماتریس به صورت همزمان وارد به لایه توجه چند سر موجود در بلوک رمزگذار می‌شوند.

۷-۳-۲ لایه خطی رمزگذار

درنهایت در بلوک رمزگذار یک لایه خطی وجود دارد که وظیفه طبقه‌بندی کلمات ورودی را دارد. تعداد نورون‌های این لایه برابر است با تعداد کلمات موجود در کلمه. وظیفه این لایه مشخص‌کردن بهترین خروجی و خروجی برنده که می‌تواند به عنوان کلمه بعدی برای تکمیل جمله پیشنهاد داده شود است.

برای مثال اگر جمله‌ای که از ابتدا به عنوان مثال بیان کردیم را به شبکه بدھیم می‌توانیم انتظار داشته باشیم که یکی از خروجی‌های شبکه به صورت زیر تولید شود:

“زمانی که شما درس می‌خوانید، **بیشترین تمرکز را دارید.**”

فصل ۳

کارهای پیشین

ماشین‌های دوار به طیف وسیعی از ماشین‌ها اطلاق می‌شود که حول یک محور می‌چرخند، مانند توربین‌ها، ژنراتورها، پمپ‌ها، کمپرسورها و موتورها [۱۸]. این ماشین‌ها در بسیاری از زمینه‌ها مانند تولید برق، نفت و گاز، حمل و نقل کاربردهای گسترده‌ای دارند [۱۹]. بلبرینگ^۱ به عنوان یکی از مهم‌ترین ابزارهای صنعتی، با تسهیل چرخش شفت از طریق یک رابط کم اصطکاک بین شفت و محفظه آن و کاهش سایش و پارگی اجزای ماشین، نقش مهمی در تضمین عملکرد روان و کارآمد ماشین‌های دوار دارد [۲۰]. با این حال، بلبرینگ‌ها در طول زمان به دلیل بار زیاد، روانکاری ناکافی و آلودگی، تخریب می‌شوند [۲۱]. تخریب بلبرینگ‌ها می‌تواند منجر به لوزش و نویز بیش از حد، بازده کم و مصرف انرژی بالا شود و اگر کنترل نشود، می‌تواند منجر به خرابی‌های فاجعه‌بار شود که باعث تعمیرات پرهزینه و ازکارافتادن ماشین‌آلات شود [۲۲] [بنابراین؛ نظارت و پیش‌بینی وضعیت سلامت و عمر مفید باقی‌مانده بلبرینگ‌ها برای کاهش زمان ازکارافتادن دستگاه بسیار مهم است.

به‌طورکلی روش‌هایی پیش‌بینی RUL را می‌توان به ۳ دسته تقسیم کرد:

- روش‌های مبتنی بر مدل

- روش‌های مبتنی بر داده

- روش‌های ترکیبی

¹Bearing

۱-۳ روش‌های مبتنی بر مدل

روش‌های مبتنی بر مدل، نیازمند مدل‌سازی دقیق دینامیک مسئله و اطلاع از جزئیات دقیق ابزار صنعتی مورد بررسی است. ساختار تجهیزات صنعتی در مقیاس‌های بزرگ با روابط غیرخطی بین سیستم‌ها و قطعات مختلف پیچیده می‌شود که اغلب ارائه یک مدل دقیق از آن کار بسیار دشواری است. هدف اصلی در روش‌های مبتنی بر داده، ایجاد یک رابطه نگاشت بین فضای ویژگی‌های ورودی و RUL هدف است. این روش نیازی به شناخت دقیق فیزیکی مدل مکانیکی ندارد.

برای مثال در [۵]، لی و همکاران، با استفاده از ارائه یک مدل احتمالاتی مبتنی بر روابط مارکوف^۲ که بخشی از آنها را در ادامه آورده‌ایم:

$$RUL = \sum_{\substack{i=1 \\ \text{وضعیت فعلی}}}^N [\mu(D(s_i)) + cf \cdot \sigma(D(s_i))] \quad (1-3)$$

$$RUL = \sum_{\substack{i=1 \\ \text{وضعیت فعلی}}}^N \mu(D(s_i)) \quad (2-3)$$

$$RUL = \sum_{\substack{i=1 \\ \text{وضعیت فعلی}}}^N [\mu(D(s_i)) - cf \cdot \sigma(D(s_i))] \quad (3-3)$$

سعی در مدل‌سازی دقیق یک سیستم مکانیکی، برای پیش‌بینی عمر باقی‌مانده مفید داشتند. نتایج ارائه شده در این مقاله «شکل ۱-۳» نشان می‌دهد که نویسندهاند نتوانسته‌اند مقدار خطای RMS را تا حد خوبی کاهش دهند.

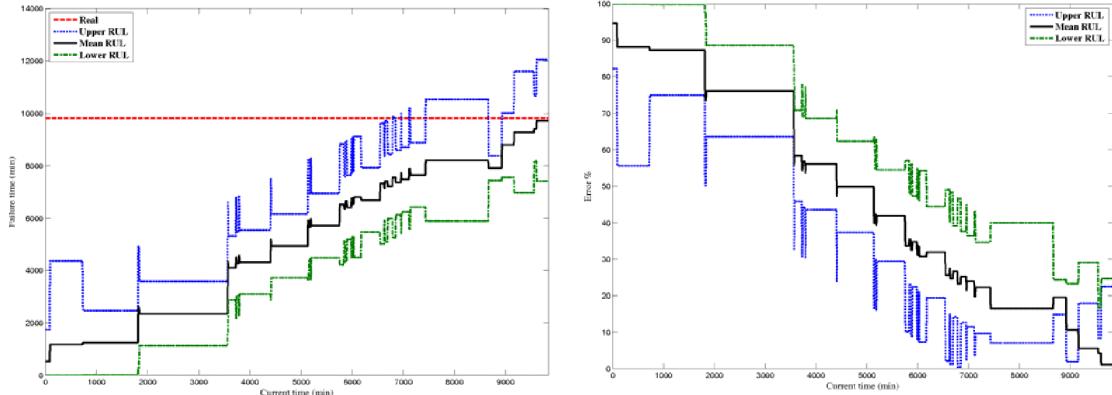
Method	HMM	HSMM	WPD/MoG-HMM
RMSE	1.2322×10^5	2.8229×10^4	4.6679×10^3

شکل ۱-۳: نتایج ارائه شده [۵]

دلیل این موضوع را می‌توان، پیچیدگی روابط برای مدل‌سازی و عدم امکان مدل‌سازی دقیق مسئله دانست که همان‌طور نیز که قبلاً بیان شد در همه مقالاتی که دسته روش‌های مبتنی بر مدل قرار می‌گیرند این ضعف دیده می‌شود.

²Markov

در این مقاله، برای مدل‌سازی روابط حاکم بر مسئله، از ماشین حالت^۳ استفاده شده است. نمودارهای ارائه شده در این مقاله ضعف را نشان می‌دهند که ماشین حالت انتخاب مناسبی برای این مسئله نمی‌باشد. چرا که سیستم ارائه شده نمی‌تواند با حداکثر سرعت خودکار کند و به صورت برخط پیش‌بینی را انجام دهد. این ضعف خودش را در دستگاه‌های بلادرنگ^۴ نشان می‌دهد.



شکل ۲-۳: نمودار پیش‌بینی شده [۵]

۲-۳ روش‌های مبتنی بر داده

به دلیل پیچیدگی روابط و شرایط موجود در مسئله و دشواری تحلیل کامل سیستم در روش‌های مبتنی بر مدل، روش‌های مبتنی بر تحلیل داده‌ها مورد اقبال قرار گرفته‌اند. هدف در روش‌های مبتنی بر داده، ایجاد یک رابطه نگاشت بین ورودی و خروجی، بدون داشتن دانش قبلی نسبت به سیستم است. این روش‌ها به طور گستردگی برای پیش‌بینی RUL بلبرینگ‌ها مورداستفاده قرار گرفته‌اند، زیرا می‌توانند پیش‌بینی را با استفاده از داده‌های نظارت بر وضعیت، بدون دانش و شناخت قبلی نسبت به سیستم انجام دهنند. روش‌های مبتنی بر داده را می‌توان به دو دسته گروه‌بندی کرد:

- روش‌های مبتنی بر یادگیری ماشین سنتی
- روش‌های مبتنی بر یادگیری عمیق

روش‌های یادگیری ماشین سنتی شامل رگرسیون بردار پشتیبان^۵ (SVR) [۲۳]، جنگل تصادفی^۶ (RF)

³State Machine

⁴Real-Time

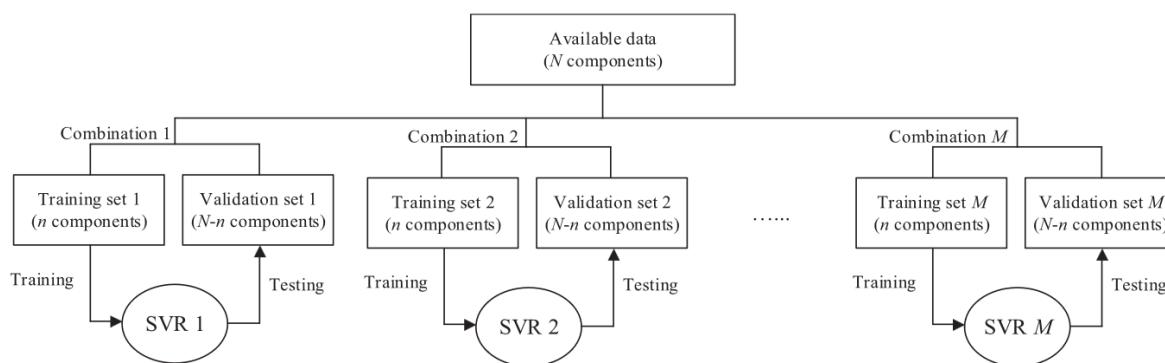
⁵Support Vector Regression

⁶Random Forest

[۲۴]، یادگیری افراطی ماشین^۷ (ELM) [۲۵]، یادگیری گروهی^۸ (GP) [۲۶]، فرایند گاوی^۹ (EL) [۲۷]، یادگیری افراطی ماشین^{۱۰} (MM) [۲۸] و موارد دیگر است.

۱-۲-۳ روش‌های مبتنی بر یادگیری ماشین سنتی

به عنوان مثال، وانگ و همکاران [۶] یک رویکرد رگرسیون برداری چند پشتیبانی را برای به دست آوردن پارامترهای مدل فرعی بهینه برای پیش‌بینی RUL بلبرینگ‌ها معرفی کرد. آن‌ها از مدلی مانند شکل «۳-۳» استفاده کردند.



شکل ۳-۳: مدل ارائه شده مبتنی بر بردارهای پشتیبان در [۶]

پس از به دست آوردن پارامترهای مدل بهینه، یک سازوکار به روزرسانی وزن خودکار برای ارزیابی مناسب بودن هر مدل فرعی برای عملکرد پیش‌بینی قوی‌تر پیشنهاد شد. در این کار از مجموعه داده CMAPASS برای ارزیابی عملکرد پیش‌بینی روش ارائه شده استفاده شد و نتایج ارائه شده «شکل ۴-۳» نشان می‌دهد که داد که RMSE روش ارائه شده، ۹۸/۱۴ به دست آمده است.

منگ و همکاران [۵] مدل مارکوف خاکستری را با نظریه طیف فراکتال^{۱۱} ادغام کرد تا مسیر تخریب بلبرینگ‌ها را پیش‌بینی کند. برای انجام پیش‌بینی در این مقاله از مجموعه داده PRONOSTIA استفاده شده است. روش ارائه شده با ویژگی ذرات مورفولوژی تعمیم‌یافته مقایسه شده است که نشان داده روش ارائه شده می‌تواند RMSE را تا ۴ درصد کاهش دهد. نتایج این روش در شکل «۵-۳» آورده شده است.

وانگ و همکاران در [۸] روشی مبتنی بر ترکیب ویژگی‌های استخراجی در حوزه زمان و فرکانس را بر روی شبکه LSTM ارائه کردند. ویژگی‌های زمانی و فرکانسی مورد استفاده در این مقاله در شکل «۷-۳» آورده شده است.

⁷Extreme Learning Machine

⁸Gaussian Process

⁹Ensemble Learning

¹⁰Markov Model

¹¹Fractal Spectrum

Prediction errors for component 17.

Model	MRE	RMSE	Score	Convergence
SVR 1	0.1538	0.1129	1.9882	192.8072
SVR 2	0.1666	0.1371	2.3389	<u>163.3994</u>
SVR 3	0.1313	<u>0.0777</u>	<u>1.4517</u>	215.2919
SVR 4	0.2333	0.1606	2.8810	196.4635
SVR 5	0.1704	0.1177	2.0641	184.7899
SVR 6	0.1603	0.1255	2.1551	181.1349
SVR 7	0.2481	0.1529	2.6155	313.4165
SVR 8	0.1817	0.1031	1.8914	232.6900
SVR 9	0.2473	0.1326	2.3010	339.7910
SVR 10	0.1806	0.1328	2.2448	277.0041
Multi-SVR fusion	0.1283	0.1160	1.8548	165.6945

[۶] شکل ۴-۳: نتایج ارائه شده

Prediction method	Component	RMSE(%)
GMMP	IMF1	13.38
	IMF2	14.55
Fractal spectrum parameters	IMF1	9.33
	IMF2	10.53

[۷] شکل ۴-۵: نتایج ارائه شده

در این مقاله از مجموعه داده PRONOSTIA استفاده شده است و نتایج ارائه شده نشان از کاهش مقدار RMSE دارد. «شکل ۴-۲»

۲-۲-۳ روش‌های مبتنی بر یادگیری عمیق

روش‌های یادگیری ماشین سنتی در یادگیری روابط پیچیده و غیرخطی مؤثر نیستند. برای پرداختن به این موضوع، در سال‌های اخیر، روش‌های یادگیری عمیق به دلیل توانایی آنها در یادگیری الگوهای پیچیده و مدیریت حجم زیادی از داده‌ها، به طور فزاینده‌ای برای پیش‌بینی RUL محبوب شده‌اند. این روش‌های یادگیری عمیق را می‌توان به‌طورکلی به دو زیر گروه تقسیم کرد:

- شبکه‌های تکراری^{۱۲}

- شبکه‌های غیرتکراری

¹²Recurrent

Time-domain Features		Frequency-domain Features	
$X_{max} = \max(\mathbf{x}_i)$	$X_{min} = \min(\mathbf{x}_i)$	$X_{maxf} = \max(\boldsymbol{\beta}_i)$	$X_{minf} = \min(\boldsymbol{\beta}_i)$
$X_{sra} = \left(\frac{1}{M} \sum_{j=1}^M \sqrt{ x_{ij} } \right)^2$	$X_{mean} = \frac{1}{M} \sum_{j=1}^M x_{ij}$	$X_{rmsf} = \left(\frac{1}{L} \sum_{l=1}^L \beta_{il}^2 \right)^{\frac{1}{2}}$	$X_{meanf} = \frac{1}{L} \sum_{l=1}^L \beta_{il}$
$X_{rv} = \left(\frac{1}{M} \sum_{j=1}^M (x_{ij} - X_{mean})^2 \right)^{\frac{1}{2}}$	$X_{mav} = \frac{1}{M} \sum_{j=1}^M x_{ij} $	$X_{rvf} = \left(\frac{1}{L} \sum_{l=1}^L (\beta_{il} - X_{meanf})^2 \right)^{\frac{1}{2}}$	$X_{fc} = \frac{1}{L} \sum_{l=1}^L \frac{(f_{il} \cdot \beta_{il})}{X_{meanf}}$
$X_{sv} = \left(\frac{1}{M} \sum_{j=1}^M \frac{(x_{ij} - X_{mean})}{X_{rv}} \right)^3$	$X_{rms} = \left(\frac{1}{M} \sum_{j=1}^M x_{ij}^2 \right)^{\frac{1}{2}}$	$X_{svf} = \left(\frac{1}{L} \sum_{l=1}^L \left(\frac{\beta_{il} - X_{meanf}}{X_{rvf}} \right) \right)^3$	$\left(\frac{1}{L} \sum_{l=1}^L (f_{il} - X_{fc})^2 \cdot \beta_{il} \right)^{\frac{1}{2}}$
$X_{kv} = \left(\frac{1}{M} \sum_{j=1}^M \frac{(x_{ij} - X_{mean})}{X_{rv}} \right)^4$	$X_{mf} = \frac{\max(\mathbf{x}_i)}{X_{sra}}$	$X_{kuf} = \left(\frac{1}{L} \sum_{l=1}^L \frac{(\beta_{il} - X_{meanf})}{X_{rvf}} \right)^4$	$\left(\frac{1}{L} \sum_{l=1}^L \frac{(f_{il}^2 \cdot \beta_{il})}{X_{meanf}} \right)^{\frac{1}{2}}$
$X_{kf} = X_{kv} / X_{rms}^4$	$X_{skf} = X_{kv} / X_{rv}^3$	$X_{kff} = X_{kuf} / X_{rmsf}^4$	$X_{skff} = X_{kuf} / X_{rvf}^3$
$X_{if} = \max(\mathbf{x}_i) / X_{mav}$	$X_{shf} = X_{rv} / X_{mav}$		
$X_{cf} = \max(\mathbf{x}_i) / X_{rms}$	$X_{ppv} = X_{max} - X_{min}$		

شکل ۳-۶: ویژگی‌های استفاده شده در [۸]

Bearing	LR	DT	SVR	RF	Bag	Ada	RS	FFEM
1	40.13	17.79	11.09	14.09	7.05	9.15	6.69	6.32
2	26.73	26.08	16.53	17.79	16.53	16.26	17.17	15.98
3	20.66	16.42	16.17	9.91	8.23	9.22	8.32	8.21
4	34.02	26.33	17.87	11.48	9.90	18.85	9.08	7.62
5	23.44	19.95	18.08	16.67	14.39	15.32	14.32	13.85
6	22.48	24.21	22.92	19.00	17.14	17.65	17.13	17.10
7	32.90	25.81	24.88	17.57	17.21	22.17	17.24	16.98

شکل ۷-۳: نتایج ارائه شده در [۸]

تکنیک‌های یادگیری عمیق غیرتکراری شامل شبکه‌های عصبی مصنوعی^{۱۳} [۲۹] شبکه عصبی کانولوشنی^{۱۴} [۳۰]، رمزگذار های خودکار^{۱۵} [۱۰]، شبکه عصبی گرافی^{۱۶} [۳۱، ۳۲] و مدل‌های تولیدی مانند شبکه مولد متخاصم^{۱۷} [۳۳].

به عنوان مثال، زو و همکاران [۹] یک CNN چند مقیاسی «شکل ۸-۳» را برای استخراج ویژگی‌های عمیق از داده‌های پایش وضعیت پیشنهاد کرد، و ویژگی‌های استخراج شده با ویژگی‌های زمان - فرکانس برای پیش‌بینی RUL بلبرینگ‌ها ترکیب شدند. در این مقاله، از مجموعه داده PRONOSTIA استفاده شده است. نتایج نشان داد که CNN چند مقیاسی ارائه شده عملکرد پیش‌بینی را افزایش می‌دهد. نتایج این کار در شکل

¹³Artificial Neural Network

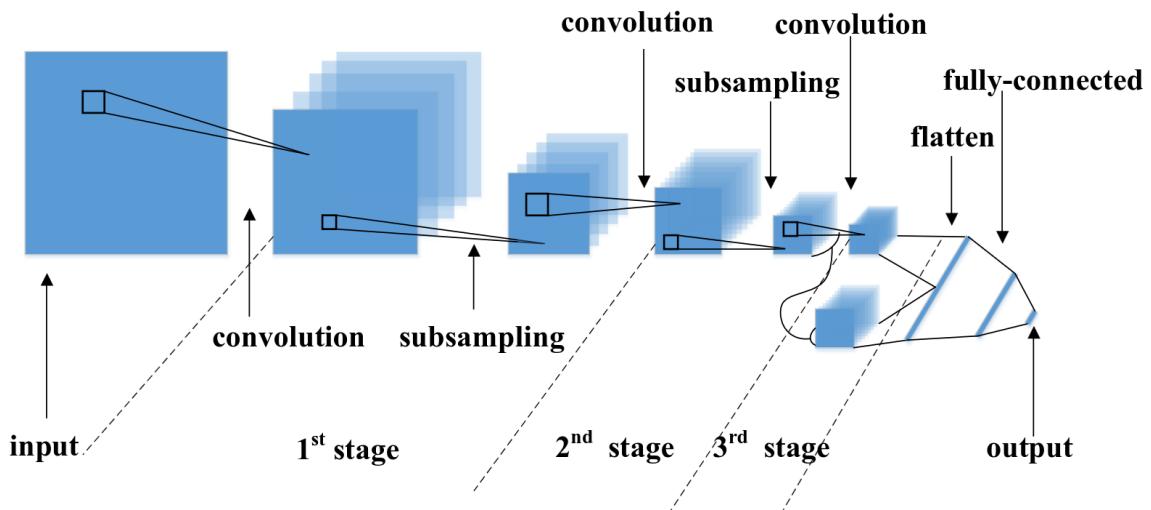
¹⁴Convolutional Neural Network

¹⁵Autoencoders

¹⁶Graph Neural Network

¹⁷Generative Adversarial Network

۹-۳) آورده شده است.



شکل ۸-۳: ساختار ارائه شده در [۹]

Testing dataset	Current time(s)	Actual RUL(s)	MSCNN RUL(s)	RNN-HI RUL(s)	SOM-HI RUL(s)	SVR-HI RUL(s)
Bearing1_3	18010	5730	4731	3250	5790	5970
Bearing1_4	11380	2900	2590	1100	410	1200
Bearing1_5	23010	1610	3996	1980	6080	5040
Bearing1_6	23010	1460	1744	1150	1180	1230
Bearing1_7	15010	7570	6090	6220	8110	9120
Score			0.3624	0.2798	0.3605	0.2657
MAE			1091.8	1262	1568	1430
NRMSE			0.3514	0.5522	0.5342	0.4107

شکل ۹-۳: نتایج ارائه شده در [۹]

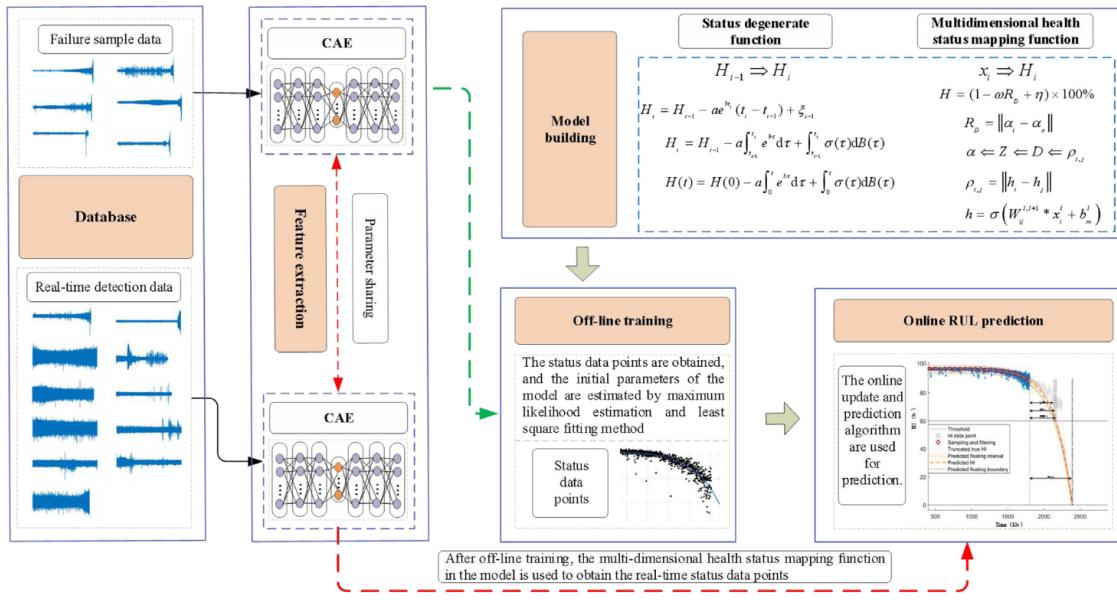
خو و همکاران در [۱۰] یک رمزگذار خودکار کانولوشنی برای استخراج ویژگی‌ها از داده‌های جمع‌آوری شده از بلبرینگ‌های مستهلك شده ارائه کرد، «شکل ۱۰-۳» و یک تابع مقیاس‌بندی شاخص سلامت برای کاهش مقیاس ویژگی‌های استخراج شده به کار گرفته‌اند. در این مقاله نیز از مجموعه داده PRONOSTIA برای آموزش و ارزیابی مدل استفاده شده است. نتایج آن‌ها نشان می‌دهد «شکل ۱۱-۳» که روش ارائه شده در پیش‌بینی RUL و ارزیابی مراحل تخریب بلبرینگ‌ها کارآمد است.

یانگ و همکاران نیز در [۳۴] از یک GNN برای تخمین RUL بلبرینگ‌ها استفاده کردند.

در حالی که بسیاری از مطالعات انجام شده، استفاده از شبکه‌های عصبی غیرتکراری را برای پیش‌بینی RUL مورد بررسی قرار داده‌اند، این روش‌ها اغلب در برخورد با داده‌های سری زمانی مؤثر نیستند؛ زیرا قادر به تشخیص ویژگی‌های متوالی و وابستگی‌های بلندمدت در داده‌های پایش وضعیت نیستند. در مقابل، الگوریتم‌های یادگیری عمیق مکرر مانند شبکه عصبی بازگشتی [۳۵]، حافظه کوتاه‌مدت [۱۸، ۳۶، ۳۷]، شبکه بازگشتی دروازه‌ای^{۱۹} [۳۸] و نسخه‌های دوطرفه آنها مانند LSTM دوطرفه [۳۹] و GRU دوطرفه نشان داده شده است که در پیش‌بینی

¹⁸Long Short-Term Memory

¹⁹Gated recurrent network



شکل ۱۰-۳: مدل ارائه شده در [۱۰]

Summary of experimental results with nine prediction methods.

Testing bearing	Wang [19] Er (%)	Liu [38] Er (%)	Lei [39] Er (%)	Zhang [29] Er (%)	Yoo [40] Er (%)	She [41] Er (%)	Kumar [42] Er (%)	Zhuang [43] Er (%)	Proposed method Er (%)
Bearing 1-3	5.06	2.58	-0.35	2.27	1.05	-4.36	-1.40	3.66	-2.62
Bearing 1-4	23.30	-9.14	5.6	5.60	20.35	70.50	29.20	2.65	17.40
Bearing 1-5	4.35	-0.99	100	12.42	11.18	6.21	-2.48	-4.35	5.59
Bearing 1-6	0.68	6.03	28.08	10.96	34.93	-4.11	8.90	-0.68	3.42
Bearing 1-7	-42.54	-0.70	-19.55	-22.46	29.19	18.63	25.36	10.57	1.06
Bearing 2-3	17.40	55.44	-20.19	0.99	57.24	17.40	-10.49	0.93	26.96
Bearing 2-4	12.23	15.56	8.63	5.76	-1.44	-1.44	-6.47	-5.76	-2.88
Bearing 2-5	-0.32	49.19	23.3	25.89	-0.65	5.18	20.71	-2.27	7.77
Bearing 2-6	-2.33	38.53	58.91	-10.85	-42.64	16.25	4.65	-5.43	13.95
Bearing 2-7	8.62	5.17	5.17	1.72	8.62	10.34	1.72	56.90	-8.62
Bearing 3-3	-3.66	2.56	40.24	-3.66	-1.22	6.10	3.44	2.44	3.66
SD	17.0173	22.2020	35.4135	12.5670	25.5912	20.8134	13.0920	17.7509	10.1942
Score	0.6682	0.6101	0.4285	0.6442	0.5689	0.6190	0.6233	0.6928	0.6860

شکل ۱۱-۳: نتایج ارائه شده در [۱۰]

RUL بلبرینگ‌ها مؤثرتر هستند.

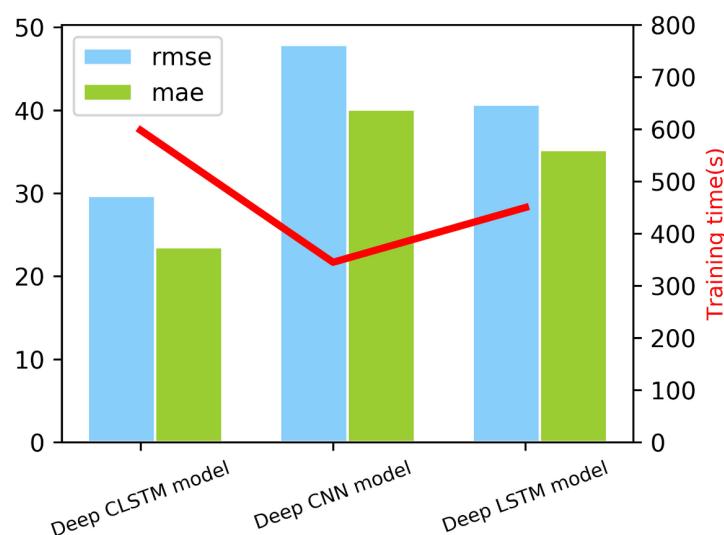
به عنوان مثال، ما و همکاران در [۱۱] یک LSTM کانولوشنی عمیق «شکل ۱۲-۳» برای پیش‌بینی RUL معرفی کرد. عملیات کانولوشن در سلول LSTM پیشنهادی قادر به استخراج ویژگی‌های زمان - فرکانس و حفظ وابستگی‌های طولانی مدت به طور همزمان بود. نتایج عددی ارائه شده در این مقاله «شکل ۱۳-۳»، نشان می‌دهد که LSTM کانولوشنی پیشنهادی بهتر از سلول CNN و LSTM عمیق است.

ژانگ و همکاران [۱۲] یک روش یادگیری عمیق ترکیبی موازی «شکل » را توسعه داد که یک CNN یک بعدی و GRU دوطرفه را برای پیش‌بینی‌های RUL ترکیب می‌کند و امکان استخراج موازی ویژگی‌های مکانی و زمانی از داده‌های نظارت بر شرایط را فراهم می‌کند. نتایج ارائه شده «شکل ۱۵-۳» در مقاله نشان می‌دهد که این روش بهبود خوبی نسبت به سایر مقاله‌های بررسی شده در این کار داشته است.

Summary of experimental results with nine prediction methods.

Testing bearing	Wang [19] Er (%)	Liu [38] Er (%)	Lei [39] Er (%)	Zhang [29] Er (%)	Yoo [40] Er (%)	She [41] Er (%)	Kumar [42] Er (%)	Zhuang [43] Er (%)	Proposed method Er (%)
Bearing 1-3	5.06	2.58	-0.35	2.27	1.05	-4.36	-1.40	3.66	-2.62
Bearing 1-4	23.30	-9.14	5.6	5.60	20.35	70.50	29.20	2.65	17.40
Bearing 1-5	4.35	-0.99	100	12.42	11.18	6.21	-2.48	-4.35	5.59
Bearing 1-6	0.68	6.03	28.08	10.96	34.93	-4.11	8.90	-0.68	3.42
Bearing 1-7	-42.54	-0.70	-19.55	-22.46	29.19	18.63	25.36	10.57	1.06
Bearing 2-3	17.40	55.44	-20.19	0.99	57.24	17.40	-10.49	0.93	26.96
Bearing 2-4	12.23	15.56	8.63	5.76	-1.44	-1.44	-6.47	-5.76	-2.88
Bearing 2-5	-0.32	49.19	23.3	25.89	-0.65	5.18	20.71	-2.27	7.77
Bearing 2-6	-2.33	38.53	58.91	-10.85	-42.64	16.25	4.65	-5.43	13.95
Bearing 2-7	8.62	5.17	5.17	1.72	8.62	10.34	1.72	56.90	-8.62
Bearing 3-3	-3.66	2.56	40.24	-3.66	-1.22	6.10	3.44	2.44	3.66
SD	17.0173	22.2020	35.4135	12.5670	25.5912	20.8134	13.0920	17.7509	10.1942
Score	0.6682	0.6101	0.4285	0.6442	0.5689	0.6190	0.6233	0.6928	0.6860

شکل ۱۲-۳: ساختار ارائه شده در [۱۱]



شکل ۱۳-۳: نتایج ارائه شده در [۱۱]

هان و همکاران در [۴۰]، خود کدگذار پشتهای^{۲۰} و RNN را برای پیش‌بینی RUL ترکیب کردند. خود کدگذار برای ترکیب ویژگی‌ها در شاخص‌های سلامتی استفاده شد و سپس RNN برای پیش‌بینی شده مدل پیاده‌سازی شده است.

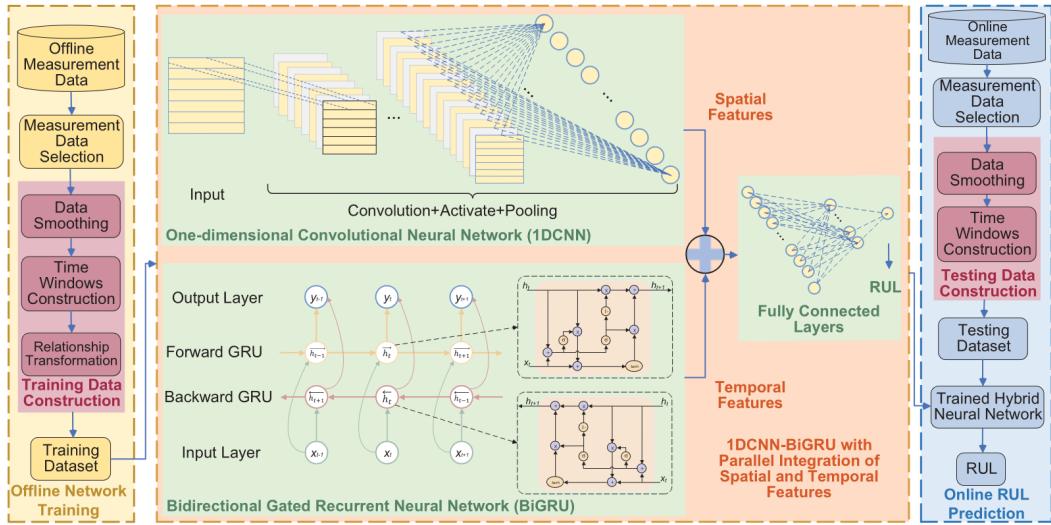
اگرچه شبکه‌های RNN به طور گسترده برای پیش‌بینی‌های RUL مورد استفاده قرار گرفته‌اند، اما در ثبت و درک وابستگی‌های بلندمدت در داده‌ها مؤثر نیستند و در فرآیند آموزش دچار محوش‌گی گرادیان^{۲۱} و فراموشی می‌شوند.

برای پرداختن به این موضوع، مکانیزم‌های توجه، بهویژه مکانیزم توجه به خود^{۲۲} [۴۱]، به طور گسترده استفاده شده است. مکانیزم توجه در شبکه‌های مختلفی استفاده شده است، اما در میان شبکه‌هایی که از سازوکار

²⁰Stacked Autoencoder

²¹Gradient Vanishing

²²Self-Attention



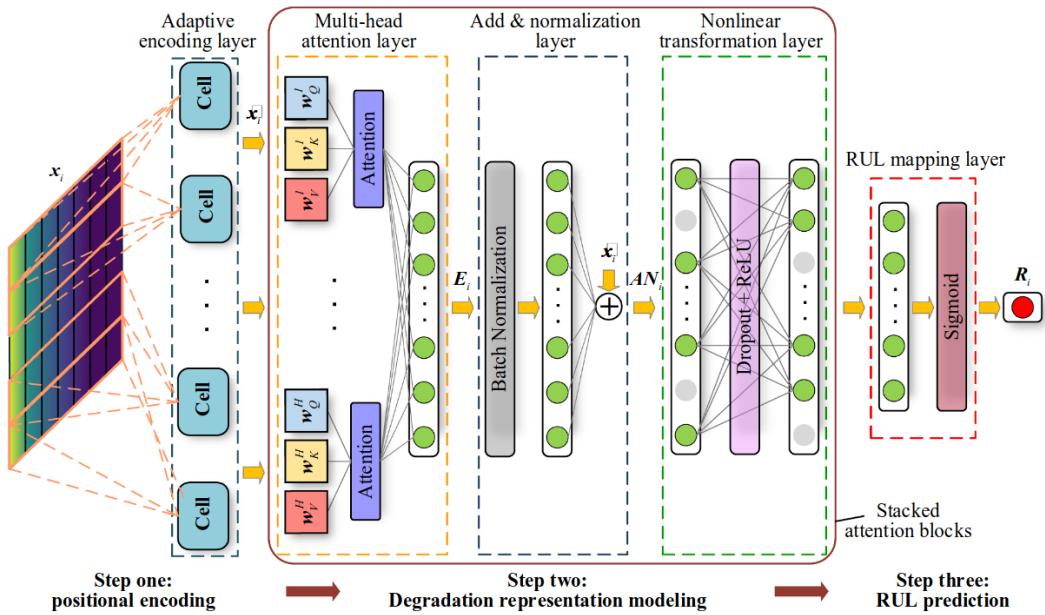
شکل ۱۴-۳: ساختار ارائه شده در [۱۲]

Approach	S-Score					RMSE		
	FD001	FD002	FD003	FD004	FD001	FD002	FD003	FD004
SVM [19]	7703.33	316483.31	22541.58	141122.19	40.72	52.99	46.32	59.96
ETR [19]	1667.86	803203.04	2240.70	345956.76	23.76	36.05	25.66	40.01
CNN-FNN [20]	1286.70	13570.00	1596.20	7886.40	18.45	30.29	19.82	29.16
Tandem CNN-RNN [22]	820.67	15917.00	950.94	7212.20	16.89	30.97	17.82	29.73
HELM [19]	562.96	434056.50	456.32	105492.77	17.43	34.84	17.69	37.98
MLP [19]	560.59	14026.72	479.85	10444.35	16.78	28.78	18.47	30.96
ELM [22]	523.00	498149.97	573.78	121414.47	17.27	37.28	18.90	38.43
LSTMBS [43]	481.10	7982.00	493.40	5200.00	14.89	26.86	15.11	27.11
DBN [19]	417.59	9031.64	442.43	7954.51	15.21	27.12	14.71	29.88
Deep LSTM [21]	338.00	4450.00	852.00	5550.00	16.14	24.49	16.18	28.17
MODBNE [19]	334.23	5585.34	421.91	6557.62	15.04	25.05	12.51	28.66
IESGP [44]	331.90	4245.40	355.20	6280.80	14.72	24.81	14.99	28.61
BiGRU [23]	-	6352.00	-	6339.00	-	26.54	-	29.13
Attention-LSTM [25]	322.44	-	-	5649.14	14.53	-	-	27.06
Tandem CNN-LSTM [24]	303.00	3440.00	1420.94	4630.00	16.12	20.46	17.12	23.26
BiLSTM [45]	295.00	4130.00	317.00	5430.00	13.65	23.18	13.74	24.86
Tandem CNN-LSTM-DA [46]	290.00	9869.00	316.00	6594.00	14.40	27.23	14.32	26.69
AdaBN-DCNN [47]	279.00	2020.00	817.00	3690.00	13.17	20.87	14.97	24.57
BiLSTM-ED [48]	273.00	3099.00	574.00	3202.00	14.74	22.07	17.48	23.49
CEED-DLSTM [49]	262.00	6953.00	452.00	15069.00	14.72	29.00	17.72	33.43
Ours	223.30	1201.35	268.83	2304.08	12.99	16.10	12.48	18.29

شکل ۱۵-۳: نتایج ارائه شده در [۱۲]

توجه به خود استفاده می‌کنند، ترانسفرمر یکی از قدرتمندترین و مؤثرترین شبکه‌ها است. معماری ترانسفرمر توجه را با ویژگی‌های دیگری مانند رمزگذاری موقعیتی و شبکه‌های عصبی پیش‌خور ترکیب می‌کند تا همبستگی غیرخطی در داده‌ها را آشکار کند، در نتیجه عملکرد پیش‌بینی را به طور قابل توجهی بهبود می‌بخشد.

برای مثال، سو و همکاران در [۱۳] از رمزگذار ترانسفرمر برای پیش‌بینی RUL بلبرینگ‌ها استفاده کردند. «شکل ۱۶-۳» روش پیشنهادی آنها شامل دو مرحله بود که در مرحله اول ویژگی‌های سطح پایین را با استفاده از مکانیزم استخراج ویژگی استخراج می‌کرد و مرحله دوم از رمزگذار ترانسفرمر برای تخمین RUL استفاده می‌کرد. در این مقاله از دو مجموعه داده عمومی FEMTO و XJTU-SY برای تأیید اثربخشی روش ارائه شده استفاده شده و نتایج نشان داده است که رمزگذار ترانسفرمر منجر به افزایش عملکرد پیش‌بینی می‌شود. «شکل ۱۷-۳» دینگ و همکاران در [۱۴] یک ترانسفرمر - کانولوشن را معرفی کرد که عملکرد کانولوشن و مکانیسم

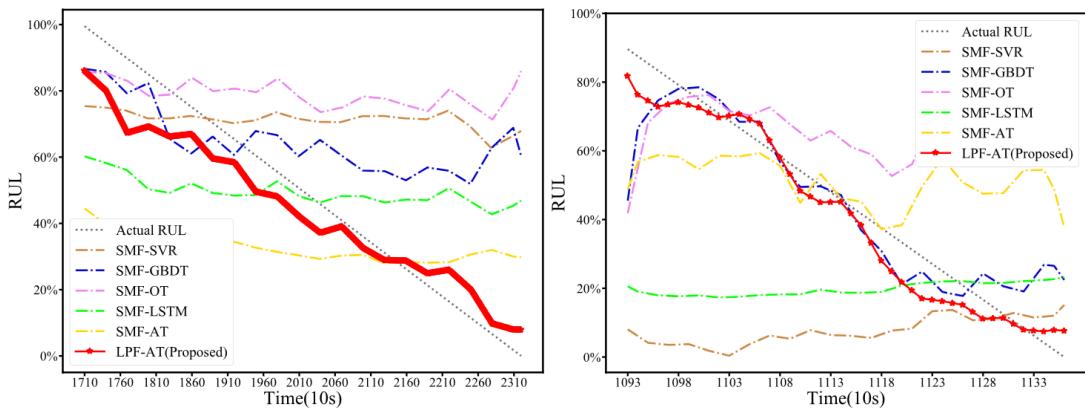


شکل ۱۶-۳: معماری ارائه شده در [۱۳]

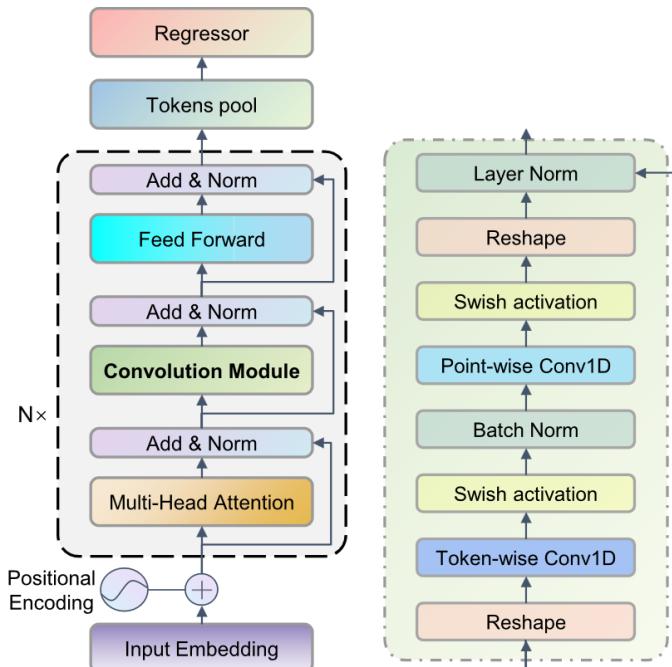
خودتوجهی را برای تخمین RUL بلبرینگ‌ها یکپارچه می‌کند. عملیات کانولوشن برای آشکارکردن وابستگی‌های محلی در داده‌ها استفاده شده است، و مکانیسم توجه به خود برای آشکارکردن وابستگی‌های سراسری داده‌ها به کار گرفته شده است. معماری این مدل در شکل «۱۸-۳» آورده شده است. نتایج ارائه شده در این مقاله «» نیز نشان از بهبود عملکرد و کاهش خطای پیش‌بینی نسبت به سایر روش‌ها دارد.

ژانگ و همکاران در ^{۲۳} یک مدل ترنسفرمر جدید «شکل ۲۰-۳» برای پیش‌بینی RUL پیشنهاد کرد که در آن یک مکانیزم چندسر پراکنده توجه به خود برای بهبود کارایی محاسباتی معرفی شد. نتایج تجربی نشان داده است که روش پیشنهادی از ترنسفرمر معمولی و سایر روش‌های مبتنی بر داده بهتر عمل می‌کند.

²³zhang2023integrated



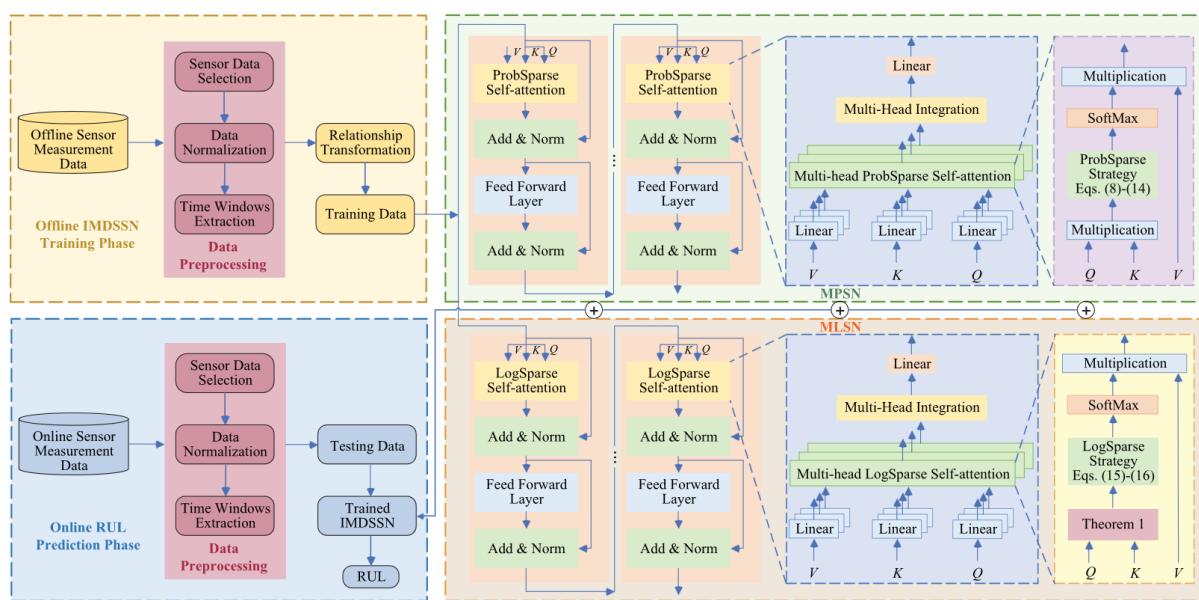
شکل ۱۷-۳: خروجی پیش‌بینی ارائه شده در [۱۳]



شکل ۱۸-۳: معماری ارائه شده در [۱۴]

	RMSE	MAE	Score
Ren's [37]	0.3385 ± 0.0147	0.2761 ± 0.0109	0.2218 ± 0.0124
Wu's [11]	0.3007 ± 0.0081	0.2541 ± 0.0058	0.2431 ± 0.0153
Chen's [12]	0.2657 ± 0.0079	0.2051 ± 0.0087	0.1865 ± 0.0114
Zhu's [14]	0.1987 ± 0.0043	0.1542 ± 0.0054	0.2784 ± 0.0271
Cheng's [10]	0.1373 ± 0.0051	0.1010 ± 0.0046	0.2759 ± 0.0185
Wang's [38]	0.1475 ± 0.0017	0.1283 ± 0.0021	0.3217 ± 0.0117
Chen's [39]	0.1291 ± 0.0034	0.0938 ± 0.0025	0.4397 ± 0.0127
CoT (base)	0.1250 ± 0.0019	0.0927 ± 0.0018	0.4515 ± 0.0139
CoT (large)	0.1096 ± 0.0008	0.0769 ± 0.0007	0.4822 ± 0.0103

شکل ۱۹-۳: نتیجه ارائه شده در [۱۴]



شکل ۲۰-۳: ساختار ارائه شده در [۱۵]

فصل ۴

چالش‌ها و نوآوری‌ها

۱-۴ تعریف دقیق مسئله

یک سیستم از ابتدایی که شروع به کار می‌کند، تا زمانی که اولین علائم خرابی را از خود بروز می‌دهد، به صورت خطی کار می‌کند. پس از نمایانشدن اولین علامت خرابی می‌توان عمر سیستم را با یک تابع خطی با شیب ۱- مدل کرد [۴۲]. مدت زمانی که اولین زمان خرابی نمایان می‌شود تا زمانی که سیستم به طور کامل از کار بیفتد را عمر باقی‌مانده مفید می‌نامند. این زمان در سیستم‌های مختلف بسته به شرایط و کاربردهای آن متفاوت است؛ اما در اپلیکیشن مورد بررسی ما این زمان را می‌توان در بازه ۴ دقیقه تا ۱۲۴ دقیقه در نظر گرفت [۴۳]. با توجه به اینکه حداقل زمان ۴ دقیقه است، الزام وجود و استفاده از FPGA خودش را نشان می‌دهد، زیرا می‌بایست تقریب و پیش‌بینی با بالاترین سرعت ممکن انجام شود و ازانجاكه در خودروهای خودران نمی‌توان از پردازنده‌های گرافیکی استفاده کرد، می‌بایست از FPGA استفاده نمود تا کمترین میزان توان مصرفی را با بالاترین سرعت پردازش داشته باشیم.

۲-۴ راه حل پیشنهادی و ایده‌ی اولیه

همان‌طور که در بخش کارهای پیشین «۳» مطرح شد، پژوهش‌های گسترده‌ای در این زمینه انجام شده است. اما تمامی کارای انجام شده صرفاً با دیدگاه نرم‌افزاری و آفلاین انجام شده است و پژوهش‌های کمی با این دیدگاه انجام شده است که پیش‌بینی به صورت آنلاین انجام شود. ما در این پژوهش قصد داریم با درنظر گرفتن کاربرد پیش‌بینی عمر باقی‌مانده بلبرینگ‌های خودروای خودران آن را بر روی FPGA شبیه‌سازی و پیاده‌سازی آن را امکان‌سنجی کنیم. فاز آموزش شبکه ترنسفرمر به صورت نرم‌افزاری و بیرون از FPGA انجام خواهد شد و

شرایط محیطی تأثیرگذار در عمر بلبرینگ در نظر گرفته نمی‌شود.

پیاده‌سازی‌های مختلفی از شبکه‌های Transformer بر روی FPGA انجام شده است و همگی آن‌ها به افزایش سرعت و کاهش انرژی مصرفی نسبت به GPU و CPU اشاره داشتند. برای مثال در [۴۴] فاز استنتاچ^۱ شبکه‌های BERT و GPT بر روی FPGA پیاده‌سازی شده است. در این مقاله گزارش شده است برای شبکه GPT، سرعت $1/9$ و انرژی $5/7$ نسبت به پیاده‌سازی مشابه بر روی کارت گرافیک NVIDIA A100 بهبودیافته است.

در معماری شبکه Transformer نیاز است که تمام داده‌ها به صورت کامل وارد شبکه شده و بتواند پیش‌بینی RUL را انجام دهد. به عبارتی دیگر یعنی داده‌ای که در زمان t_{n-10} وارد شبکه شده است بر داده‌ای که اکنون در زمان t_n وارد شده است تأثیر می‌گذارد. این خود را زمانی نشان می‌دهد که بخواهیم این سیستم را بر روی FPGA پیاده‌سازی کنیم. به دلیل آنکه در FPGA حافظه بزرگی در اختیار نداریم نمی‌توانیم تمامی سیگنال را ذخیره کنیم و سپس پردازش را انجام دهیم و از آنجایی که همه قسمت‌های سیگنال برای پیش‌بینی RUL بادقت بالا مهم است، نمی‌توانیم بخش‌هایی از سیگنال را دور بریزیم تا با این روش محدودیت حافظه را دور بزنیم؛ زیرا ممکن است اطلاعات مهمی از تخریب سیگنال در آن نواحی باشد؛ بنابراین نیاز داریم مکانیزمی را پیشنهاد دهیم تا بتوانیم سیگنال را ذخیره و پردازش کنیم به‌ نحوی که با محدودیت حافظه مواجه نشویم.

برای حل این چالش پیشنهاد می‌شود یک FIFO در نظر گرفته شود که حجم آن بر اساس زمان پردازش سیستم تعیین می‌شود به‌طوری که در مدت زمانی که طول می‌کشد تا FIFO پر شود، پردازش تکمیل شده باشد تا مجدداً FIFO با داده جدید پر شود. با اعمال این تکنیک امید داریم تا بتوانیم تمام داده‌های ورودی را پردازش کنیم و داده‌ای از دست نرود. یکی دیگر از ضرورت‌های استفاده از FPGA اینجا خودش را نشان می‌دهد، چرا که در پردازنده‌ها حجم FIFO محدود است و نمی‌توانیم آن را متناسب با کاربرد خودمان تنظیم کنیم. حجم FIFO بر اساس کاربرد تعیین می‌شود و این امکان را به ما می‌دهد که مناسب با نیازمان، FIFO را تنظیم کنیم.

¹Inference

فصل ۵

نتیجه‌گیری

پیش‌بینی عمر باقی‌مانده مفید یکی از مسائل کلیدی در نگهداری پیش‌گیرانه و مدیریت سلامت سیستم‌های صنعتی است. با افزایش پیچیدگی سیستم‌ها و اهمیت بهینه‌سازی عملکرد، استفاده از روش‌های پیشرفته برای پیش‌بینی دقیق‌تر و سریع‌تر RUL ضروری شده است. شبکه‌های عصبی مصنوعی، بهویژه شبکه‌های ترانسفرمر، به دلیل توانایی بسیار بالای آنها در درک وابستگی‌ها در داده‌های سری زمانی به عنوان یکی از ابزارهای مؤثر در این زمینه مطرح شده‌اند. شبکه‌های ترانسفرمر با بهره‌گیری از مکانیسم توجه، قادر به یادگیری وابستگی‌های بلندمدت در داده‌ها هستند و این ویژگی آن‌ها را برای کاربردهایی که نیاز به تحلیل دقیق و سریع داده‌های حسگر دارند، بسیار مناسب می‌سازد.

استفاده از FPGA برای پیاده‌سازی شبکه‌های عصبی می‌تواند بسیاری از محدودیت‌های موجود در زمینه پردازش داده‌ها و مصرف انرژی را برطرف کند. FPGA‌ها به دلیل قابلیت پیکربندی مجدد و پردازش موازی، می‌توانند سرعت پردازش را به طور قابل توجهی افزایش و مصرف انرژی را کاهش دهند. این ویژگی‌ها بهویژه در کاربردهای حیاتی مانند خودروهای خودران که نیاز به پیش‌بینی‌های بلادرنگ و با دقت بالا دارند، بسیار حائز اهمیت است. به عنوان مثال، پیاده‌سازی فاز استنتاج شبکه‌های BERT و GPT بر روی FPGA نشان داده است که سرعت و کارایی این شبکه‌ها به‌طور چشمگیری بهبود یافته و مصرف انرژی کاهش یافته است.

یکی از چالش‌های اصلی در پیاده‌سازی شبکه‌های ترانسفرمر بر روی FPGA، مدیریت حافظه محدود این سخت‌افزارهاست. برای حل این مشکل، می‌توان از تکنیک‌هایی مانند FIFO استفاده کرد که به ما امکان می‌دهد تا داده‌ها را به صورت مرحله‌ای پردازش و ذخیره کنیم. با تنظیم حجم FIFO براساس زمان پردازش سیستم، می‌توان اطمینان حاصل کرد که تمام داده‌های ورودی بدون از دست رفتن اطلاعات مهم پردازش شوند. این روش نه تنها کارایی پردازش را بهبود می‌بخشد بلکه نیاز به حافظه بزرگ را کاهش داده و از محدودیت‌های موجود در FPGA بهره‌برداری می‌کند. به این ترتیب، ترکیب شبکه‌های ترانسفرمر با FPGA می‌تواند راه حلی

کارآمد و بهینه برای پیش‌بینی RUL در سیستم‌های صنعتی و خودروهای خودران ارائه دهد.

Bibliography

- [1] Y. Wei and D. Wu. Conditional variational transformer for bearing remaining useful life prediction. *Advanced Engineering Informatics*, 59:102247, 2024.
- [2] B. Wang, Y. Lei, N. Li, and N. Li. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability*, 69(1):401–412, 2018.
- [3] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier. Pronostia: An experimental platform for bearings accelerated degradation tests. In *IEEE International Conference on Prognostics and Health Management, PHM’12.*, pages 1–8. IEEE Catalog Number: CPF12PHM-CDR, 2012.
- [4] Q. Wang, S. Zheng, A. Farahat, S. Serita, and C. Gupta. Remaining useful life estimation using functional data analysis. In *2019 ieee international conference on prognostics and health management (icphm)*, pages 1–8. IEEE, 2019.
- [5] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot. Hidden markov models for failure diagnostic and prognostic. In *2011 Prognostics and System Health Managment Confernece*, pages 1–8. IEEE, 2011.
- [6] Y. Li, X. Huang, C. Zhao, and P. Ding. A novel remaining useful life prediction method based on multi-support vector regression fusion and adaptive weight updating. *ISA transactions*, 131:444–459, 2022.
- [7] Z. Meng, J. Li, N. Yin, and Z. Pan. Remaining useful life prediction of rolling bearing using fractal theory. *Measurement*, 156:107572, 2020.
- [8] G. Wang, H. Li, F. Zhang, and Z. Wu. Feature fusion based ensemble method for remaining useful life prediction of machinery. *Applied Soft Computing*, 129:109604, 2022.

- [9] J. Zhu, N. Chen, and W. Peng. Estimation of bearing remaining useful life based on multiscale convolutional neural network. *IEEE Transactions on Industrial Electronics*, 66(4):3208–3216, 2018.
- [10] W. Xu, Q. Jiang, Y. Shen, F. Xu, and Q. Zhu. Rul prediction for rolling bearings based on convolutional autoencoder and status degradation model. *Applied Soft Computing*, 130:109686, 2022.
- [11] M. Ma and Z. Mao. Deep-convolution-based lstm network for remaining useful life prediction. *IEEE Transactions on Industrial Informatics*, 17(3):1658–1667, 2020.
- [12] J. Zhang, J. Tian, M. Li, J. I. Leon, L. G. Franquelo, H. Luo, and S. Yin. A parallel hybrid neural network with integration of spatial and temporal features for remaining useful life prediction in prognostics. *IEEE Transactions on Instrumentation and Measurement*, 72:1–12, 2022.
- [13] X. Su, H. Liu, L. Tao, C. Lu, and M. Suo. An end-to-end framework for remaining useful life prediction of rolling bearing based on feature pre-extraction mechanism and deep adaptive transformer model. *Computers & Industrial Engineering*, 161:107531, 2021.
- [14] Y. Ding and M. Jia. Convolutional transformer: An enhanced attention mechanism architecture for remaining useful life estimation of bearings. *IEEE Transactions on Instrumentation and Measurement*, 71:1–10, 2022.
- [15] J. Zhang, X. Li, J. Tian, H. Luo, and S. Yin. An integrated multi-head dual sparse self-attention network for remaining useful life prediction. *Reliability Engineering & System Safety*, 233:109096, 2023.
- [16] A. Saxena, K. Goebel, D. Simon, and N. Eklund. Damage propagation modeling for aircraft engine run-to-failure simulation. In *Proceedings of the 1st International Conference on Prognostics and Health Management (PHM08)*, Denver, CO, Oct 2008.
- [17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [18] M. Yu, Y. Zhang, and C. Yang. Rolling bearing faults identification based on multiscale singular value. *Advanced Engineering Informatics*, 57:102040, 2023.

- [19] Y. Lei, J. Lin, Z. He, and M. J. Zuo. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical systems and signal processing*, 35(1-2):108–126, 2013.
- [20] D.-T. Hoang and H.-J. Kang. A survey on deep learning based bearing fault diagnosis. *Neurocomputing*, 335:327–335, 2019.
- [21] Z. Liu and L. Zhang. A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings. *Measurement*, 149:107002, 2020.
- [22] D. Neupane and J. Seok. Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. *Ieee Access*, 8:93155–93178, 2020.
- [23] M. M. Islam, A. E. Prosvirin, and J.-M. Kim. Data-driven prognostic scheme for rolling-element bearings using a new health index and variants of least-square support vector machines. *Mechanical Systems and Signal Processing*, 160:107853, 2021.
- [24] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan. Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. volume 28, pages 2306–2318. IEEE, 2016.
- [25] Z. Liu, Y. Cheng, P. Wang, Y. Yu, , and Y. Long. A method for remaining useful life prediction of crystal oscillators using the bayesian approach and extreme learning machine under uncertainty. volume 305, pages 27–38. Elsevier, 2018.
- [26] J. Zhou, Y. Qin, D. Chen, F. Liu, and Q. Qian. Remaining useful life prediction of bearings by a new reinforced memory gru network. *Advanced Engineering Informatics*, 53:101682, 2022.
- [27] J. Shi, T. Yu, K. Goebel, and D. Wu. Remaining useful life prediction of bearings using ensemble learning: The impact of diversity in base learners and features. *Journal of Computing and Information Science in Engineering*, 21(2):021004, 2021.
- [28] T. Wang, Z. Liu, and N. Mrad. A probabilistic framework for remaining useful life prediction of bearings. *IEEE Transactions on Instrumentation and Measurement*, 70:1–12, 2020.
- [29] G. Wang and J. Xiang. Remain useful life prediction of rolling bearings based on exponential model optimized by gradient method. *Measurement*, 176:109161, 2021.

- [30] C. Wang, W. Jiang, X. Yang, and S. Zhang. Rul prediction of rolling bearings based on a dcae and cnn. *Applied Sciences*, 11(23):11516, 2021.
- [31] Y. Wei and D. Wu. Model-based real-time prediction of surface roughness in fused deposition modeling with graph convolutional network-based error correction. *Journal of Manufacturing Systems*, 71:286–297, 2023.
- [32] Y. Wei and D. Wu. State of health and remaining useful life prediction of lithium-ion batteries with conditional graph convolutional network. *Expert Systems with Applications*, 238:122041, 2024.
- [33] H. Lu, V. Barzegar, V. P. Nemanic, C. Hu, S. Laflamme, and A. T. Zimmerman. Joint training of a predictor network and a generative adversarial network for time series forecasting: A case study of bearing prognostics. *Expert Systems with Applications*, 203:117415, 2022.
- [34] X. Yang, Y. Zheng, Y. Zhang, D. S.-H. Wong, and W. Yang. Bearing remaining useful life prediction based on regression shapaelet and graph neural network. *IEEE Transactions on Instrumentation and Measurement*, 71:1–12, 2022.
- [35] L. Liu, X. Song, K. Chen, B. Hou, X. Chai, and H. Ning. An enhanced encoder–decoder framework for bearing remaining useful life prediction. *Measurement*, 170:108753, 2021.
- [36] Y. Zhu, J. Wu, X. Liu, J. Wu, K. Chai, G. Hao, and S. Liu. Hybrid scheme through read-first-lstm encoder-decoder and broad learning system for bearings degradation monitoring and remaining useful life estimation. *Advanced Engineering Informatics*, 56:102014, 2023.
- [37] B. Zhang, S. Zhang, and W. Li. Bearing performance degradation assessment using long short-term memory recurrent network. *Computers in Industry*, 106:14–29, 2019.
- [38] Q. Ni, J. Ji, and K. Feng. Data-driven prognostic scheme for bearings based on a novel health indicator and gated recurrent unit network. *IEEE Transactions on Industrial Informatics*, 19(2):1301–1311, 2022.
- [39] J. Luo and X. Zhang. Convolutional neural network based on attention mechanism and bi-lstm for bearing remaining life prediction. *Applied Intelligence*, pages 1–16, 2022.

- [40] T. Han, J. Pang, and A. C. Tan. Remaining useful life prediction of bearing based on stacked autoencoder and recurrent neural network. *Journal of Manufacturing Systems*, 61:576–591, 2021.
- [41] Y. Cao, Y. Ding, M. Jia, and R. Tian. A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings. *Reliability Engineering & System Safety*, 215:107813, 2021.
- [42] F. O. Heimes. Recurrent neural networks for remaining useful life estimation. In *2008 international conference on prognostics and health management*, pages 1–6. IEEE, 2008.
- [43] H. Lu, V. Barzegar, V. P. Nemani, C. Hu, S. Laflamme, and A. T. Zimmerman. Gan-lstm predictor for failure prognostics of rolling element bearings. In *2021 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pages 1–8. IEEE, 2021.
- [44] H. Chen, J. Zhang, Y. Du, S. Xiang, Z. Yue, N. Zhang, Y. Cai, and Z. Zhang. Understanding the potential of fpga-based spatial acceleration for large language model inference. *ACM Transactions on Reconfigurable Technology and Systems*, 2024.

واژه‌نامه

partition	تقسیم‌بندی	الف
distributed	توزیع شده	اولین تخریب اولین تخریب
big data analytics.....	تحلیل داده‌های بزرگ	آستانه threshold
transformer	ترنسفرمر	ب
generated	تولید شده	برخط online
end of life.....	تخریب کامل	برنامه‌ریزی خطی linear programming
self-attention	توجه به خود	بهینه optimum
multi-head attention.....	توجه چند سر	بیشینه maximum
word to vector.....	تبدیل کلمات به بردار	بلبرینگ bearing
ج		
separable	جداپذیر	پ
black box	جعبه سیاه	پرت outlier
data stream	جویبار داده	پرسمان query
alternating current	جریان متناوب	پوشش cover
ح		
extreme	حدی	پیچیدگی complexity
greedy.....	حریصانه	پردازش سیگنال signal processing
sensor.....	حسگر	پردازش زبان طبیعی natural language processing
خ		
	density	تراکم density
	approximation	تقریب approximation
ت		

dominate.....	غلبه	cluster	خوشه
		linear	خطی

ف

distance	فاصله
----------------	-------

د

space	فضا	داده	داده
		data mining.....	داده‌کاوی

ک

efficient	کارا
minimum.....	کمینه
key	کلید
convolution	کانولوشن

ر

set.....	مجموعه	formal	رسمی
dataset	مجموعه‌داده	decoder	رمزگشا
parallelization.....	موازی‌سازی	encoder	رمزگذار
buffer	میان‌گیر	support vector regression .	رگرسیون بردار پشتیبان

م

model	مدل	sublinear	زیرخطی
wave.....	موج		

س

rate.....	نرخ	hierarchichal	سلسه‌مراتبی
		signal	سیگنال
		noise.....	سروصدا

ه

artificial intelligence.....	هوش مصنوعی
machine learning.....	یادگیری ماشین

ش

object	شیء
network	شبکه
neural network	شبکه عصبی

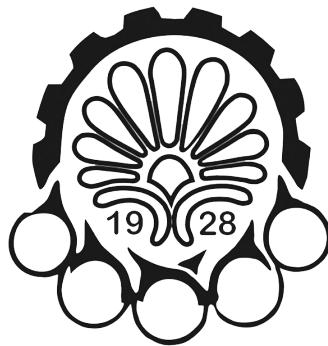
غ

Abstract

In industrial environments and factories, one of the most important concerns is the maintenance and repair of tools and industrial equipment. Each piece of equipment plays a crucial role in the production line of any company, and the smallest damage or malfunction can disrupt the entire production process. Therefore, it is essential to continuously monitor the current health status of the equipment and predict the time of potential failures and the remaining useful life.

The prediction of Remaining Useful Life (RUL) is a critical task in the field of Prognostics and Health Management (PHM) for industrial tools and equipment. Accurate RUL prediction is vital because it enables timely maintenance and repair, reduces downtime, and allows for the optimization of operational efficiency by loading the equipment according to its remaining useful life.

Keywords: Remaining Useful Life, RUL, Transformer, FPGA



Amirkabir University of Technology

(Tehran Polytechnic)

Department of Computer Engineering

M.Sc. Seminar Report

FPGA-Based Hardware Acceleration of Remaining Useful Life Prediction of Rotating Machinery Using Transformer Neural Network

By:

Reza Adinepour

Supervisor:

Prof. Morteza Saheb Zamani

August 2024