دوره ۱۳، شماره پاییز و زمستان ۱۳۹۹

مجله فناوری اطلاعات در طراحی مهندسی
Information Technology in Engineering Design
http://ited.sinaweb.net

پیش‌بینی سرعت خودروها با استفاده از تجزیه مد تجربی مبتنی بریادگیری جمعی

محمد علی کیانی فرد[۱] حسن مطلبی*[۲] وحید خطیبی بردسیری[۳]

(۱) گروه مهندسی کامپیوتر، دانشکده علوم، واحد کرمان، دانشگاه آزاد اسلامی، کرمان، ایران*

(۲) دانشکده مهندسی برق و کامپیوتر، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته ، کرمان، ایران

(۳) گروه مهندسی کامپیوتر، دانشکده علوم، واحد کرمان، دانشگاه آزاد اسلامی، کرمان، ایران

(تاریخ دریافت : ۱۴۰۰/۰۱/۱۷ تاریخ پذیرش : ۱۴۰۰/۰۴/۲۶)

چکیده

پیش‌بینی متوسط سرعت حرکت خودروها، که یکی از مهم‌ترین پارامترهای ترافیکی است، یک موضوع تحقیقاتی کلیدی در سیستم‌های حمل‌ونقل هوشمند است. روش‌های یادگیری جمعی با بهره‌گیری از چندین یادگیر و ترکیب نتایج پیش‌بینی آن‌ها باعث افزایش دقت پیش‌بینی و در نتیجه کاهش خطای پیش‌بینی می‌شوند. در این مقاله، یک مدل جمعی با بکارگیری تجزیه مد تجربی، رگرسیون بردار پشتیبان و متایادگیر ارائه شده است. در این مدل، سری زمانی سرعت خودروها به چندین تابع مد ذاتی و یک باقیمانده تجزیه و به‌عنوان ویژگی های جدید به مجموعه داده اضافه شده است. در مدل جمعی پیشنهادی، مخزن یادگیرها با استفاده از ترکیب روش‌های انتخاب زیرمجموعه تصادفی و انتخاب گام‌به‌گام جهت انتخاب مؤثرترین ویژگی‌های تصادفی از مجموعه داده برای هر یادگیر ساخته می‌شود. نتایج پیش‌بینی دسته‌بندهای پایه متا ویژگی‌ها را ایجاد می‌کند که به یک متایادگیر جهت مشخص کردن نتیجه نهایی سپرده می‌شوند. جهت ارزیابی مدل پیشنهادی، برخی لینک‌های ارتباطی شبکه بزرگراه‌های استراتژیک انگلستان بکار گرفته شده است. نتایج ارزیابی نشان دهنده کاهش خطای روش پیشنهادی نسبت به روش‌های پیشین است.

واژه‌های کلیدی: پیش‌بینی سرعت خودروها، رگرسیون بردار پشتیبان، تجزیه مد تجربی، یادگیری جمعی

* عهده‌دار مکاتبات:

حسن مطلبی

نشانی: دانشکده مهندسی برق و کامپیوتر، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته ، کرمان، ایران

تلفن : ۰۹۱۲۵۷۹۹۰۴۶ پست الکترونیکی: h.motallebi.p@gmail.com

۱- مقدمه

پیش‌بینی دقیق پارامترهای ترافیکی از قبیل جریان ترافیک، متوسط زمان سفر و سرعت حرکت خودروها در جاده‌های ارتباطی می‌تواند به رانندگان و تصمیم‌گیرندگان جهت برنامه‌ریزی مسیر حرکت و مدیریت ترافیک کمک کند [۱]. از آنجایی که متوسط سرعت خودروها در یک لینک ارتباطی نشان دهنده وضعیت ترافیک از لحاظ روان و یا سنگین بودن ترافیک است، پیش‌بینی دقیق سرعت خودروها باعث کاهش هزینه حمل‌ونقل، افزایش سرعت و کاهش آلایندگی می‌شود. تاکنون، روش‌های یادگیری ماشین متعددی جهت پیش‌بینی کوتاه‌مدت [۲] و بلندمدت [۳] ترافیک ارائه گردیده است. در این مقاله، به ارایه روشی برای پیش‌بینی سرعت خودروها با استفاده از تجزیه مد تجربی مبتنی بر یادگیری جمعی می‌پردازیم.

روش‌های یادگیری جمعی (دسته‌بندی الگوها، سری‌های زمانی و رگرسیون) با بهره‌گیری از چندین الگوریتم یادگیری، دقت پیش‌بینی بیشتری نسبت به روش‌های تک یادگیری بدست می‌آورند. با استفاده از روش‌های آماری، محاسباتی، تجزیه واریانس و قدرت همبستگی می‌توان این ادعا را اثبات کرد [٤]. روش‌های یادگیری جمعی به چند گروه تقسیم می‌شوند: جنگل تصادفی، بوستینگ، بگینگ، استکینگ، روش‌های تجزیه سری زمانی، روش‌های یادگیری همبستگی منفی، روش‌های جمعی مبتنی بر بهینه‌سازی چند هدفی، روش‌های جمعی فازی، روش‌های یادگیری چندهسته‌ای و روش‌های جمعی مبتنی بر یادگیری عمیق [۵].

در روش جمعی استکینگ، با به‌کارگیری نتایج پیش‌بینی یادگیرهای پایه ویژگی‌های جدیدی بنام متاویژگی ایجاد می‌شود که با حفظ برچسب نمونه‌های اصلی، مجموعه داده جدیدی ایجاد می‌نمایند. یک متایادگیر مجموعه داده جدید را مورد آموزش قرار داده و نتیجه پیش‌بینی را مشخص می‌کند [٦].

انتخاب زیرمجموعه تصادفی در روش‌های یادگیری جمعی جهت انتخاب تصادفی یک‌سری از ویژگی‌های مجموعه داده مورد استفاده قرار می‌گیرد. در این روش زیرمجموعه‌های به وجود آمده به یادگیرهای مختلف جهت ساخت یادگیری جمعی سپرده می‌شوند [۷]. ازآنجایی‌که برخی از ویژگی‌های ایجادشده تأثیر چندانی درنتیجه پیش‌بینی ندارند، برای انتخاب مؤثرترین ویژگی‌ها از روش رگرسیون گام‌به‌گام استفاده شده می‌شود. در این روش همه متغیرهای مستقل وارد تحلیل می‌شوند و در روند محاسبات، متغیر مستقلی که تأثیری محسوس در ارتباط با متغیر وابسته نداشته باشد از تحلیل حذف شده و وارد معادله نمی‌شود.

روش‌های یادگیری جمعی مبتنی بر تجزیه به دو دسته روش‌های سلسله مراتبی جمعی و مدل‌های تقسیم و حل تقسیم می‌شوند. در تجزیه مد تجربی[۱]، سری زمانی به مجموعه‌ای از سری‌های زمانی ساده‌تر تبدیل می‌شود. سپس هرکدام از سری‌های زمانی آموزش داده می‌شوند و درنهایت جهت به‌دست آوردن نتیجه نهایی پیش‌بینی، نتایج تک‌تک یادگیرها با هم جمع‌بندی می‌شوند [۸].

[۱] Empirical Mode Decomposition

تجزیه مد تجربی، تعدادی تابع مد ذاتی[۲] بر روی فرکانس و یک باقیمانده[۳] نهفته در سیگنال را ایجاد می‌کند، بنابراین، تجزیه را می‌توان به‌عنوان یک گسترش داده از طریق توابع مد ذاتی و یک باقیمانده به‌عنوان یک تحلیل کلی فرکانس زمان در نظر گرفت [۹]. تحقیقات متعددی در خصوص پیش‌بینی با استفاده از تجزیه مد ذاتی [۸-۱۱ و٤] انجام شده است. در همه آن‌ها توابع مد ذاتی را به یک یادگیر واگذار کرده و نتایج به‌دست‌آمده را با هم جمع می‌نمایند، ما در این مقاله با استفاده از تجزیه مد تجربی، سری زمانی سرعت خودروها را تجزیه و ویژگی‌های جدیدی ایجاد کرده‌ایم.

دنباله سرعت حرکت خودروها یک سری زمانی است که شامل داده‌های ترتیبی وابسته به زمان است. از طرفی رگرسیون خطی به دلیل نیاز به زمان پردازش و حافظه کم، یک روش محبوب برای پیش‌بینی این نوع داده است، ولی از آنجایی که داده‌های ترافیکی از نوع داده‌های غیرخطی هستند از روش رگرسیون بردار پشتیبان (SRN) که نوعی از روش ماشین بردار پشتیبان[٤] است استفاده کرده‌ایم. رگرسیون بردار پشتیبان ایده‌ای شبیه به ماشین بردار پشتیبان دارد. رگرسیون بردار پشتیبان از توابع کرنل[٥] جهت انتقال داده به فضای ویژگی با ابعاد بالاتر استفاده می‌کند [۱۲].

در این تحقیق، مجموعه داده بزرگراه‌های انگلستان[٦]، که شامل نه ویژگی است، با هدف ارزیابی میزان افزایش دقت پیش‌بینی مورد استفاده قرار گرفته است. ویژگی سرعت حرکت خودروها با استفاده از تجزیه مبتنی بر تجربه به چند ویژگی جدید تجزیه می‌شود که در کنار داده‌های بلادرنگ[۷] و داده‌های تاریخی[۸] (سرعت خودروها در هشت دوره زمانی (۱۵ دقیقه) قبل) جهت پیش‌بینی سرعت خودروها در دوره یا دوره‌های زمانی بعدی استفاده می‌شود. این مقاله در پنج بخش تنظیم شده است، در ادامه ابتدا در قسمت دوم پیشینه تحقیق و سپس در قسمت سوم به بحث در خصوص مدل پیشنهادی شامل تجزیه مد تجربی، داده‌های مورداستفاده و روش کار پرداخته می‌شود. در قسمت چهارم مقاله روش پیشنهادی مورد ارزیابی قرارگرفته و در قسمت پایانی نتیجه‌گیری کلی ارائه می‌شود.

۲- پیشینه تحقیق

تحقیقات گوناگونی در زمینه پیش‌بینی الگوهای ترافیکی تلاش دارند پارامترهای مختلف شامل تعداد خودرو در کیلومتر، جریان ترافیک و متوسط زمان سفر [۲] را پیش‌بینی کنند. در ادامه، مختصراً به برخی از این تحقیقات می‌پردازیم.

هدف از پیش‌بینی جریان[۹] ترافیک، تخمین تعداد وسایل نقلیه در یک فاصله زمانی و یک منطقه خاص است. در [۱۳] یک روش پیش‌بینی جریان ترافیک کوتاه‌مدت بر مبنای SVR و PSO پیشنهاد شده است که در آن به‌منظور دستیابی به‌دقت بالا و کاهش زمان، از بهینه‌سازی حرکت ذرات جهت جستجوی بهینه پارامترهای رگرسیون بردار پشتیبان استفاده شده است. طول دوره پیش‌بینی با دقت بالا در پیش‌بینی ترافیک کوتاه‌مدت به فاصله زمانی پنج تا پانزده دقیقه محدود می‌شود [١٤-١٥]. در

---

۲ Intrinsic Mode Function
۳ Residue
٤ Support Vector Regression
٥ Kernel Function
٦ Strategy Road Network (SRN)
۷ Real time
۸ Historical
۹ Flow

[۳] نویسندگان یک الگوریتم پیش‌بینی جریان ترافیک با دقت بالا و زمان پیش‌بینی طولانی‌تر ارائه داده‌اند. آن‌ها یک مدل ترکیبی بر مبنای LSTM و ARIMA ارائه کرده‌اند که می‌تواند نرخ دقت بالا بر اساس افزایش زمان پیش‌بینی و طول بازه زمانی را تضمین کند. همچنین، در[۱۵] مدلی برای پیش‌بینی جریان ترافیک با دقت بالاتر و زمان پیش‌بینی طولانی‌تر ارائه شده است.

در [۱٤] نویسندگان از روش gradient boosting جهت تحلیل و مدل کردن زمان سفر[۱۰] استفاده کرده‌اند. همچنین، [۱٦] بر اساس ماشین بردار پشتیبان یک مدل پیش‌بینی زمان سفر پویا پیشنهاد داده است. در [۱۷] یک روش برای تخمین میانگین زمان سفر ارائه شده است.

برخی تحقیقات، ازجمله [۱۲]، با استفاده از مقدار پارامتر ترافیکی در بازه‌های زمانی قبلی، سعی در پیش‌بینی کوتاه‌مدت پارامتر در بازه یا بازه‌های زمانی بعدی دارند. در مقاله مذکور، با استفاده از رگرسیون بردار پشتیبان، پارامتر زمان سفر در پنج دقیقه آینده بر اساس مقادیر زمان سفر در ٤۰ دقیقه (هشت دوره زمانی پنج دقیقه‌ای) قبل پیش‌بینی شده است.

شبکه‌های عصبی در تحقیقات فراوانی برای پیش‌بینی ترافیک بکار گرفته شده است، در [۲] نویسندگان ساختار انواع مختلف شبکه‌های عصبی بکار گرفته شده برای پیش‌بینی حالت ترافیکی کوتاه‌مدت را بررسی کرده‌اند. همچنین، در تحقیقات مختلفی از رگرسیون جهت پیش‌بینی ترافیک استفاده شده است. رگرسیون بردار پشتیبان در [۱۲] حداقل مربعات جزئی PLS تعمیم رگرسیون خطی چندگانه MLR در [۱۸] جهت پیش‌بینی ترافیک کوتاه‌مدت بکار گرفته شده است. در [۱۹] یک مدل یادگیری عمیق برای پیش‌بینی کوتاه‌مدت سرعت ترافیک پیشنهاد شده است.

در مقالات متعددی، که در آن‌ها داده‌ها از نوع سری زمانی هستند، از تجزیه مد تجربی استفاده شده است. در زمینه‌های گوناگونی مثل پیش‌بینی تقاضای بار [۱۰]، پیش‌بینی سرعت باد [۸] و همچنین در خصوص پیش‌بینی پارامترهای ترافیکی [۲۰و۱۱] تجزیه مد تجربی به‌عنوان روشی قدرتمند برای کاهش نویزها، فیلتر کردن روند و پیش‌بینی در تجزیه و تحلیل داده‌های تجربی و پیچیده استفاده شده است. در [۱۱] یک روش پیش‌بینی ترکیبی مبتنی بر شبکه‌های عصبی و تجزیه مد تجربی ارائه شده است. همچنین، [۲۰] با استفاده از روش تجزیه مد ذاتی و بهینه‌سازی حرکت ذرات و ماشین بردار پشتیبان اقدام به پیش‌بینی جریان ترافیک کرده است. در همه مقاله‌هایی که از روش‌های مبتنی بر تجزیه استفاده کرده‌اند، هر کدام از سری‌های زمانی ایجاد شده به یک یادگیر اختصاص داده می‌شوند و نتایج یادگیرها تجمیع شده و نتیجه نهایی را ایجاد می‌کند. برخلاف این روش‌ها، در این مقاله از سری‌های زمانی ایجادشده به‌عنوان ویژگی‌های جدید استفاده و مجموعه داده جدیدی ایجاد کرده‌ایم.

برخی تحقیقات، از جمله [۲۱-۲۳]، به استفاده از مدل‌های یادگیری جمعی در پیش‌بینی پارامترهای ترافیکی پرداخته‌اند. [۲۳] با ترکیب شبکه‌های عصبی مصنوعی (ANN) و روش‌های آماری جریان ترافیک یک ساعته را پیش‌بینی کرده است. در [۲۱] یک چارچوب آنلاین ارائه شده که قادر است با استفاده از موقعیت فعلی و داده‌های ترافیک تاریخی، آموزش ببیند و ترافیک آینده را پیش‌بینی کند. یک روش جمعی ناهمگون در [۲۲] ارائه شده که در آن پیش‌بینی زمان سفر طولانی[۱۱] با الگوریتم‌ها و پارامترهای متفاوتی انجام شده و در نهایت بهترین روش برای هر مسیر مشخص می‌شود.

---

۱۰ Journey time

۱۱ Long term

۳- شیوه پیشنهادی

۳-۱- مجموعه داده

مجموعه داده مورد استفاده در این مقاله، مجموعه داده شبکه جاده‌های استراتژیک انگلستان است[۱۲]، که از طریق سرویس اطلاعاتی ترافیک ملی انگلستان در دسترس است [۲۵]. این مجموعه داده شامل متوسط زمان سفر، سرعت خودروها و جریان ترافیک است که هر ۱۵ دقیقه یک‌بار جمع‌آوری‌شده است. نام لینک ارتباطی، شرح لینک، طول لینک، بازه زمانی، تاریخ و کیفیت جمع‌آوری اطلاعات، دیگر ویژگی‌های موجود در مجموعه داده هستند. در جدول ۱ مشخصات این ویژگی‌ها شرح داده شده‌اند.

جدول ۱: ویژگی‌های مجموعه داده اولیه

ویژگی‌ها	شرح
LinkRef	شناسه لینک
LinkDescription	شرح لینک
TimePeriod	بازه زمانی (۱۵ دقیقه)
Date	تاریخ سفر
AverageJT	متوسط زمان سفر
AverageSpeed	سرعت خودروها
DataQuality	کیفیت داده‌ها
LinkLenght	طول لینک (کیلومتر)
Flow	متوسط تعداد خودرو

[۱۲] https://data.gov.uk/dataset/dft-eng-srn-routes-journey-times

در این مقاله، نام لینک که از حروف الفبا تشکیل شده، به عدد تبدیل و ویژگی شرح لینک را حذف کرده‌ایم، زمان که به‌صورت ۰۰:۰۰–۲٤:۰۰ است به ۹٦ بازه، (۰–۹۵) که نمایانگر ۹٦ ربع ساعت در طول ۲٤ ساعت شبانه‌روز است، گسسته‌سازی شده است [۲٤]. از ویژگی تاریخ، ویژگی‌های روز هفته، روز از ماه و روز از سال ایجاد شده است.

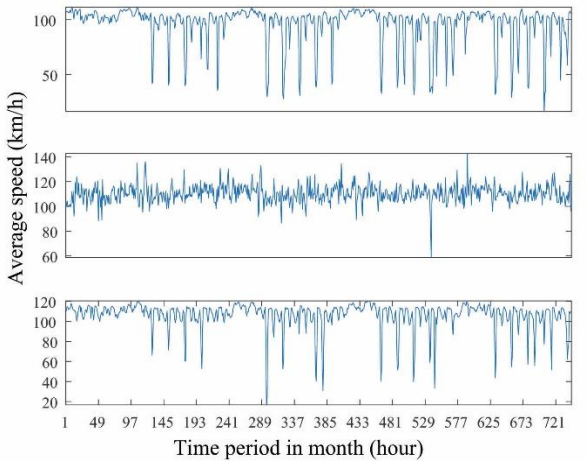
ویژگی سرعت خودروها در هشت بازه زمانی ربع ساعته قبل به‌عنوان هشت ویژگی تاریخی به مجموعه داده اضافه شده است. با استفاده از تجزیه مد تجربی، n تابع مد ذاتی و یک ویژگی باقیمانده ایجاد و به‌عنوان ویژگی‌های جدید به مجموعه داده اضافه شده است. این ویژگی‌ها در جدول ۲ نشان داده شده‌اند.

جدول ۲: ویژگی‌های مجموعه داده جدید

ویژگی	شرح
LinkNumber	شماره لینک
TimePeriod	بازه زمانی ۹٦–۰
DayOfWeek	روز هفته
AverageJT	متوسط زمان سفر
AverageSpeed	سرعت خودروها
DataQuality	کیفیت داده‌ها
LinkLenght	طول لینک
Flow	متوسط تعداد خودرو در کیلومتر
Q۱-Q۸	سرعت خودروها در ۸ بازه زمانی قبل
IMF۱-IMFn	ویژگی ایجادشده با تجزیه مد ذاتیN
Residue	باقیمانده در تجزیه مد ذاتی

Target	**سرعت خودروها در ۱ تا ۸ بازه پیش رو**

ویژگی متوسط سرعت حرکت خودروها از نوع سری‌های زمانی است و در طول هفته و روز یک الگوی تکراری ایجاد می‌کند. به عنوان مثال، الگوی تکراری سرعت خودروها در طول یک ماه برای سه لینک متفاوت در شکل ۱ نشان داده شده است.



شکل ۱: سرعت خودروها در سه لینک تصادفی در دوره‌های زمانی یک ماه

۲-۳- تجزیه مد تجربی

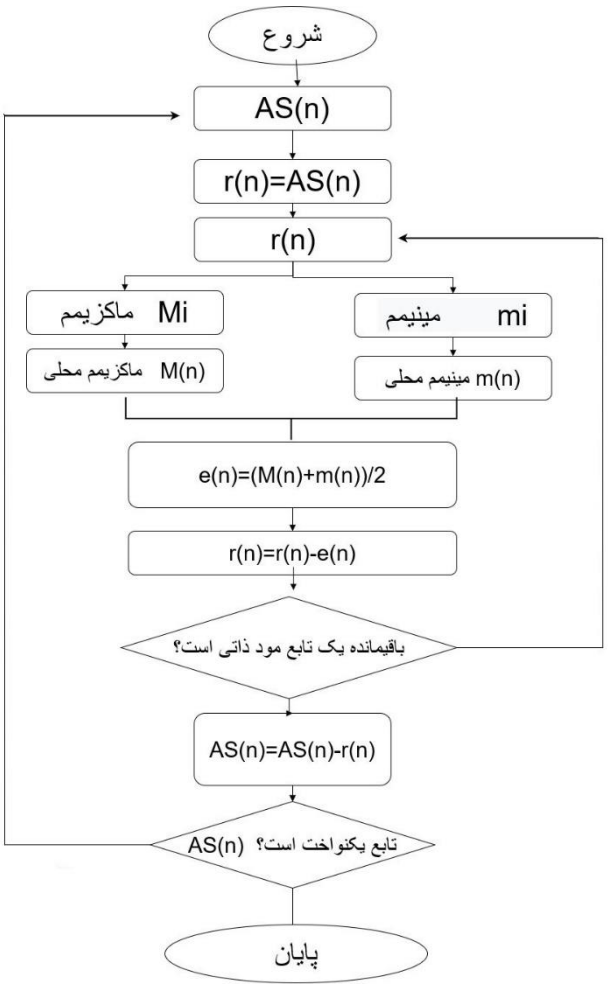
تجزیه مد تجربی، یک روش پردازش سیگنال جهت تجزیه و تحلیل سیگنال‌های غیرخطی است که توانایی تجزیه سیگنال به تعدادی سیگنال تک قطعه‌ای به نام تابع مد ذاتی را دارد. تابع مد ذاتی نشان‌دهنده حالت‌های نوسانی ساده تعبیه‌شده در سیگنال است [۲۰]. مراحل انجام کار در این روش مطابق شکل ۲ است.

در این بحث، AS(n) بیانگر سری زمانی مربوط به سرعت خودروها در یک لینک است، به‌وسیله درون‌یابی ماکزیمم و مینیمم‌ها، ماکزیمم محلی و مینیمم محلی Mi, mi تولید می‌شوند، سپس e(n) یعنی میانگین آن‌ها محاسبه می‌شود. r(n) یک کاندید جهت ایجاد تابع مد ذاتی است که از کم کردن مقدار سرعت خودروها و میانگین به دست می‌آید. در هر مرحله، اگر کاندید موردنظر مد ذاتی بود مقدار r(n) از مقدار متوسط کسر می‌گردد. این مراحل تا رسیدن به معیارهای توقف از پیش تعیین‌شده ادامه دارد [۹]. در نهایت سری زمانی سرعت خودروها مطابق فرمول ۱ است

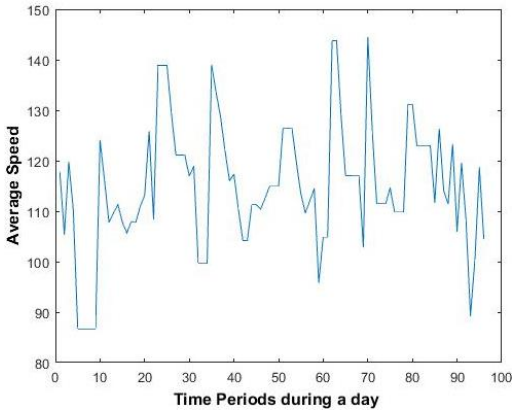
$$\boldsymbol{AS(n) = \sum_{k=1}^{n} IMF(k) + r(n)} \qquad (۱)$$

در این مقاله، تعداد هفته‌های آموزش به‌عنوان پنجره زمانی در نظر گرفته می‌شوند، سری زمانی سرعت خودروها در پنجره زمانی تجزیه شده و مدل آموزش می‌بیند. شکل ۳ نمونه‌ای از سری زمانی سرعت خودروها و شکل ٤ توابع مد ذاتی و باقیمانده حاصل از تجزیه شکل

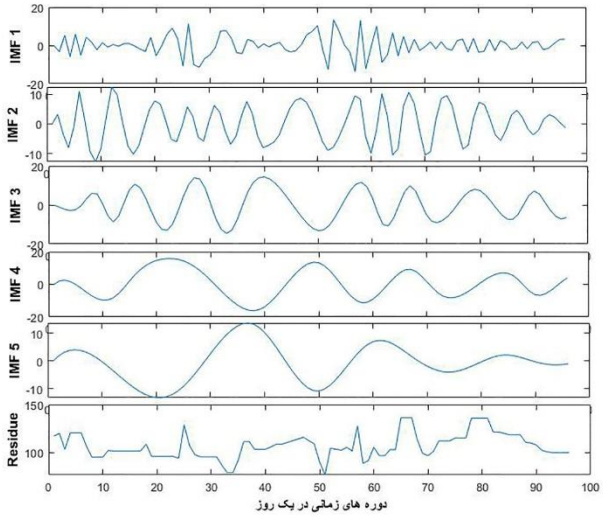
۳ را نشان می‌دهد. با ورود نمونه تست جهت ارزیابی، سابقه تاریخی آن به نحوی در نظر گرفته می‌شود که پنجره زمانی کامل شود و سپس مقادیر ویژگی‌های حاصل از تجزیه مربوط به نمونه تست ساخته می‌شوند.



شکل ۲: مراحل کار در تجزیه مد تجربی



شکل ۳: متوسط سرعت در لینک LM۲۷ در ۹۶ دوره زمانی یک روز



شکل ٤: نمونه‌ای از توابع مد ذاتی و باقیمانده بعد از اعمال تجزیه مد تجربی در محدوده زمانی یک روز

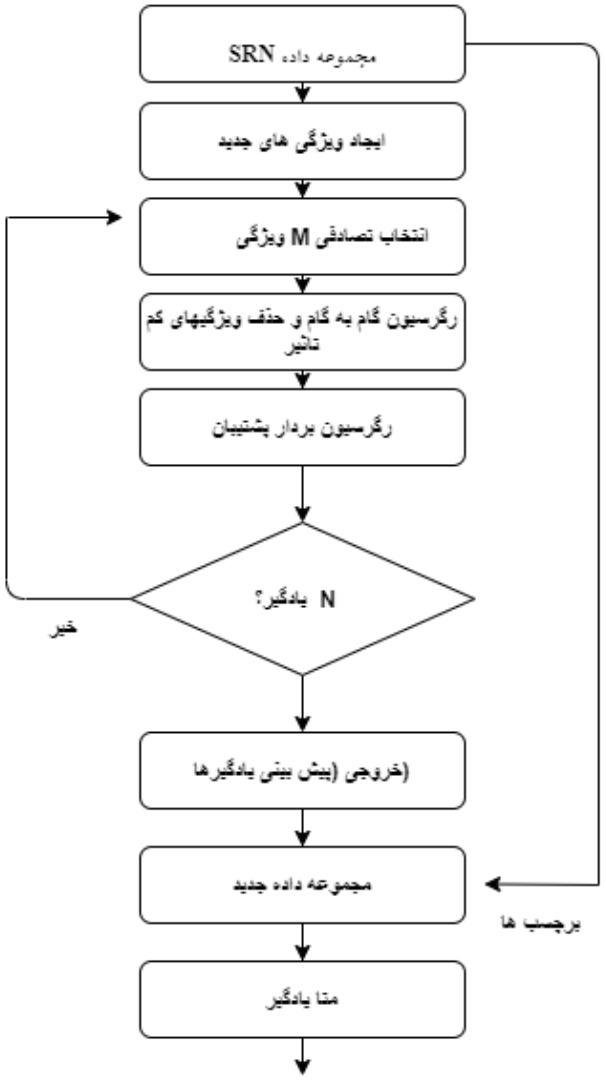
۳-۳- روش کار

جهت پیش‌بینی سرعت خودروها در دوره‌های زمانی ربع ساعته پیش رو از مجموعه داده ایجاد شده مطابق جدول ۲ استفاده کرده‌ایم. طرح کلی روش در شکل ۵ نشان داده شده است. برای ایجاد مدل جمعی و ساخت دسته‌بندها روش‌های متفاوتی وجود دارد که شامل سه مرحله ۱) ساخت مخزن یادگیرها[۱۳]، ۲) انتخاب یادگیرهای شایسته‌تر و ۳) جمع‌بندی نتایج یادگیرها هستند. در این مقاله، از روش استکینگ و تفاوت در ویژگی‌ها استفاده کرده‌ایم. ابتدا با استفاده از روش انتخاب زیرمجموعه‌های تصادفی، تعدادی ویژگی (f) را انتخاب کرده و سپس با روش گام‌به‌گام مؤثرترین آن‌ها انتخاب شده‌اند. خروجی یادگیرهای پایه با حفظ برچسب، متا ویژگی‌ها را تشکیل می دهند و در اختیار یک متا یادگیر جهت آموزش و مشخص کردن خروجی نهایی پیش‌بینی قرارمی‌گیرند. رگرسیون بردار پشتیبان با تابع

[۱۳] Pool of leaners

کرنل

خطی به‌عنوان یادگیر پایه و متا مورداستفاده قرار گرفته است.



شکل ۵: طرح کلی مدل پیشنهادی

٤- ارزیابی کارایی مدل پیشنهادی

در این مقاله، برای ارزیابی دقت پیش‌بینی مدل پیشنهادی از معیارهای خطای جذر میانگین مربعات و درصد میانگین مطلق خطا استفاده شده است، فرمول‌های مربوط به این دو معیار در معادلات ۲ و ۳ آورده شده است.

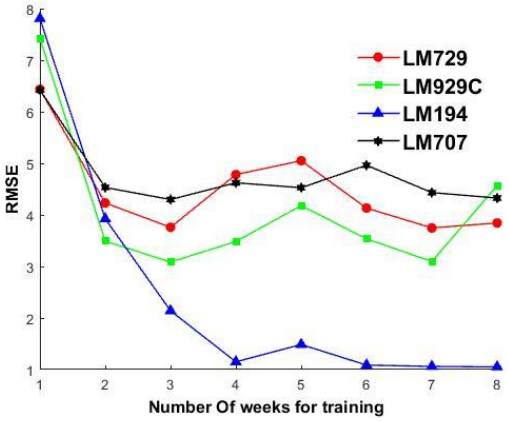
$$(۲)\quad RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|Pi - Ai|^2}$$

$$(۳)\quad MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Ai-Pi}{Ai}\right|$$

در این فرمول‌ها **P** بیانگر مقدار پیش‌بینی شده و **A** بیانگر مقدار واقعی نمونه‌ها است. در ادامه، ابتدا تأثیر عواملی از جمله تعداد روزهای آموزش، تعداد یادگیر و تعداد ویژگی‌ها بر نتایج پیش‌بینی مورد بررسی قرار می‌گیرد و سپس مدل پیشنهادی با مدل‌های پایه و **ARIMA** مقایسه می‌شود. در نهایت مدل پیشنهادی برای پیش‌بینی یک تا هشت گام پیش‌رو مورد ارزیابی قرار می‌گیرد.

٤-١- تأثیر افزایش روزهای آموزش بر نتایج پیش‌بینی

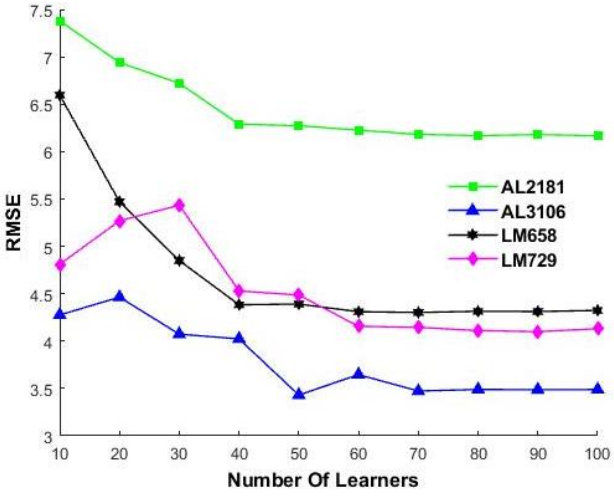
روش پیشنهادی برای چهار لینک تصادفی از مجموعه داده **SRN** مورد ارزیابی قرارگرفته و نتایج ارزیابی در شکل ٦ نمایش داده شده است. در این ارزیابی تعداد هفته مجموعه داده آموزش که همان پنجره زمانی است از یک تا هشت هفته مورد آموزش قرار گرفته است. میزان خطای روش پیشنهادی نشان دهنده کاهش خطا با افزایش تعداد هفته داده آموزش است.



شکل ٦: تأثیر افزایش هفته‌های آموزش بر کاهش خطا

٤-٢- تأثیر تعداد یادگیر بر کاهش خطای پیش‌بینی

در این مرحله، روش پیشنهادی برای لینک‌های مرحله قبل با ثابت نگه داشتن پنجره‌زمانی به چهار هفته و تغییر تعداد یادگیر در مخزن دسته‌بندها مورد ارزیابی قرار گرفته است.

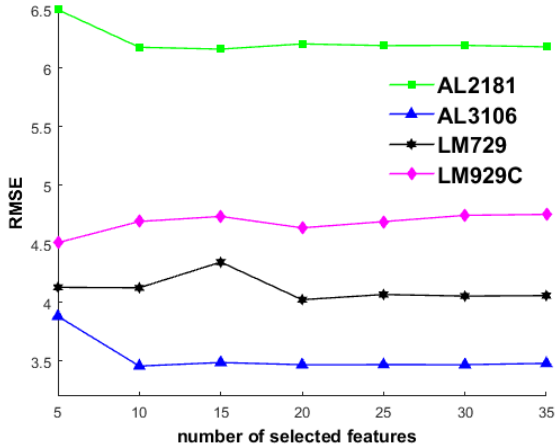


شکل ۷: تأثیر تعداد یادگیر یادگیری جمعی بر کاهش خطا

نتایج این ارزیابی در شکل ۷ نشان داده‌شده است. نوسانات نمودارها نشان‌دهنده تأثیر متفاوت تعداد یادگیر بر میزان خطای لینک‌های متفاوت است. در نمودار فوق بهترین حالت برای چهار لینک از شش لینک مورد بررسی تعداد ۵۰ یادگیر است.

٤-۳- تأثیر تعداد ویژگی تصادفی انتخاب‌شده در هر مرحله بر نتایج پیش‌بینی

تقریباً ٤٦ ویژگی در مجموعه داده جدید ایجاد شده وجود دارد. مدل پیشنهادی برای چهار لینک تصادفی در حالات مختلف (تعداد ویژگی انتخابی برای یادگیرها) مورد ارزیابی قرار گرفته است. برای برخی لینک‌ها تعداد ویژگی تأثیر بسزایی در میزان خطا ندارد. شکل ۸ نشان دهنده تأثیر تعداد ویژگی انتخابی بر میزان خطای RMSE است. نمودار نشان دهنده این واقعیت است که بهترین تعداد ویژگی در لینک‌های مورد آموزش بین ۵ تا ۱۵ است.



شکل ۸: تأثیر تعداد ویژگی بر خطای RMSE

۵-۱- مقایسه روش پیشنهادی و روش‌های موجود

در قسمت‌های ۱-٤ تا ۳-٤ اثر سه متغیر تعداد هفته آموزش، تعداد ویژگی در هر مرحله و تعداد یادگیر جهت ساخت مخزن یادگیرها مورد بررسی و ارزیابی قرار گرفت، در این قسمت با در نظر گرفتن مقدار بهینه برای پارامترهای فوق برای هر لینک اقدام به مقایسه روش پیشنهادی با روش‌های دیگر می‌کنیم.

پیش‌بینی سرعت خودروها در این تحقیق مورد ارزیابی قرار گرفته است. در این آزمایش، داده‌های چهار هفته اول سال ۲۰۱٤ به‌عنوان داده‌های آموزشی و داده‌های هفته پنجم به‌عنوان داده‌های تست در نظر گرفته‌شده است. چند لینک ارتباطی به‌صورت تصادفی انتخاب و روش پیشنهادی را برای آن‌ها بکار گرفته‌ایم. نتایج پیش‌بینی هرکدام با روش رگرسیون درختی، رگرسیون بردار پشتیبان، روش جمعی بگینگ، شبکه عصبی لونبرگ[۱٤]، روش میانگین حرکت[۱۵] و روش حاصل از ترکیب ARIMA با EMD (که در اکثر مقاله‌ها بکار گرفته شده است) مقایسه و نتایج حاصل از آن در جدول ۳ نشان داده‌شده است. نتایج ارزیابی حاکی از آن است که در صورت استفاده از تجزیه مد تجربی به‌عنوان ویژگی‌های آموزش بهبود قابل‌توجهی خواهیم داشت و در صورتی که علاوه بر تجزیه مد ذاتی از یادگیری جمعی استفاده کنیم نتیجه بهتری حاصل خواهد شد.

روش‌های رگرسیون درختی، رگرسیون بردار پشتیبان، روش جمعی بگینگ، شبکه عصبی لونبرگ به دو صورت عادی و با استفاده از ویژگی های به‌دست‌آمده از تجزیه مد تجربی مورد ارزیابی قرارگرفته‌اند. جهت نمایش بهتر خطای پیش‌بینی، مقادیر واقعی سرعت خودروها و مقادیر پیش‌بینی شده برای دو لینک ارتباطی در طی یک روز (۹٦ بازه زمانی ربع ساعته ۲۹ ژانویه) مقایسه و در شکل‌های ۹ و ۱۰ نشان داده شده است.

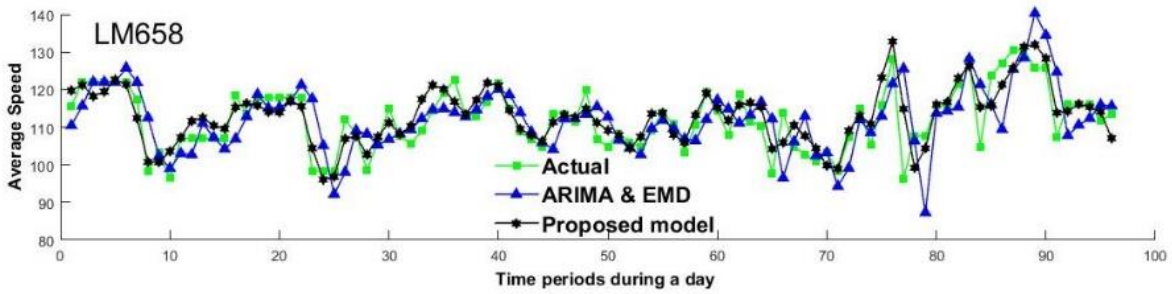
جدول ۳: مقایسه خطای پیش‌بینی مدل پیشنهادی و مدل‌های پیشین برای چند لینک ارتباطی بر اساس معیارهای RMSE و MAPE

Link	AL۳۱۰٦		LM٦۵۸		LM۷۰۷		LM۷۲۹		LM۹۲۹C		AL۲۱۸۱	
Method	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Ensemble bagging	۸,۲٦٦	۰,۰۸۳	۱۱,۳۵۲	۰,۰۸۱	۱۰,۱۷٦	۰,۰۷۰	۱۱,۷۹٤	۰,۰۹۰	٤,۸۸۷	۰,۰٤۸	۱۲,۲۱۷	۰,۱٤۰
Tree regression	۸,۲۷٤	۰,۰۷۹	۹,٦۲۲	۰,۰۷۰	۸,۸۷۲	۰,۰٦۱	۱۱,۲٦۸	۰,۰۸۱	۵,۷۷۹	۰,۰۵٦	۱۱,۳۹۵	۰,۱۲۹
SVR	۷,۰۹۰	۰,۰٦۵	۷,۲۷۲	۰,۰۵۲	۷,۹۳۰	۰,۰٤۹	۹,۱۵۸	۰,۰٦۳	۳,٦۸۲	۰,۰۳۵	۱۰,۵۵۲	۰,۱۳٤
NN	۷,۱۳۹	۰,۰۷٦	۸,۳۳۵	۰,۰٦۱	۷,۵۵۹	۰,۰۵۳	۸,۳۱۵	۰,۰٦۱	۵,۱٤٤	۰,۰٤۸	۱۱,۱۱٦	۰,۱۱۸
Ensemble & EMD	۷,۷۱۲	۰,۰۸٤	۷,۲۱۹	۰,۰۵۰	۹,۲۰٤	۰,۰٦٤	۸,۷۸۷	۰,۰٦۵	۳,۸۲٤	۰,۰۳۷	۱۲,٦۹٦	۰,۱۳۷

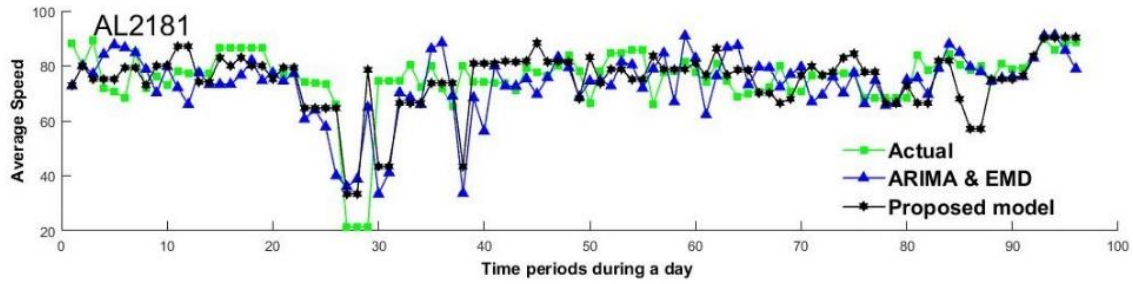
[۱٤] Levenberg-Marquardt backpropagation

[۱۵] Autoregressive integrated moving average

Tree & EMD	۷,٤۳۹	۰,۰۷٦	۹,۲۵۲	۰,۰٦٤	۸,۳۲۰	۰,۰٤۹	۷,۹٤٦	۰,۰۵۹	۳,۹۲۲	۰,۰۳۷	۱۱,۸٤۲	۰,۱٤۵
SVR & EMD	٤,۱٦٦	۰,۰٤۱	٤,۵۵٤	۰,۰۳۳	۵,٦۱۵	۰,۰۳٦	۳,۹۷۱	۰,۲۸۷	۲,۳۳٦	۰,۰۲۳	۹,۰۰۰	۰,۱۱٦
NN &EMD	٦,۹٤۳	۰,۰۷۸	۵,۲۷۹	۰,۰۳۸	٦,۷۳٦	۰,۰٤۷	۸,۰۱۵	۰,۰٦۱	۲,۷۱٦	۰,۰۲٦	۸,۹۳۱	۰,۰۹۵
ARIMA	٦,۸۸۸	۰,۰٦۳	۷,٤۷۳	۰,۰۵۱	۷,۸۸۹	۰,۰٤۹	۹,۱۱۹	۰,۰٦٤	۳,۷۷۰	۰,۰۳٤	۱٤,۵۸۱	۰,۱۷٤
ARIMA-EMD	۳,٤٤٤	۰,۰۳٦	۵,۱۰۸	۰,۰۳٤	۵,۸۱۹	۰,۰۳۷	٦,۸۹۸	۰,۰٤۹	۲,٤۱۹	۰,۰۲۳	۹,۵۰۱	۰,۱۰٤
Proposed Ensemble	۲,۸۹۹	۰,۰۲۸	٤,٤۰٤	۰,۰۳۱	۵,۳۳۱	۰,۰۳۰	۳,٤٦۳	۰,۰۲۲۸	۲,۲۸۲	۰,۰۲۱	۸,۹۸٤	۰,۱۱۷



شکل ۹: مقادیر واقعی و مقادیر پیش‌بینی شده با روش پیشنهادی و روش ARIMA & EMD



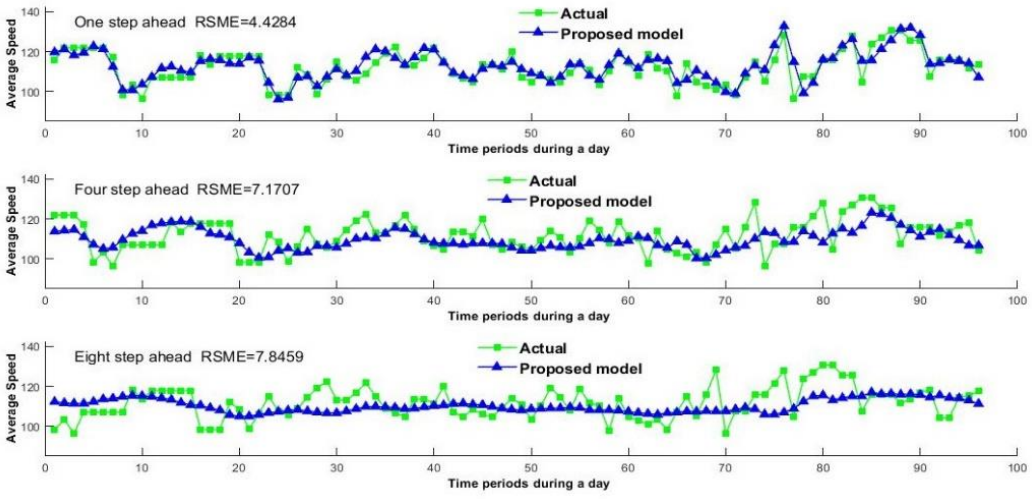
شکل ۱۰: مقادیر واقعی و مقادیر پیش‌بینی با روش پیشنهادی و روش ARIMA & EMD

٦-۱- پیش‌بینی گام‌های پیش رو

تاکنون به بحث پیش‌بینی یک مرحله پیش رو یعنی ربع ساعت آینده پرداختیم. نتایج نشان داده شده در قسمت‌های قبل نیز فقط مربوط به پیش‌بینی یک مرحله پیش‌رو است. در این قسمت، مدل‌های مختلف را برای پیش‌بینی یک تا هشت مرحله پیش رو مورد مقایسه قرار داده‌ایم، نتایج پیش‌بینی برای لینک LM۶۵۸ در جدول ۴ نشان داده شده است.

# Steps ahead	۱	۲	۳	٤	٥	٦	۷	۸
Ensemble bagging	۱۱,۳٥۲	۹,۲۹۹	۱۱,٥۰٦	۱۱,٤۰۸	۱۰,۷۳۲	۱۱,۱۸۹	۱۳,٥۲٤	۱۱,۰۹٤
Tree regression	۹,٦۲۲	۱۰,۳٤۹	۱۱,۲۹۰	۱۱,٥٤۱	۱۰,٦٤۸	۱۱,۹٦۹	۱۰,۰۹۱	۹,٥۳۹
SVR	۷,۲۷۲	۸,۱٥۲	۸,۳۳۷	۸,٥۲۷	۸,٤۷۳	۸,۳۷۲	۸,۱۰۲	۸,۱۷۹
NN	۹,٦٥۲	۸,۲۸۷	۸,٥۷۹	۷,۹۱۲	۸,۰٤۰	۸,٦٤٤	۸,۲۸٥	۸,٤۹۲
Ensemble & EMD	۷,۲۱۹	۹,۱٤٥	۱۳,۷۳۲	۹,۹۰٤	۱۱,٥۷۲	۱۲,۱۲٦	۱۱,٥۸۷	۱۳,۱۰۱
Tree & EMD	۹,۲٥۲	۱۱,۲۷۱	۱۲,۰۱۲	۱۱,٦۷۸	۱۱,۸۰٤	۱۲,٤٤۳	۱۱,۳۱٤	۱۳,۲٤۸
SVR & EMD	٤,٥٥٤	٦,۱٤٥	٦,٥۳۰	۷,۱۲۸	۷,۲٤۰	۷,۸٥۰	۷,٥۷۸	۷,٦۱۰
NN &EMD	٥,۳۲٦	۱۰,۷۹۳	۸,٥۸۱	۲۰,٦٥٦	۸,٤۱۹	۱٤,۰۲٤	۹,۹۷۸	۸,۳۹٥
ARIMA	۷,٤۷۳	۸,٤۷۸	۸,۹۹٥	۹,٤۸٦	۹,۹٥٦	۱۰,۰۸٦	۹,۹۹۷	۱۰,۳٥۰
ARIMA EMD	٥,۱۰۸	٥,۹۲۰	٦,٦۱٥	۷,٦۰۰	۷,۹۸۳	۸,٦٤۲	۸,۳٦۰	۸,۰۸٤
Proposed	٤,٤۱۰	٦,۱٦۱	٦,٤۲۹	۷,٤۳٤	۷,۰۳۰	۷,۳٦۲	۷,٥۸٦	۸,۰۱۸

جدول ٤: مقایسه خطای پیش‌بینی مدل پیشنهادی و مدل‌های پیشین در یک تا هشت گام پیش رو



شکل ۱۱: مقادیر واقعی و مقادیر پیش‌بینی روش پیشنهادی برای یک لینک تصادفی در یک، چهار و هشت گام پیش رو

همچنین، مقادیر واقعی سرعت خودروها برای یک لینک تصادفی در مدت ۹۶ بازه زمانی در طی یک شبانه روز با مقادیر پیش‌بینی شده برای آن با روش پیشنهادی در شکل ۱۱ نشان داده شده است. این مقادیر در سه حالت الف) یک گام (۱۵ دقیقه) پیش‌رو ب) ٤ گام پیش‌رو و ج) ۸ گام پیش‌رو مورد ارزیابی قرار گرفته است. نمودارها نشان دهنده افزایش میزان خطا با افزایش تعداد گام‌های پیش‌رو است.

۷- نتیجه گیری و پیشنهادها

در این مقاله، یک روش یادگیری جمعی استکینگ جهت پیش‌بینی سرعت خودروها ارائه دادیم. یادگیری جمعی مراحل مختلفی دارد: ساخت یادگیرها، انتخاب یادگیر مناسب و تجمیع نتایج. در این مقاله ترکیبی از روش انتخاب زیرمجموعه تصادفی ویژگی‌ها برای مرحله اول (ساخت یادگیرها) و استکینگ برای مرحله سوم استفاده کردیم. با استفاده ازتجزیه مد تجربی سرعت خودروها به چند سری زمانی تجزیه و ویژگی‌های جدیدی ایجاد کردیم. همچنین، ترکیبی از رگرسیون بردار پشتیبان و رگرسیون گام‌به‌گام جهت پیش‌بینی یادگیرها بکار گرفته شده و در نهایت پیش‌بینی سرعت خودروها انجام شده است. نتایج ارزیابی نشان‌دهنده کاهش خطای پیش‌بینی سرعت در روش پیشنهادی نسبت به روش‌های رگرسیون درختی، رگرسیون بردار پشتیبان، روش جمعی بگینگ، شبکه عصبی لونبرگ، روش میانگین حرکت و روش ترکیبی ARIMA با EMD است.

۶- مراجع

[۱] Angarita-Zapata, Juan S., Antonio D. Masegosa, and Isaac Triguero. "Evaluating automated machine learning on supervised regression traffic forecasting problems." Computational Intelligence in Emerging Technologies for Engineering Applications. Springer, Cham, (۲۰۲۰). ۱۸۷-۲۰٤.

[۲] Do, Loan NN, Neda Taherifar, and Hai L. Vu. "Survey of neural network-based models for short-term traffic state prediction." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery ۹,۱ (۲۰۱۹): e۱۲۸۵.

[۳] Liu, Boyi, et al. "Singular point probability improve LSTM network performance for long-term traffic flow prediction." National Conference of Theoretical Computer Science. Springer, Singapore, ۲۰۱۷.

[٤] Qiu, Xueheng, et al. "Empirical mode decomposition based ensemble deep learning for load demand time series forecasting." Applied Soft Computing ٥٤ (۲۰۱۷): ۲٤٦-۲٥٥.

[٥] Ren, Ye, Le Zhang, and Ponnuthurai N. Suganthan. "Ensemble classification and regression-recent developments, applications and future directions." IEEE Computational intelligence magazine ۱۱,۱ (۲۰۱٦): ٤۱-٥۳.

[٦] Wolpert, David H. "Stacked generalization." Neural networks ٥,۲ (۱۹۹۲): ۲٤۱-۲٥۹.

[۷] Ho, Tin Kam. "The random subspace method for constructing decision forests." IEEE transactions on pattern analysis and machine intelligence ۲۰,۸ (۱۹۹۸): ۸۳۲-۸٤٤.

[۸] Ren, Ye, P. N. Suganthan, and Narasimalu Srikanth. "A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods." IEEE Transactions on Sustainable Energy ٦,۱ (۲۰۱٤): ۲۳٦-۲٤٤.

[۹] Xiong, Hui, Pengjian Shang, and Songhan Bian. "Detecting intrinsic dynamics of traffic flow with recurrence analysis and empirical mode decomposition." Physica A: Statistical Mechanics and its Applications ٤۷٤ (۲۰۱۷): ۷۰-۸٤.

[۱۰] Qiu, Xueheng, et al. "Empirical mode decomposition based ensemble deep learning for load demand time series forecasting." Applied Soft Computing ٥٤ (۲۰۱۷): ۲٤٦-۲٥٥.

[۱۱] Wei, Yu, and Mu-Chen Chen. "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks." Transportation Research Part C: Emerging Technologies ۲۱,۱ (۲۰۱۲): ۱٤۸-۱٦۲.

[۱۲] Philip, Anna Mary, Gitakrishnan Ramadurai, and Lelitha Vanajakshi. "Urban arterial travel time prediction using support vector regression." Transportation in Developing Economies ٤,۱ (۲۰۱۸): ۱-۸.

[۱۳] Hu, Wenbin, et al. "A short-term traffic flow forecasting method based on the hybrid PSO-SVR." Neural Processing Letters ٤۳,۱ (۲۰۱٦): ۱٥٥-۱۷۲.

[١٤] Zhang, Yanru, and Ali Haghani. "A gradient boosting method to improve travel time prediction." Transportation Research Part C: Emerging Technologies ٥٨ (٢٠١٥): ٣٠٨-٣٢٤.

[١٥] Liu, Boyi, et al. "A long short-term traffic flow prediction method optimized by cluster computing." Electrical & Electronic Engineering (٢٠١٨).

[١٦] Bai, Cong, et al. "Dynamic bus travel time prediction models on road with multiple bus routes." Computational intelligence and neuroscience ٢٠١٥ (٢٠١٥).

[١٧] Egea, Alvaro Cabrejas, Peter de Ford, and Colm Connaughton. "Estimating Baseline Travel Times for the UK Strategic Road Network." ٢٠١٨ ٢١st International Conference on Intelligent Transportation Systems (ITSC). IEEE, ٢٠١٨.

[١٨] Li, Wan, et al. "Short-term traffic state prediction from latent structures: Accuracy vs. efficiency." Transportation Research Part C: Emerging Technologies ١١١ (٢٠٢٠): ٧٢-٩٠.

[١٩] Wang, Wei, and Xucheng Li. "Travel Speed Prediction with a Hierarchical Convolutional Neural Network and Long Short-Term Memory Model Framework." arXiv preprint arXiv:١٨٠٩,٠١٨٨٧ (٢٠١٨).

[٢٠] Yue-sheng, Gu, Wei Ding, and Zhao Ming-fu. "A New Intelligent Model for Short Time Traffic Flow Prediction via EMD and PSO–SVM." Green Communications and Networks. Springer, Dordrecht, (٢٠١٢). ٥٩-٦٦.

[٢١] Xu, Jie, et al. "Mining the situation: Spatiotemporal traffic prediction with big data." IEEE Journal of Selected Topics in Signal Processing ٩,٤ (٢٠١٥): ٧٠٢-٧١٥.

[٢٢] Mendes-Moreira, João, et al. "Improving the accuracy of long-term travel time prediction using heterogeneous ensembles." Neurocomputing ١٥٠ (٢٠١٥): ٤٢٨-٤٣٩.

[٢٣] Moretti, Fabio, et al. "Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling." Neurocomputing ١٦٧ (٢٠١٥): ٣-٧.

[٢٤] Wibisono, Ari, et al. "Traffic big data prediction and visualization using fast incremental model trees-drift detection (FIMT-DD)." Knowledge-Based Systems ٩٣ (٢٠١٦): ٣٣-٤٦.

[٢٥] The Highways Agency National Transport Information System Publish Services", Technical Report, ٢٠١١.

Abstract

Forecasting the traffic speed, as one of the most important traffic parameters, is a key research topic in intelligent transportation systems. Ensemble learning methods, use several base-learners and combine their prediction results to increase the accuracy of prediction and thus reduce prediction error. In this paper, an ensemble learning method is presented based on Empirical Mode Decomposition (EMD), Support Vector Regression (SVR), and meta-learning methods. In this model, the traffic speed time series is decomposed into several intrinsic mode functions and a residual which are considered as new features of the data. In the proposed ensemble model, in order to select the most effective subset of features for each learner, the pool of learners is constructed using a combination of random subset selection methods and stepwise selection techniques. The prediction results of the base learners are used as the meta-features that are given to a meta-learner to determine the final result. To evaluate the proposed model, some traffic links of the UK Strategic Road Network (SRN) have been used. The evaluation results show a reduction in the error of the proposed method compared to baseline methods.