

امکان‌سنجی پیش‌بینی خیلی کوتاه‌مدت وضع هوا به کمک الگوریتم ژنتیک بر پایه داده‌های ایستگاه‌های همدیدی ایران

محسن مشکی^۱، پیمان کبیری^۱ و علیرضا محباالججه^۲

^۱دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران

^۲ مؤسسه ژئوفیزیک، دانشگاه تهران، تهران، ایران

چکیده

جو زمین یک سامانه آشوبناک است و پدیده‌هایی که در آن رخ می‌دهند، پیچیدگی بالایی دارند. روزانه حجم ابسوهی از داده‌های آب‌وهوایی در صدها ایستگاه همدیدی کشور ثبت می‌شوند. در این مقاله، امکان به کارگیری این داده‌ها برای پیش‌بینی کوتاه‌مدت وضع هوا مورد بررسی قرار گرفته است. چهارچوبی جهت پیش‌بینی بر مبنای این داده‌ها، پیشنهاد شده است که مبتنی بر به کارگیری مجموعه‌ای از نگاشت‌های وضعیت کنونی به وضعیت سه ساعت آینده است. در بخشی از مقاله امکان پیش‌بینی وضعیت سه ساعت آینده ایستگاه همدیدی همدان بر پایه وضعیت جاری این ایستگاه و ۹ ایستگاه پیرامونش مورد بررسی قرار گرفته است. برای نگاشت از وضعیت کنونی به وضعیت آینده، نخست بارامترهای ورودی مهم‌تر در توصیف وضعیت کنونی انتخاب و سپس از یک رگرسیون‌کننده برای نگاشت به وضعیت آینده استفاده شده است. نتایج تجربی اختلاف زیادی بین پیش‌بینی‌های انجام شده بر مبنای داده‌های ایستگاه‌های همدیدی نسبت به پیش‌بینی‌های عملیاتی نشان می‌دهد. اگر چه ضعف مدل می‌تواند یکی از دلایل این اختلاف باشد، اما به نظر می‌رسد مشکل اصلی ناشی از کمبود داده‌های ضروری، به خصوص در لایه‌های بالایی جو و تفکیک مکانی پایین داده‌ها باشد.

واژگان کلیدی: ایستگاه سینوپتیک، سامانه آشوبناک، پیش‌بینی کوتاه مدت، انتخاب ویژگی، رگرسیون، پیش‌بینی غیر عددی.

۱- مقدمه

جو یک سامانه آشوبناک است و پدیده‌هایی که در آن رخ می‌دهند پیچیدگی بالایی دارند. پیش‌بینی سامانه‌های آشوبناک با دشواری بسیاری همراه است. در سامانه‌های آشوبناک، تغییرات کوچک در اندازه‌گیری یا تخمین شرایط اولیه می‌تواند به خطای بزرگی در پیش‌بینی بینجامد (Lorenz, 1993). اندکی بالا و پایین‌شدن کمیت‌های اندازه‌گیری شده هنگام تعیین شرایط اولیه می‌تواند تفاوت چشم‌گیری در خروجی مدل به کار گرفته شده برای پیش‌بینی پدید آورد. به بیان دیگر، خطاهای کوچک در شرایط آغازین به گونه‌ای فزآینده بزرگ می‌شوند و پیش‌بینی

را تحت تأثیر قرار می‌دهند. افزون بر این، پیش‌بینی‌پذیری به خطاهای مدل نیز محدود است که به طور معمول به دلیل شبیه‌سازی تقریبی فرآیندهای جوی رخ می‌دهند (Buizza, 2002).

این دو منبع عدم قطعیت، قابلیت‌های یک سامانه منفرد را محدود می‌کنند و میزان قطعیت در پیش‌بینی را به گونه‌ای چشم‌گیر کاهش می‌دهند. این ضعف در پیش‌بینی، خود را به صورت نوسان در دقت پیش‌بینی نشان می‌دهد به شکلی که گاهی پیش‌بینی‌ها با دقت بالا و گاهی با دقت کم همراه است. شکل (۱)، سه پیش‌بینی برای شرایط آغازین بسیار نزدیک به هم را نشان می‌دهند (پارامتر ارتفاع ژئوپتانسیل ۱۰۰۰ هکتوپاسکال). تفاوت میان شرایط آغازین این سه پیش‌بینی کمتر از خطای (اندازه‌گیری شرایط

پیش‌بینی‌های کوتاه‌مدت که از چند ساعت تا ۳ الی ۵ روز را شامل می‌شود. پیش‌بینی‌های میان‌مدت از ۳ الی ۵ روز تا ۱۰ روز را شامل می‌شود. پیش‌بینی‌های بلند‌مدت نیز پیش‌بینی‌های از ۱۰ روز تا چند ماه یا گاهی ۱ الی ۲ سال را شامل می‌شود (تقربی بودن بازه‌ها بدلیل گوناگونی تعریف آنها در مراجع مختلف است) (WMO, 2010).

ایستگاه سینوپتیک، ایستگاه‌هایی هستند که به‌طور همزمان در سراسر دنیا براساس ضوابط و مقررات سازمان جهانی هواشناسی به‌طور ۲۴ ساعته موظف به اندازه‌گیری و تهییه پارامترهای جوی و ارسال آنها در شبکه مخابراتی، بهترتیب سه ساعت به سه ساعت بهنام "سینوب" و ساعت به ساعت به نام "متار" می‌باشند و بر حسب مورد استفاده انواع مختلفی دارد.

تهییه، نصب و راهاندازی انواع ایستگاه‌های سینوپتیک شامل سنجنده‌ها و دستگاه‌های ثبت داده جهت اندازه‌گیری پارامترهای سرعت و جهت باد، دما و رطوبت، میزان تابش، فشار هوای میزان بارش باران، دمای عمق خاک، دمای سطح خاک و رطوبت خاک می‌شود.

در این مقاله، امکان استفاده از داده‌های ایستگاه‌های هم‌دیدی موجود در کشور برای پیش‌بینی خیلی کوتاه‌مدت آب‌وهوا^۱ مورد بررسی قرار خواهد گرفت. به عنوان یک مطالعه موردنی، پیش‌بینی دمای سه ساعت آینده در ایستگاه همدان بر پایهٔ بیش از نود پارامتر آب‌وهوا بیان شده در ایستگاه‌های همسایه همدان انجام می‌شود. در ادامه، نکاتی درخصوص مدل‌های پیش‌بینی و انواع آن خواهد آمد. در بخش سه کارهای انجام شده مرتبط مورد بررسی قرار خواهد گرفت. سپس، در فصل چهار روش پیشنهادی معرفی خواهد شد که در آن از الگوریتم ژنتیک برای انتخاب ویژگی و از شبکه عصبی برای تخمین استفاده شده است. بخش‌های بعدی به معرفی دادگان، نتایج تجربی و کارهای آینده اختصاص دارد.

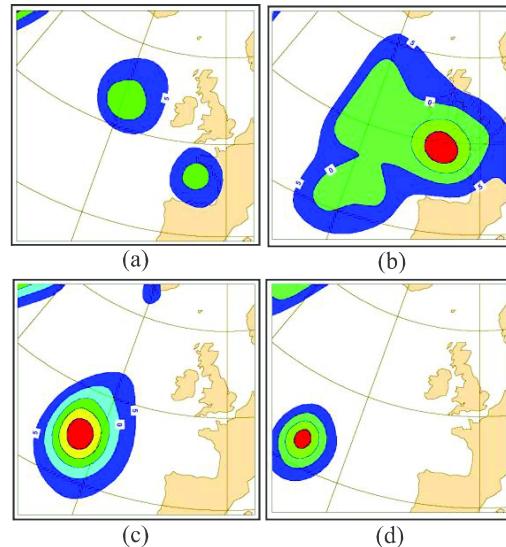
۲- پیش‌بینی وضع هوای

مدل‌های پیش‌بینی وضع هوای را می‌توان به دو دستهٔ کلی مدل‌های پیش‌بینی عددی و غیرعددی تقسیم کرد. در مدل‌های پیش‌بینی عددی، مجموعه‌ای از معادلات دیفرانسیلی دینامیک سامانه را نمایش می‌دهند و چون حل

^۱ ۱ گاهی از اصطلاح Nowcasting هم برای چنین

پیش‌بینی‌های استفاده می‌شود.

آغازین) بررسی شده تخمینی است و هر یک از آنها با احتمال یکسان می‌توانند به عنوان تخمینی از شرایط اولیه واقعی درنظر گرفته شوند. این شرایط اولیه به‌طور تقریبی یکسان توسعه یک سامانه پیش‌بینی پس از پنج روز با استفاده از یک پیش‌بینی پنج روزه به شرایط جوی بسیار ناهم‌گونی تبدیل شده‌اند. نخستین پیش‌بینی (a) دو سامانه هوایی ضعیف را نشان می‌دهد که از شرق اقیانوس اطلس به سمت غرب و جنوب بریتانیا نزدیک می‌شوند. دومین پیش‌بینی (b) سامانه‌ای قوی‌تر را نشان می‌دهد که به نواحی مرکزی‌تر نفوذ پیدا کرده است و پیش‌بینی سوم (c) نشان می‌دهد که سامانه هوایی بر روی اقیانوس اطلس باقی مانده است. از میان این سه، پیش‌بینی آخر دقیق‌تر و مشاهده‌های صورت‌گرفته پس از پنج روز (d) گواهی بر این مدعای است. مثال آورده شده نمونهٔ خوبی از شرایط اولیه نزدیکی است که با گذشت زمان نسبت به هم واگرا می‌شوند (Buizza, 2002). هم‌اکنون پیش‌بینی‌های آب‌وهوا بی (ونه اقلیمی) تا حدود دو هفته انجام می‌شود. از آنجا که این مقاله، بر پیش‌بینی خیلی کوتاه‌مدت تمرکز دارد، چندان به چالش‌های ناشی از آشوب‌ناکی جو نمی‌پردازد.



شکل ۱- a تا c پیش‌بینی‌های ۵ روزه وضع هوای برای سه شرایط اولیه تقریباً یکسان، همراه با (d) وضعیت واقعی پس از ۵ روز (Buizza, 2002)

گسترۀ وسیعی از کاربردها از مدل‌های پیش‌بینی بهره می‌برند. این کاربردها را از نظر فاصله بین زمان پیش‌بینی و رخداد مسورد نظر می‌توان به سه دسته تقسیم کرد.



می‌دهد. معادلات دیفرانسیلی چگونگی تعامل هر سلول با سلول‌های مجاور را مشخص می‌کنند.

داده‌های وضع هوا در هر سلول با استفاده از مشاهدات صورت‌گرفته در ایستگاه‌های زمینی، ماهواره‌ها و دیگر سنجنده‌های هواشناسی مشخص می‌شود. مدل پیش‌بینی با درنظر گرفتن تعامل سلول‌های شبکه با پیش‌روی زمان آنها را به روزرسانی می‌کند و بدین ترتیب وضعیت این سلول‌ها را در ساعتها و روزهای آینده پیش‌بینی می‌کنند.

کار با این حجم از داده و انجام پردازش‌های پیچیده روی آنها نیازمند بهره‌گیری از ابرایانه‌های قدرتمند است. امروزه، با وجود افزایش چشم‌گیر توان ابرایانه‌ها، پیش‌بینی قابل قبول تا کمتر از ۱۰ روز امکان‌پذیر است و پس از آن پیش‌بینی دقت کافی ندارد. جو، سامانه‌ای آشوبناک است و وجود خطاهای هر چند کوچک در داده‌ها یا مدل، با زمان رشد کرده و دقت پیش‌بینی را کاهش می‌دهد. پس از اجرای شبیه‌سازی به عنوان بخش اصلی پردازش، تکنیک‌های پس‌پردازشی روی خروجی مدل انجام می‌شود تا خطا در پیش‌بینی عددی هر چه بیش‌تر سنجیده شود و تحت کنترل در آید.

چالش‌های ساخت مدل، تنها محدود به یافتن معادلات و حل هر چه دقیق‌تر آنها نمی‌شود. مدل عمومی به‌طور معمول به پارامترهایی وابسته است که اثرگذاری بالایی روی آن دارند. برخی از آنها عبارتند از تابش خورشید، فرآیندهای مربوط به رطوبت (ابر و بارش)، مبادله گرمای، نوع خاک، پوشش گیاهی، آب سطحی و اثرات عوارض زمینی (Woods, 2006).

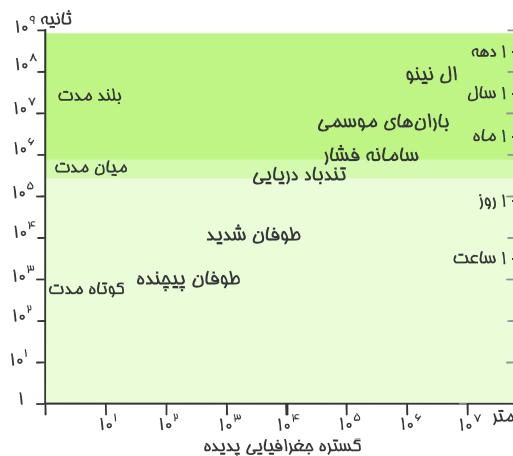
فرآیند پیش‌بینی عددی شامل چند مرحله است که در ادامه مورد بررسی قرار می‌گیرند.

۲-۱-۱- آغازگری

فرآیند وارد کردن داده‌ها به مدل و ساخت شرایط اولیه را "آغازگری" گویند. داده‌هایی که توسط سنجنده‌های مختلف ثبت شده‌اند باید به شکل مناسبی تبدیل شوند که امکان پردازش توسط مدل را داشته باشند. این سنجنده‌ها در ایستگاه‌های زمینی، روی کشتی‌های اقیانوس‌پیما، بالن‌های هواشناسی و ماهواره‌ها نصب شده‌اند. به این ترتیب، داده‌های هواشناسی و ماهواره‌ها مجموعه‌ای ناهمگن از داده‌هاست که باید به قالب مناسب درآیند. این ناهمگنی هم به نوع پارامترهای هواشناختی ثبت شده توسط سنجنده‌های مختلف و هم به

این معادلات به صورت تحلیلی میسر نیست، از روش‌های عددی برای تخمین نتایج استفاده می‌شود. می‌توان این دسته را مدل‌های "جعبه سفید"^۱ نیز نامید، چرا که چگونگی کار مدل و معادلات حاکم بر آن روش ن است. بیش‌تر سامانه‌های عملیاتی موجود از این مدل‌ها برای پیش‌بینی وضع هوا استفاده می‌کنند. دسته‌های دوم، مدل‌های غیر عددی هستند که به طور معمول با استفاده از یک ریافت مبتنی بر داده و به کمک تکنیک‌ها و ابزارهای یادگیری ماشین ساخته می‌شوند. این دسته از مدل‌ها بیش‌تر برای پیش‌بینی‌های میان‌مدت و بلند‌مدت استفاده می‌شوند (WMO, 2010).

شکل دو، فاصله زمانی و گستره جغرافیای چند پدیده جوی را نشان می‌دهد. با افزایش فاصله زمانی بین پیش‌بینی و رخداد مورد نظر، عدد قطعیت پیش‌بینی افزایش می‌یابد، از این رو، پژوهش جاری به پیش‌بینی کوتاه‌مدت خواهد پرداخت.



(شکل ۲)- فاصله زمانی پیش‌بینی و گستره جغرافیایی چند پدیده جوی (WMO, 2010)

۲-۲- مدل‌های عددی

مدل‌های پیش‌بینی عددی، مجموعه‌ای از معادلات دیفرانسیلی را استفاده می‌کنند که بر پایه قوانین فیزیک، حرکت شاره‌ها و شیمی تنتظیم شده‌اند. این معادلات روی یک دستگاه مختصات اعمال می‌شوند که ناحیه مورد مطالعه را به یک شبکه سه‌بعدی تقسیم می‌کند. دو بعد افقی، مختصات جغرافیایی و بعد قائم، لایه‌های جوی را مشخص می‌کنند. هر سلول از این شبکه، مجموعه‌ای از پارامترها مانند سرعت و جهت باد، دما، فشار، رطوبت و ... را نشان

^۱ White Box

یا مبتنی بر داده قرار می‌گیرند که مدل را بر پایه حجم انبوهی از داده‌ها می‌سازند. مدل‌های غیر عددی به طور معمول به کمک روش‌ها و ابزارهای یادگیری ماشین ساخته می‌شوند. این مدل‌ها زمانی مفیداند که مدل‌های تحلیلی، موجود یا به اندازه کافی به سامانه واقعی نزدیک نباشند. این دسته از مدل‌ها بیشتر برای پیش‌بینی‌های میان‌مدت و بلندمدت استفاده می‌شوند (Nandar, 2009).

۳- کارهای انجام شده

گوارد داده‌ها، یکی از گام‌های نخستین در پیش‌بینی عددی وضع هواست. ناون (2009)، پس از بررسی تاریخی گوارد داده، یک بررسی جامع روی روش‌های گوارد داده انجام داده است. د-رزنی (de Rosnay, 2008)، روشی برای ایجاد داده‌های گواردشده مربوط به رطوبت خاک به کمک سنجش از دور پیشنهاد کرده است. او در کارش، از برخی نواحی در تصاویر ماهواره SMOS برای تشخیص میزان رطوبت خاک استفاده نموده است. مرکز اروپایی پیش‌بینی میان‌مدت آب‌وهوا (ECMWF)، علاوه‌بر اینکه توسعه‌دهنده یکی از مهم‌ترین سامانه‌های عملیاتی موجود است، به لحاظ تکنیکی هم تأثیر قابل توجهی بر سامانه‌های دیگر گذاشته است (Eliassen, 2005; Persson & Grazzini, 2005). در این کار پژوهشی-عملیاتی، بر دو ایده پیش‌بینی ترکیبی^۲ و مشاهدات هدفمند تأکید شده است. پیش‌بینی ترکیبی روشنی برای مجتمع سازی مجموعه‌ای از پیش‌بینی‌های منفرد و قطعی است که هر یک دارای تابع توزیع احتمال برای فضای حالت پیش‌بینی است. همه پیش‌بینی‌های منفرد با مدل یکسان تولید می‌شوند، اما به‌ازای حالت آغازین متفاوت. ساختار همادی می‌تواند قدرت پیش‌گویی خود را تخمین بزند. به این ترتیب که اگر نتیجه پیش‌بینی‌های منفرد تفاوت چشم‌گیر داشته باشند، قطعیت در پیش‌بینی کم است و بر عکس. ایده اصلی در مشاهده‌های هدفمند، آن است که در نواحی ناپایدار روش بردار تکینه^۳ به کار گرفته شود. می‌توان با شناسایی این نقاط، مشاهده‌ها را در این نقاط بیشتر و دقیق‌تر کرد. هاگل (Hagel, 2009)، پیاده‌سازی یک سامانه پیش‌بینی عددی وضع هوای منطقه‌ای کوتاه‌مدت را برای مجارستان گزارش کرده است. در این کار، روشی مشابه با ECMWF پیشنهاد شده و

پراکندگی ناهمگن مکانی آنها باز می‌گردد. فرآیند آماده‌سازی داده‌ها برای استفاده در یک مدل پیش‌بینی را "گوارد داده"^۴ گویند.

۲-۱-۲- پردازش

یک مدل جوی، برنامه‌ای رایانه‌ای است که اطلاعات هواشناسی را برای زمان‌های آینده تولید می‌کند. مدل‌های نوینی که امروزه به کار می‌روند، مجموعه رابطه‌های ریاضی به نام معادلات بسیط را حل می‌کنند. این معادلات برگرفته از قوانین حاکم بر گازهای ایده‌آل است و به کمک آنها می‌توان چگالی، فشار، دما، سرعت جابه‌جایی هوا (باد) و برخی پارامترهای دیگر را پیش‌بینی کرد (Pielke, 2002). می‌توان به این مجموعه، قوانین دیگری را افزود که در کاربردهای خاصی ضروری هستند. برای نمونه، جهت پیش‌بینی آسودگی هوا، قوانین حاکم بر ذرات معلق به مجموعه معادلات بسیط افزوده می‌شود تا مدل جزئیات مربوط به این ذرات را در نظر گیرد. این معادلات، معادلات دیفرانسیل پارهای غیرخطی هستند که حل دقیق آنها با روش‌های تحلیلی غیرممکن است؛ مگر در برخی مدل‌های ساده شده و خاص. حل ناپذیری این معادلات، به کارگیری روش‌های عددی را تبدیل به یک ضرورت کرده است.

۲-۱-۳- پارامترسازی

برخی فرآیندهای جوی بسیار کوچک‌مقیاس یا پیچیده‌تر از آن هستند که در مدل‌های عددی پیش‌بینی وضع هوا در نظر گرفته شوند. پارامتردهی، روالی است برای نمایش و جای دادن چنین فرآیندهایی در قالب متغیرهایی که در مدل استفاده می‌شوند. به عنوان نمونه، فاصله نقاط روی شبکه در مدل‌های وضع هوا و اقلیمی بین پنج تا سیصد کیلومتر تغییر می‌کند. از سوی دیگر، یک ابر کومه‌ای نوعی به مقیاسی کوچک‌تر از یک کیلومتر نیاز دارد تا بتوان آن را با دقیق مناسب به کمک روابط حرکت شاره‌ها مدل کرد (Woods, 2006).

۲-۲- مدل‌های غیر عددی

مدل‌های عددی، زیرمجموعه‌ای از مدل‌های تحلیل هستند که بر پایه روابط شناخته شده در سامانه واقعی کار می‌کنند (Woods, 2006). در مقابل رویکرد تحلیلی که بر استخراج روابط یک سامانه واقعی استوار است؛ رویکردهای غیر عددی

² European Centre For Medium-Range Weather Forecasts (Ecmwf)

³ Ensemble Forecasting

⁴ Singular Vectors

^۱ Data Assimilation



شبکه‌ای دو بعدی شامل ۲۴۹۱ گره ایجاد کرده‌اند. همچنین، کارهای مشابه زیادی با بهره‌گیری از مدل‌های یادگیر و ابزارهای بهینه‌سازی گوناگون برای نواحی مختلف انجام شده است. هانگ و همکاران (2009)، از یک شبکه عصبی ساده برای پیش‌بینی بارندگی در بانکوگ پایتخت تایلند استفاده کرده‌اند. وو و ژو (Wu & Zhou, 2009) از تخمین زنده ماشین بردار پشتیبان^۲ و روش‌های بهینه‌سازی مبتنی بر هوش جمعی برای ایجاد و بهینه‌سازی یک مدل پیش‌بینی بارندگی استفاده کرده‌اند. لونام (Luenam et al., 2010)، از یک شبکه عصبی-فازی برای پیش‌بینی بارندگی در تایلند استفاده کرده است.

مطابق بررسی‌های صوت‌گرفته و همچنین نظر کارشناسان خبره در حوزه هواشناسی، تا زمان تهیه این مقاله هیچ مدل بومی جامع برای پیش‌بینی وضع هوای کشور ساخته نشده است. نیازهای کنونی کشور با بهره‌گیری از سامانه‌های پیش‌بینی عمومی و منطقه‌ای غیر بومی برآورده می‌شود. با این حال، تلاش‌های پراکنده‌ای برای تخمین یا پیش‌بینی محدود برخی پارامترهای آب‌وهوا بیانی-مانند بارندگی صورت گرفته است. جهانبخش و ترابی (جهانبخش و ترابی، ۱۳۸۳)، ایران را بر مبنای داده‌های ۴۱ ایستگاه هم‌دیدی به پنج منطقه اقلیمی تقسیم کرده و برای هر یک از این نواحی پنج گانه، یک ایستگاه شاخص تعیین کرده‌اند. سپس، هر یک از این نواحی را به لحاظ بارندگی ماهانه در یک بازه زمانی سی ساله مورد بررسی قرار دادند. نتایج آنها تغییرات معنی‌داری در دما و بارندگی نشان نمی‌دهند.

دسته‌ای از کارها به بررسی پارامترهای مهم در پیش‌بینی بارندگی اختصاص دارد. کارآموز (کارآموز، ۱۳۸۵)، اثرگذاری ارتفاع ژئوپتانسیل سطح پانصد هکتوپاسکال را بر بازش در غرب ایران مورد بررسی قرار داده است. ایشان با بهره‌گیری از روش تحلیل عامل فضای ارتفاع ژئوپتانسیلی سطح پانصد هکتوپاسکال را طبقه‌بندی کرده است. نتایج کار، رضایت‌بخش بودن این طبقه‌بندی را نشان می‌دهند.

۴- روش پیشنهادی

وضع هوای هر نقطه جغرافیایی در آینده نزدیک، واسته به وضعیت آن نقطه و نقاط پیرامون آن است. شکل (۳)، این موضوع را به خوبی نشان می‌دهد.

سپس، تکنیک‌هایی برای بهبود نتایج سامانه‌های پیش‌بینی عمومی با استفاده از سامانه جدید پیشنهاد شده است. مونتوپلی و مارزانو (Montopoli & Marzano, 2009)، روشی برای ریزمقیاس‌نمایی خروجی مدل‌های پیش‌بینی پیشنهاد کرده‌اند. در این کار، خروجی مدل MM5 که دارای مقیاس بزرگ است، برای ایجاد نقشه‌های بارندگی با استفاده از یک روش آماری ریزمقیاس شده است.

در حوزه مدل‌سازی غیر عددی کارهای کمتری نسبت به حوزه مدل‌سازی عددی انجام شده است. استفاده از مدل‌های پیش‌بینی غیر عددی بیشتر در پیش‌بینی‌های میان‌مدت گزارش شده است. با این وجود پژوهش‌هایی وجود دارند که در آنها پیش‌بینی کوتاه‌مدت مورد نظر بوده است. آنسل (Ancell, 2009) روشی برای پیش‌بینی کوتاه‌مدت بارندگی پیشنهاد کرده است. از آنجا که باران یک پدیده ریزمقیاس است و مدل‌های پیش‌بینی عددی از مقیاس‌های بزرگ استفاده می‌کنند، نمی‌توان تنها بر حسب خروجی مدل‌های عددی توزیع بارندگی را با دقت مناسب مشخص کرد (Ancell, 2009). او از شبکه بیز برای یافتن ارتباط زمانی-مکانی بین بارندگی در یک‌صدا ایستگاه هواشناسی در شمال غربی اسپانیا استفاده کرد؛ تا روابط علت و معلولی بین آنها را شناسایی و بازنمایی کند. او شبکه بیز ساخته شده را با یک مدل پیش‌بینی عددی ادغام و امکان پیش‌بینی بارش بر مبنای خروجی مدل عددی را در ایستگاه‌های مورد نظر فراهم کرد. این کار، نوعی پس‌پردازش غیر عددی روی خروجی مدل عددی محسوب می‌شود.

بخش عمده‌ای از کارهای انجام شده در حوزه مدل‌سازی غیر عددی به پیش‌بینی میان‌مدت و به خصوص ماهانه اختصاص دارد. نادر (Nadar, 2009)، از یک مدل احتمالی مبتنی بر شبکه بیز برای پیش‌بینی بارندگی ماهانه در میانمار استفاده کرده است. شبکه بیز احتمال‌های شرطی یا روابط علت و معمولی متغیرهای آب‌وهوا بیانی را بازنمایی می‌کند. در این کار از داده‌های ثبت شده در ایستگاه‌های هواشناسی میانمار و برخی داده‌های عمومی مانند دمای سطح دریا استفاده شده است. کی و همکاران (Qi et al., 2009)، روشی ریاضی برای ایجاد یک مدل بهینه برای پیش‌بینی بارندگی در چین پیشنهاد کرده‌اند. آنها پیش از ایجاد مدل، نخست داده‌های ۹۱ ایستگاه هواشناسی را با استفاده از روش "درونسایی دومکعبی"^۱ گوارد نموده و

² Support Vector Regressor

سال ۱۳۹۲ شماره ۲ پیاپی

¹ Bicubic Interpolation Method

به بیان دیگر:

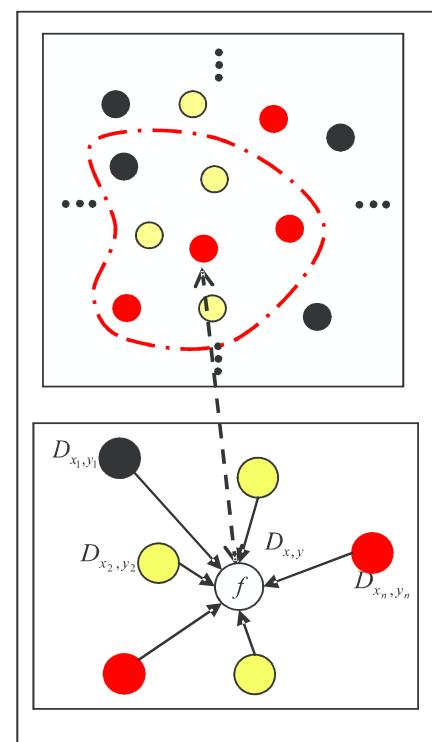
$$D_{x,y}^{t+\Delta t} = f(D_{x,y}^t, D_{x_1,y_1}^t, D_{x_2,y_2}^t, \dots, D_{x_n,y_n}^t)$$

به ازای هر نقطه داده‌ای مانند y تابعی به صورت $f(x,y)$ وجود دارد که مشخص می‌کند وضعیت نقطه (x,y) در لحظه $t+\Delta t$ چگونه بر مبنای وضعیت آن نقطه و نقاط مجاور در لحظه t به دست می‌آید. به عبارت دیگر، وضعیت نقطه (x,y) در آینده، تابعی از وضعیت این نقطه و نقاط پیرامونی در لحظه کنونی است. هدف ما تخمین هر چه دقیق‌تر این تابع $(f(x,y))$ است. داده‌های هوشمناسی ثبت شده، اطلاعات ارزشمندی در مورد این تابع به ما می‌دهند. این اطلاعات شامل تعدادی جفت‌های ورودی-خروجی است، که داده‌های لازم برای آموزش مدل‌های یادگیر را فراهم می‌کند. شکل (۳)، فرآیند ایجاد مجموعه داده برای نقاط مختلف را نشان می‌دهد.

زمان‌های متوالی با فواصل زمانی یکسان است (وضعیت هوا در لحظه‌های $Wt+s\Delta t$, $Wt+\Delta t$, Wt , ...). ساختن مدل، با استفاده از تکنیک‌های یادگیری ماشین مستلزم آن است که مجموعه داده‌ای برای هر نقطه ساخته شود که شامل جفت‌هایی به صورت وضعیت فعلی-وضعیت آینده باشد (که نحوده تغییر وضعیت را بازنمایی کنند). برای یک نقطه مانند P_i مجموعه داده ایجاد شده DS_i نامیده می‌شود. برای نمونه در شکل (۴)، DS_i مجموعه داده‌ای است که برای نقطه داده‌ای p_i با مختصات (x,y) ساخته شده است. هر رکورد در این مجموعه داده، یک دوتایی است که بردار اول شامل پارامترهایی است که وضعیت جاری نقطه مورد نظر و پیرامون آن را مشخص می‌کند و بردار دوم شامل پارامترهایی است که وضعیت آینده آن نقطه را بازنمایی می‌کنند.

بررسی داده‌های در دسترس نشان داد، بیشترین تراکم و پوشش ایستگاه‌های همدیدی پیرامون شهر همدان وجود دارد. به همین دلیل، فرآیند ایجاد مجموعه داده برای این نقطه داده‌ای (ایستگاه همدان) به کار گرفته شد. به این ترتیب، مجموعه داده‌ای ساخته شد که امکان پیش‌بینی وضعیت آب‌وهای همدان را بر پایه تعدادی از نقاط پیرامون آن فراهم می‌کند. این نقاط در شکل (۵) مشخص شده‌اند. در این شکل، همدان با یک دایره قرمزرنگ از دیگر نقاط نقشه تمایز شده است. همچنین، ۹ ایستگاه همدیدی پیرامون آن با دایره‌های آبی‌رنگ مشخص شده‌اند.

پس از تبدیل داده‌ها به قالب مورد نظر، یک مجموعه داده حاصل می‌شود که بازنمایی کننده یک نگاشت از وضعیت حال ایستگاه‌های مؤثر به وضعیت سه ساعت آینده ایستگاه هدف است. وضعیت هر ایستگاه با چند پارامتر ثبت شده در آن بازنمایی می‌شود. اگر تعداد ایستگاه‌ها را در تعداد پارامترهای ثبت شده در هر ایستگاه ضرب کنیم، تعداد پارامترهایی که وضعیت جاری را توصیف می‌کنند به دست می‌آید. این عدد قابل توجه است و باید آن را کاهش داد. به بیان دیگر، باید از بسیاری از پارامترها که در پیش‌بینی مؤثر نیستند یا با دیگر پارامترها همبستگی دارند، چشم پوشی کرد. این کار با به کارگیری یک روش انتخاب ویژگی قابل انجام است.



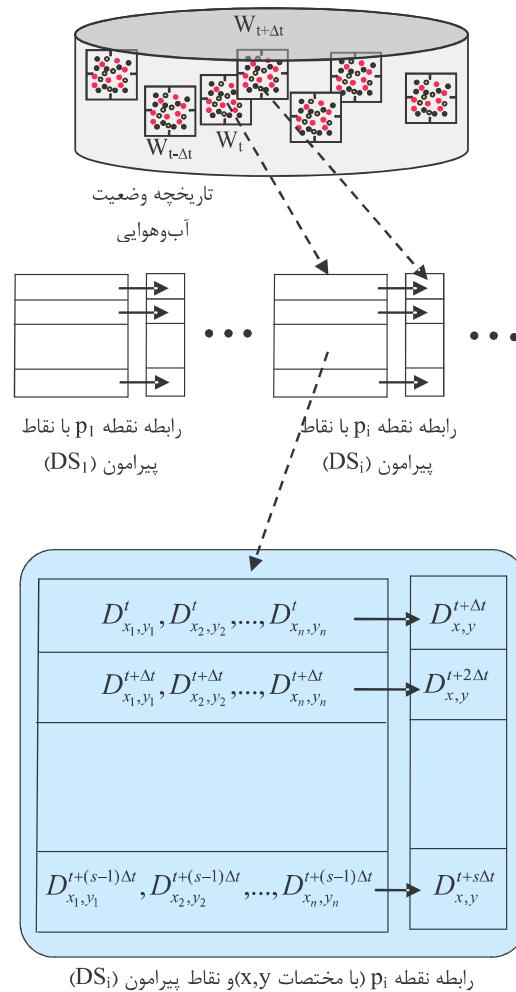
(شکل ۳)- پیش‌بینی وضعیت یک نقطه بر مبنای نقاط پیرامون در روش پیشنهادی

نخستین گام، جمع‌آوری تاریخچه وضع هوایی ناحیه مورد نظر است. این تاریخچه، شامل وضعیت هوایی ناحیه در

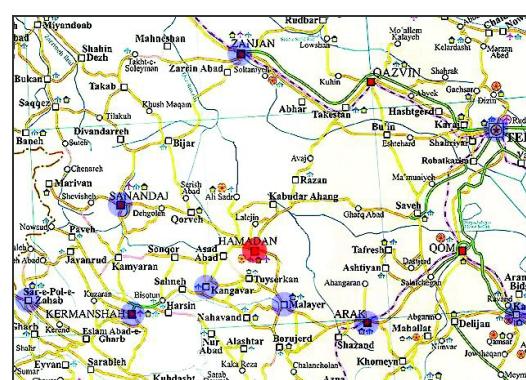
بماند. در بسیاری از مسائل واقعی، انتخاب ویژگی ضرورت دارد؛ چرا که به طور معمول برخی از ویژگی‌ها نویه‌ای^۱، نامربوط یا حتی گمراه‌کننده هستند. این ویژگی‌ها، علاوه‌بر آن که کمکی به توصیف بهتر الگوها نمی‌کنند، گاهی موجب کاهش دقت شناسایی و تخمین هم می‌شوند. بسیاری از مسائل دنیای واقعی در حوزه بیوانفورماتیک (Armananzas et al. 2011; García-Torres et al. 2013 Azam and Yao 2012; Feng et al. 2012; Imani et al. 2013 Ghosh et al. 2013; Rashedi et al. 2013; Vignolo et al. 2013 Zhang et al. 2011; Waad et al. 2013; Martis et al. 2009) بدون انتخاب ویژگی ناممکن یا دشوار هستند.

ارزش یک ویژگی یا یک مجموعه ویژگی بر پایه همبستگی^۲ و افروزنگی^۳ آن با دیگر ویژگی‌ها مشخص می‌شود. یک ویژگی همبسته یا مربوط نامیده می‌شود، اگر به کمک آن بتوان در مورد ویژگی‌های تصمیم^۴ پیش‌گویی کرد. همچنین، یک ویژگی افزونه نامیده می‌شود؛ اگر با ویژگی‌های شرطی^۵ دیگر همبستگی بالا داشته باشد؛ به نحوی که بتوان آن را با استفاده از برخی ویژگی‌های شرطی دیگر با دقت بالا پیش‌بینی کرد. به این ترتیب، می‌توان انتخاب ویژگی را جستجو به دنبال زیرمجموعه‌ای از ویژگی‌های شرطی دانست که کمترین همبستگی را با هم دارند و بیشترین همبستگی را با ویژگی‌های تصمیم دارند.

اگر در یک مجموعه داده n ویژگی وجود داشته باشد، 2^n زیرمجموعه نامزد وجود دارند که می‌توانند به عنوان زیرمجموعه برگزیده انتخاب شوند. برای یافتن تضمینی جواب بهینه باید جستجوی کامل انجام شود و همه زیرمجموعه‌های ممکن مورد بررسی قرار گیرند. از آنجا که در عمل چنین کاری میسر نیست؛ به طور معمول از روش‌های ابتکاری و جستجوی تصادفی استفاده می‌شود. این روش‌ها ادعایی در مورد یافتن روش بهینه ندارند، با این حال، به طور معمول پاسخ‌هایی ارائه می‌دهند که از کیفیت خوبی برخوردار هستند.



(شکل ۴) – فرآیند ایجاد مجموعه داده برای هر نقطه داده‌ای



(شکل ۵) – همدان و ایستگاه‌های همدیدی پیرامون آن

¹ Noisy
² Relevancy
³ Redundancy
⁴ Decision Features
⁵ Conditional Features

۱-۴- انتخاب ویژگی

هدف از انتخاب ویژگی، یافتن زیرمجموعه‌ای از ویژگی‌ها از حوزه مسئله است به‌گونه‌ای که دقت در حد قابل قبولی باقی

۴-۲-الگوریتم ژنتیک

یکی از الگوریتم‌های بهینه‌سازی پرکاربرد، الگوریتم ژنتیک است. مسئله انتخاب ویژگی را می‌توان به صورت یک مسئله بهینه‌سازی تعریف کرد. به بیان دقیق‌تر، هدف از انتخاب ویژگی، یافتن زیرمجموعه‌ای از ویژگی‌های شرطی است که بیشترین همبستگی را با ویژگی تصمیم و کمترین همبستگی را با یکدیگر داشته باشدند.

مهم‌ترین بخش از فرآیند حل یک مسئله با الگوریتم ژنتیک، تعیین شیوه بازنمایی^۱ یا رمزگذاری وتابع ارزیابی^۲ است. برای بازنمایی، می‌توان هر کروموزم (که معادل با یک راه حل نامزد است) را به صورت رشته‌ای از بیت‌ها در نظر گرفت. در این حالت، هر بیت تعلق یا عدم تعلق ویژگی شرطی متناظر به زیرمجموعه را نشان می‌دهد. تابع ارزیابی نیز می‌تواند دقت رگرسیون یا رده‌بند با کروموزم باشد. میزان دقت، موجود در زیرمجموعه متناظر با کروموزم باشد. میزان دقت، نماینده میزان همبستگی ویژگی‌های شرطی با ویژگی تصمیم است، به طوری که، هر چه همبستگی بین آنها بیشتر باشد، دقت بالاتر و هر چه همبستگی کمتر باشد، دقت پایین‌تر خواهد بود.

۴-۳-روش RELIEF

روش RELIEF، یک روش انتخاب ویژگی مبتنی بر نمونه است که برخلاف الگوریتم ژنتیک نیازی به محاسبه دقت برای هر زیرمجموعه از ویژگی‌ها ندارد. شالوده این روش، وزن دهی به ویژگی‌ها با توجه به توانایی آنها در جداسازی تبقه‌ها از هم است. شکل (۶)، الگوریتم این روش را نشان می‌دهد.

```

 $R = \{\}$ 
for  $i=1$  to  $n$ 
   $W_a = 0$ 
for  $i=1$  to  $its$ 
  choose an object in  $O$  randomly
  calculate  $x$ 's nearHit and nearMiss
  for  $j=1$  to  $n$ 
     $W_j = W_j - \text{diff}(x_j, \text{nearHit}_j)/its$ 
  +
     $diff(x_j, \text{nearMiss}_j)/its$ 
  Add the most  $k$  important features to  $R$ 

```

شکل (۶)-الگوریتم RELIEF

فصل نهم

^۱ Representation (Coding)

^۲ Fitness Function



در این الگوریتم، در یک فرآیند تکرار هر بار یک نمونه مانند x به تصادف انتخاب می‌شود و به دنبال آن $nearHit$ یا نزدیک‌ترین نمونه هم‌طبقه با x و $nearMiss$ یا نزدیک‌ترین نمونه غیر هم‌طبقه با آن مشخص می‌شوند. سپس وزن هر ویژگی بر مبنای این مقدار آن در سه نمونه x ، $nearHit$ و $nearMiss$ تغییر می‌کند. اگر ویژگی در دو نمونه x و $nearMiss$ متفاوت باشد، وزن ویژگی کاهش می‌یابد، چون $nearHit$ مطلوب نیست. به عکس، اگر ویژگی در دو نمونه x و $nearMiss$ متفاوت باشد، وزن ویژگی افزایش یافته چون به درستی این دو نمونه را از هم متمایز کرده است.

۵-مجموعه دادگان

داده‌ها شامل برخی پارامترهای ثبت‌شده بین سال‌های ۲۰۰۰ تا ۲۰۰۷ میلادی در تعدادی از ایستگاه‌های هم‌دیدی پیرامون شهر همدان است، که در بخش قبلی به آنها اشاره شد. در هر یک از ایستگاه‌ها، نزدیک به ۱۸ پارامتر هر ۳ یا ۶ ساعت یکبار ثبت می‌شوند. از این میان، برخی پارامترها در برخی ساعتها ثبت نشده‌اند و به اصلاح دارای مقدارهای از دست رفته^۳ هستند. مجموع تعداد پارامترها در ایستگاه همدان و ۹ ایستگاه پیرامونش ۱۸۱ پارامتر است که از میان آنها تنها ۹۲ پارامتر فاقد مقدارهای از دست رفته هستند و در فواصل زمانی سه ساعت بدون استثنا ثبت شده‌اند. از این میان، ۱۰ پارامتر مربوط به ایستگاه همدان و ۸۲ پارامتر مربوط به ۹ ایستگاه همسایه است. برای نمونه، پارامترهای پارامتر مربوط به همدان عبارتند از: دو پارامتر مربوط به فشار، دو پارامتر مربوط به رطوبت، دو پارامتر مربوط به دما، جهت و سرعت باد، دید عمودی و میزان ابری بودن. از آنجا که قرار است وضعیت سه ساعت آینده در همدان، بر مبنای وضعیت کنونی همدان و ۹ ایستگاه پیرامونش پیش‌بینی شود، باید تابعی از یک فضای ۹۲ بعدی (مجموع پارامترها در ۱۰ ایستگاه که توصیف کننده شرایط کنونی است) به یک فضای ۱۰ بعدی (پارامترها در ایستگاه همدان برای سه ساعت آینده) تخمین زده شود. تعداد رکوردهایی که از این تابع در دسترسی است، ۲۳۳۷۶ رکورد است (۸ سال شامل ۲۹۲۲ روز است و در هر روز هشت‌تبار

^۳ Missing Value



شرایط اجراهای مختلف در جدول (۱) آمده است.
خطای محاسبه شده در این بخش و بخش شش با استفاده از رابطه زیر محاسبه شده است:

$$E = \sum_{o=1}^{10} \frac{\sum_{i=1}^n |y_i^o - \hat{y}_i^o|}{n} \quad (1)$$

در این فرمول، y_{i0} خروجی ۵ام تخمین زننده برای الگوی آن است. در اینجا خروجی واقعی و خروجی تخمینی هر دو نرمال هستند.

(جدول ۱)- شرایط الگوریتم ژنتیک در پنج اجرا

اجرا	وزن تعیین کننده	وزن نگذاری	شبکه عصبی برای تعیین دقت تکرار آموزش	لایه محفوی		نکته
				لایه	محفوی	
۱	۱۰۰	۱۰۰	۷۰	۱۰	۴۶	۳/۷۱
۲	۱۰۰	۵۰	۴۰	۱۰	۴۱	۳/۸۴
۳	۱۰۰	۲۰۰	۴۰	۱۰	۴۱	۳/۷۱
۴	۷۰	۱۰۰	۵۰	۱۰	۳۹	۳/۷۲
۵	۷۰	۱۰۰	۷۵	۱۰	۴۷	۳/۷۰

RELIEF-۲-۶

نسخه اولیه الگوریتم برای مسائل رده‌بندی ساده پیشنهاد شده است و برای استفاده از آن در مسائل رگرسیون باید تغییراتی در آن صورت می‌گرفت. مهم‌ترین تغییر مربوط به شیوه تعیین nearHit و nearMiss است. پس از انتخاب تصادفی یک نمونه مانند X، نزدیک‌ترین نمونه در فضای ویژگی‌های شرطی که جزو p درصد (در پیاده‌سازی ۵۵درصد) نمونه‌ای است که کمترین فاصله را با X در فضای ویژگی‌های تصمیم دارد، به عنوان nearHit در نظر گرفته می‌شود.

برای انتخاب ویژگی‌ها پس از وزن دهی به آنها، کافی است وزن‌ها با یک مقدار آستانه مقایسه و ویژگی‌هایی که وزنی بیشتر از آستانه دارند، انتخاب شوند. در هر یک از دو اجرا، آستانه چنان انتخاب شده است که تعداد ویژگی‌های انتخاب شده برابر با تعداد دلخواه (۵۵ در اجرای اول و ۵۰ در اجرای دوم) شود. برای نمونه در اجرای نخست، این مقدار ۰/۰۲۹ در نظر گرفته شده است که عددی بین وزن ۵۵امین و ۶۰امین ویژگی برتر است (مراجعه شود به جدول‌های ۱ تا ۴).

پارامترها ثابت می‌شوند، بنابراین داریم $23376 = 2922 \times 8$.

۶- نتایج تجربی

برای سنجش میزان مفیدبودن داده‌های جمع‌آوری شده از ایستگاه‌های همدیدی، ابتدا داده‌ها به قالب شرح داده شده در بخش چهار در آمد. در گام بعدی از دو روش انتخاب ویژگی الگوریتم ژنتیک و الگوریتم RELIEF استفاده شد. پس از انتخاب ویژگی، شبکه عصبی پرسپترون چندلایه برای نگاشت وضعیت کنونی به وضعیت آینده به کار گرفته شد. در همه آزمایش‌ها شبکه عصبی با یک لایه مخفی با تابع فعالیت تائزانت سیگموید برای لایه پنهان و تابع خطی برای لایه خروجی استفاده شده است. در شبکه‌های عصبی به کار گرفته برای ارزیابی زیرمجموعه‌ها در الگوریتم ژنتیک، ۱۰ نرون در لایه پنهان استفاده شد و برای ارزیابی نهایی زیرمجموعه انتخابی توسط الگوریتم انتخاب ویژگی با سعی و خطا تعداد بهینه نرون‌ها تعیین شده است. همچنین مجموعه داده، به سه بخش شصت‌درصد برای آموزش، بیست‌درصد برای اعتبارسنجی و بیست‌درصد برای آزمون تقسیم شد.

۶-۱- الگوریتم ژنتیک

یافتن پاسخ بهینه از میان ۲۹۲ زیرمجموعه‌های ممکن برای یک مجموعه ۹۲ تایی از ویژگی‌ها نیازمند بررسی همه آنهاست که این کار در عمل میسر نیست. به عنوان یک الگوریتم بهینه‌سازی مؤثر، می‌توان از الگوریتم ژنتیک برای یافتن زیرمجموعه بهینه یا نزدیک بهره گرفت. در این حالت، بازنمایی مسئله به صورت یک رشته بیتی ۹۲ تایی است که هر بیت تعیین می‌کند که ویژگی متناظر با آن به کار گرفته شود یا خیر. تابع هدفی تعریف شده برای مسئله، میزان خطای شبکه عصبی پس از آموزش است که الگوریتم ژنتیک سعی می‌کند آن را کمینه کند.

چهار جدول نخست پیوست، این مسئله را به خوبی نشان می‌دهند. برای نمونه در جدول (۸)، درایه سطر GA ۱ و ستون GMT DD مقدار صفر دارد که به این معنی است که پاسخ اجرای اول الگوریتم ژنتیک نشان می‌دهد نیازی به استفاده از پارامتر جهت باد (GMT DD) نیست.

(جدول ۴)-پارامترهای توصیف کننده وضع هوای

نام کوتاه	پارامتر
GMT DD	Wind Direction
GMT FF	Wind Speed
GMT WET	Wet Bulb Temperature
GMT DRY	Dry Bulb Temperature
GMT DEW	Hourly Relative Humidity
GMT HUM	Hourly Relative Humidity
GMT QFE	Hourly Pressure QFE
GMT QFF	Hourly Pressure QFF
GMT N	Hourly Amount of Cloud
GMT WW	Present Weather
GMT VV	Vertical Visibility
GMT VP	Vapor Pressure

جدول (۵) اطلاعات مربوط به خطای هر یک از خروجی‌های تخمین‌زننده هنگامی که از ویژگی‌های انتخاب شده در اجرای نخست الگوریتم ژنتیک استفاده می‌شود، مشخص می‌کند (هر خروجی مربوط به یکی از ۱۰ پارامتر وضع هوای شهر همدان است).

همچنین جدول (۶) اطلاعات مربوط به خطای هر یک از خروجی‌های تخمین‌زننده را هنگامی که از ویژگی‌های انتخاب شده در اجرای نخست الگوریتم ژنتیک استفاده می‌شود، مشخص می‌کند.

جدول (۷) اطلاعات مربوط به خطای هر یک از خروجی‌های تخمین‌زننده را هنگامی که از ویژگی‌های انتخاب شده در اجرای نخست الگوریتم RELIEF استفاده می‌شود، را مشخص می‌کند (هر خروجی مربوط به یکی از ۱۰ پارامتر آب و هوایی شهر همدان است).

برخی اطلاعات آماری مربوط به خروجی‌ها در جدول (۸) آمده است. تخمین‌زننده ۱۰ خروجی دارد که برای هر یک از آنها میانگین، انحراف معیار، مقدار کمینه و مقدار بیشینه آن خروجی در جدول گزارش شده است.

بررسی خطای نسبی نشان می‌دهد که پیش‌بینی برخی پارامترها به خوبی انجام شده است. برای نمونه، پارامترهای دید عمودی، رطوبت و بهخصوص فشار با دقیق خوبی تخمین زده شده‌اند. از سوی دیگر، برخی پارامترها، خطای نسبی بسیار بالایی دارند، اما به نظر می‌رسد دقیق پیش‌بینی مقدار مطلق آنها چندان بدنباشد. برای نمونه، پارامتر دما (GMT WET) یا نقطه شبنم (GMT DEW) در برخی آزمایش‌ها، خطای نسبی بسیار بالایی دارند (۴/۰۳) خطای نسبی نقطه شبنم برای GA1 و خطای نسبی ۵/۲۳ برای دما در اجرای RELIEF1 در حالی که مقایسه بازه تغییرات پارامتر و خطای مطلق بیان‌گر یک خطای معمولی است. به عنوان مثال دما در

پیوست). جدول (۲)، نتایج حاصل از این دو اجرا را نشان می‌دهد.

(جدول ۲)-نتایج ۲ اجرای الگوریتم RELIEF

اجرا	تعداد ویژگی‌ها	خطای پاسخ
۱	۵۵	۲/۷۲
۲	۵۰	۳/۷۴

۶-۳-تحلیل نتایج

نتایج اجرای دو الگوریتم نزدیک به هم است با این حال، اجراهای الگوریتم ژنتیک (به جز اجرای دوم) نتایج بهتری را به جای گذاشته است، زیرا میزان خطای در آنها کم و بیش با اجراهای الگوریتم RELIEF یکسان است اما تعداد ویژگی‌های کمتری را انتخاب کرده‌اند.

در جدول (۳) شباهت بین نتایج اجراهای هفت‌گانه دو الگوریتم به نمایش در آمده است.

(جدول ۳)-ماتریس شباهت بین نتایج حاصل از اجراهای

الگوریتم ژنتیک و الگوریتم RELIEF

	GA ۱	GA ۲	GA ۳	GA ۴	GA ۵	RLF ۵۵	RLF ۵۰
GA ۱	۱	۰/۷۶	۰/۸۸	۰/۷۹	۰/۷۶	۰/۶۵	۰/۵۶
GA ۲	۰/۷۶	۱	۰/۷۱	۰/۷۴	۰/۶۸	۰/۷۱	۰/۶۶
GA ۳	۰/۸۸	۰/۷۱	۱	۰/۷۴	۰/۷۸	۰/۵۸	۰/۵۸
GA ۴	۰/۷۹	۰/۷۴	۰/۷۴	۱	۰/۶۹	۰/۷۲	۰/۶۷
GA ۵	۰/۷۶	۰/۶۸	۰/۷۸	۰/۶۹	۱	۰/۶۴	۰/۵۷
RLF۵۵	۰/۶۵	۰/۷۱	۰/۵۸	۰/۷۲	۰/۶۴	۱	۰/۹۸
RLF۵۰	۰/۵۶	۰/۶۶	۰/۵۸	۰/۶۷	۰/۵۷	۰/۹۸	۱

هر درایه از این ماتریس، شباهت ویژگی‌های انتخاب شده توسط دو اجرا را نشان می‌دهد. به عنوان مثال، درایه سطر چهار از ستون سه مشخص می‌کند که شباهت اجرای سوم الگوریتم ژنتیک با اجرای دوم آن نزدیک به ۰/۷ است. میزان شباهت، با تقسیم تعداد ویژگی‌های انتخابی مشترک بر کمینه تعداد ویژگی‌های انتخاب شده در دو اجرای مورد نظر محاسبه می‌شود.

در جدول‌های (۵ تا ۸)، اطلاعات آماری دقیق‌تری در مورد هر یک از خروجی‌ها آمده است. جدول (۴)، شامل فهرستی از پارامترهای مورد استفاده است.

فصل نهم



مقیاس مانند استفاده از دمای کلوین به جای دمای سلیسوس، می‌تواند خطای نسبی را به مقدار کوچکتری کاهش دهد. در مورد برخی پارامترها نیز، خطای نسبی و خطای مطلق هر دو بد است و این نشان از پیش‌بینی‌پذیری پایین پارامتر (و با احتمال پایین ضعف مدل) است. برای نمونه، پارامتر جهت باد (GMT DD) چنین وضعیتی دارد.

بازه ۲۵/۷-۲۳ تغییرات دارد و خطای مطلق ۱/۷۳ خطای چشم‌گیری در این بازه نیست. به نظر می‌رسد، بالابودن خطای نسبی در این موارد ناشی از وجود مقادیر نزدیک به صفر است که موجب می‌شود مخرج کسر (که مقدار واقعی است) در محاسبه خطای نسبی کوچک شده و حاصل کسر عددی بزرگ شود. لازم به ذکر است که در این موارد، تغییر

(جدول ۵)-خطا در

GA 1	Absolute Error									
	GMT DD	GMT DEW	GMT DRY	GMT FF	GMT HUM	GMT N	GMT QFE	GMT QFF	GMT VV	GMT WET
	85.72	1.69	2.83	3.32	9.13	1.05	0.76	1.51	1468.13	1.62
	Relative Error									
	GMT DD	GMT DEW	GMT DRY	GMT FF	GMT HUM	GMT N	GMT QFE	GMT QFF	GMT VV	GMT WET
	2.62	4.03	0.73	1.74	0.43	1.13	0.001	0.001	0.16	0.93
	Max Error									
	GMT DD	GMT DEW	GMT DRY	GMT FF	GMT HUM	GMT N	GMT QFE	GMT QFF	GMT VV	GMT WET
	393.25	13.49	14.95	33.46	51.80	9.06	5.12	10.18	19519.6	9.59

(جدول ۶)-خطا در

GA 2	Absolute Error									
	GMT DD	GMT DEW	GMT DRY	GMT FF	GMT HUM	GMT N	GMT QFE	GMT QFF	GMT VV	GMT WET
	86.01	1.67	2.88	3.32	9.20	1.04	0.71	1.50	1457.3	1.59
	Relative Error									
	GMT DD	GMT DEW	GMT DRY	GMT FF	GMT HUM	GMT N	GMT QFE	GMT QFF	GMT VV	GMT WET
	1.34	1.61	0.96	2.08	0.21	1.53	0.001	0.001	0.16	3.07
	Max Error									
	GMT DD	GMT DEW	GMT DRY	GMT FF	GMT HUM	GMT N	GMT QFE	GMT QFF	GMT VV	GMT WET
	364.9	14.56	14.1	34.61	49.97	8.74	4.84	12.05	19548.65	10.72

(جدول ۷)-خطا در

RELIEF 1	Absolute Error									
	GMT DD	GMT DEW	GMT DRY	GMT FF	GMT HUM	GMT N	GMT QFE	GMT QFF	GMT VV	GMT WET
	87.06	1.70	3.13	3.37	9.23	1.04	0.81	1.59	1469.82	1.73
	Relative Error									
	GMT DD	GMT DEW	GMT DRY	GMT FF	GMT HUM	GMT N	GMT QFE	GMT QFF	GMT VV	GMT WET
	1.28	2.05	1.58	1.48	0.22	0.82	0.001	0.002	0.16	5.23
	Max Error									
	GMT DD	GMT DEW	GMT DRY	GMT FF	GMT HUM	GMT N	GMT QFE	GMT QFF	GMT VV	GMT WET
	378.66	14.31	15.72	33.26	55.37	8.24	5.29	13.93	20450.89	13.01

(جدول ۸)-اطلاعات آماری مربوط به هر یک از خروجی‌های ۱۰ گانه

	GMT DD	GMT DEW	GMT DRY	GMT FF	GMT HUM	GMT N	GMT QFE	GMT QFF	GMT VV	GMT WET
Mean	102.21	0.25	12.09	4.17	51.85	2.52	825.86	1013.28	10011.58	6.37
StdDev	124.90	6.10	11.39	5.29	24.87	2.84	3.83	9.94	2782.98	7.37
min	0	-30.7	-25.5	0	5	0	810	987.8	0	-25.7
max	360	17.7	38.6	43	100	9	837.9	1049.3	30000	23

رقابت برای بقا با دیگر زیرمجموعه‌ها از جمعیت حذف شده‌اند. از سوی دیگر، می‌توان از نظر افراد خبره در حوزه هواشناسی برای ارزیابی ویژگی‌های انتخاب شده بهره گرفت. نظر فرد خبره در مورد ویژگی‌های انتخاب شده نیز حاکی از انتخاب هوشمندانه ویژگی‌هاست؛ به نحوی که پارامترهای انتخاب شده به‌طور قاطع (پارامترهایی که در اجراهای بیشتری برگزیده شده‌اند. این پارامترها را می‌توان با مراجعه به سطر آخر جدول‌های (۱ تا ۴) پیوست مشخص کرد) به‌طور معمول در پیش‌بینی وضعیت شهر همدان مؤثرند و پارامترهایی که به‌صورت قاطع نادیده گرفته شده‌اند، نیز به‌طور معمول پارامترهای نامربوط یا کم اثر هستند. البته در برخی موارد نمی‌توان به‌طور تحلیلی در مورد مؤثر بودن یا نبودن پارامترها داوری کرد.

در این پژوهش، برای پیش‌بینی پارامتر دما تنها از پارامترهای سه ساعت پیش استفاده می‌شود. در این حالت، درجه مدل پیش‌بینی یک است. می‌توان از مدل‌هایی با درجه بالاتر بهره گرفت و برای پیش‌بینی به داده‌هایی در گذشته دورتر نیز (به عنوان مثال در این کار شش ساعت گذشته) رجوع کرد.

مراجع

- جهانبخش س. و ترابی س. (۱۳۸۳)، بررسی و پیش‌بینی تغییرات دما و بارش در ایران، فصل‌نامه تحقیقات جغرافیایی، شماره ۷۴، ص. ۱۲۵-۱۰۴.
- کارآموز م. (۱۳۸۵)، پیش‌بینی دراز مدت بارش با استفاده از سیگنال‌های هواشناسی، پژوهه شماره ۳۴۲ سازمان هواشناسی کشور.

Ancell, R. (2009), "Bayesian Networks for Probabilistic Weather Forecast", PhD thesis, University of Cantabria, Spain.

Armananzas, Ruben, Saeys, Yvan, Inza, Inaki, Garcia-Torres, Miguel , Bielza, Concha , van de Peer, Yves, & Larranaga, Pedro. (2011). Peakbin Selection

روابط به کار گرفته شده برای محاسبه خطاب به صورت زیر است (Eabs خطای مطلق و Erel خطای نسبی):

$$E_{abs} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (2)$$

$$E_{rel} = \frac{\sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{n} \quad (3)$$

۷- نتیجه‌گیری و کارهای آینده

پیاده‌سازی‌های گزارش شده در این آزمایش در جهت ارزیابی کارایی دو روش انتخاب ویژگی الگوریتم ژنتیک و RELIEF انجام شده است. نتایج بدست آمده از دو جنبه قابل بررسی هستند. از دیدگاه نخست، می‌توان میزان خطاب را به عنوان یک معیار عینی برای ارزیابی نتایج به کار گرفت. از این جنبه، دقت پیش‌بینی چندان مطلوب نیست. میانگین خطای مطلق پیش‌بینی دما برای سه ساعت آینده با زیرمجموعه ویژگی‌های انتخاب شده توسط الگوریتم ژنتیک ۲/۸۳ است که چندان مطلوب به نظر نمی‌رسد؛ اما با توجه به ناکافی بودن اطلاعات چه از جهت چگالی داده (تعداد کم ایستگاه‌های هواشناسی نسبت به گستره جغرافیایی) و چه از نظر نوع و تعداد پارامترها (در هر ایستگاه به‌طور متوسط نزدیک به ۱۰ پارامتر قابل استفاده به‌صورت هر سه ساعت یکبار ثبت می‌شود) این مشکل قابل توجیه است. همچنین، کارایی تخمین‌زننده نیز در دقت پیش‌بینی دخیل است؛ اگرچه شبکه عصبی پرسپترون چندلایه یک تخمین‌زننده پرکاربرد و قابل قبول به حساب می‌آید، اما به‌طور معمول با تخمین‌زننده ایده‌آل فاصله دارد. با مفروض گرفتن تخمین‌زننده ایده‌آل می‌رسد کارایی الگوریتم‌های انتخاب مطلوب بوده است. در اجراهای الگوریتم ژنتیک همواره تعدادی از کروموزم‌های جمعیت اولیه به صورت زیرمجموعه کامل ویژگی‌ها بودند و با پیش‌روی الگوریتم ژنتیک در



- Lorenz E. N. (1993), *The Essence of Chaos*. UCL Press, pp 227.
- Luenam, P., Ingsriswang, S., Ingsrisawang, L., Aungsutatana, P., P. and Khantiyanan, W. (2010), "A Neuro-Fuzzy Approach for Daily Rainfall Prediction over the Central Region of Thailand", Hong Kong, International multiConference of Engineers and Computer Scientists, Vol I, pp. 522-525.
- Martis, R. J., Chakraborty, C., & Ray, A. K. (2009). A two-stage mechanism for registration and classification of ECG using Gaussian mixture model. *Pattern Recognition*, 42(11), pp. 2979-2988.
- Montopoli, M., Marzano, F.S. (2009), "Spatial characterization and downscaling of rain attenuation fields from numerical weather prediction models", 3rd European Conference on Antennas and Propagation, Germany.
- Nandar, A. (2009), "Bayesian Network Probability Model for Weather Prediction", International Conference of the Current Trends in Information Technology (CTIT), pp. 1-5.
- Navon, I. M. (2009), "Data assimilation for Numerical Weather Prediction: a review", *Data Assimilation for Atmospheric, Oceanic, and Hydrologic Applications*, Springer-Verlag Berlin Heidelberg, pp. 21-65.
- Persson, A., Grazzini, F. (2007), "User Guide to ECMWF forecast products", Version 4.0, European Center for medium-range Weather Forecasting.
- Pielke, R.A. (2002), "Mesoscale Meteorological Modeling". Academic Press. ISBN 0-12-554766-8.
- Qi, W., Xiaoshan, L., Sufang, A. (2009), "Mathematical model of rainfall forecast", ETP International Conference on Future Computer and Communication, pp. 112-115.
- Rashedi, Esmat, Nezamabadi-pour, Hossein, & Saryazdi, Saeid. (2013). A simultaneous feature adaptation and feature selection method for content-based image retrieval systems. *Knowledge-Based Systems*, 39, 85-94.
- Vignolo, Leandro D., Milone, Diego H., & Scharcanski, Jacob. (2013). Feature selection for face recognition based on multi-objective evolutionary wrappers. *Expert Systems with Applications*, 40(13), 5077-5084.
- Waad, Bouaguel, Ghazi, Bel Mufti, & Mohamed, Limam. (2013). A THREE-STAGE FEATURE SELECTION USING QUADRATIC PROGR-AMMING FOR CREDIT SCORING. *Applied Artificial Intelligence*, 27(8), 721-742.
- in Mass Spectrometry Data Using a Consensus Approach with Estimation of Distribution Algorithms. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(3), 760-774.
- Azam, Nouman, & Yao, JingTao. (2012). Comparison of term frequency and document frequency based feature selection metrics in text categorization. *Expert Systems with Applications*, 39(5), 4760-4768.
- Buizza, R. (2002), "Chaos and weather prediction", European Centre for Medium-Range Weather, Internal Report, Meteorological Training Course, pp. 1-28.
- de Rosnay P., Drusch, M., Wigneron, J.P., Holmes, T., Balsamo, G., Boone, A., Rudiger, C., Calvet, J.C., Kerr, Y. (2008), "Soil Moisture Remote Sensing for Numerical Weather Prediction: L-Band and C-Band Emission Modeling Over Land Surfaces, the Community Microwavw Emission Model", IGARRS IEEE International Geoscience & Remote Sensing Symposium, p. 563-566, Boston.
- Eliassen, A., Marbouty, D., Fedi, F. (2005), "Medium-Range Weather Prediction: The European Approach", ISBN-10: 0387269282, Springer, USA.
- Feng, Guozhong, Guo, Jianhua, Jing, Bing-Yi, & Hao, Lizhu. (2012). A Bayesian feature selection paradigm for text classification. *Information Processing & Management*, 48(2), 283-302.
- García-Torres, Miguel, Armañanzas, Rubén, Bielza, Concha, & Larrañaga, Pedro. (2013). Comparison of metaheuristic strategies for peakbin selection in proteomic mass spectrometry data. *Information Sciences*, 222, 229-246.
- Ghosh, Ashish, Datta, Aloke, & Ghosh, Susmita. (2013). Self-adaptive differential evolution for feature selection in hyperspectral image data. *Applied Soft Computing*, 13(4), 1969-1977.
- Hagel, E. (2009), "Development and operational application of a short-range ensemble prediction system based on the ALADIN limited area model", PhD Thesis, Eötvös Loránd University, Faculty of Science.
- Hung, N.Q., Babel, M.S., Weesakul, S., and Tripathi N. K. (2009), "An artificial neural network model for rainfall forecasting in Bangkok, Thailand", *Hydrology and Earth System Sciences*, Volume 13, Issue 8, pp.1413-1425.
- Imani, Maryam Bahojb, Keyvanpour, Mohammad Reza, & Azmi, Reza. (2013). A Novel Embedded Feature Selection Method: A Comparative Study in the Application of Text Categorization. *Applied Artificial Intelligence*, 27(5), 408-427.



نشانی رایانمۀ ایشان عبارت است از:
peyman.kabiri@iust.ac.ir



علیرضا محب‌الحمدۀ دوره

کارشناسی خود را در سال ۱۳۶۸ در دانشگاه تهران در رشته فیزیک کاربردی پشت سر گذاشت و مدرک کارشناسی ارشد خود در رشته

هواشناسی در سال ۱۳۷۰ از دانشگاه تهران دریافت کرد. وی در سال ۱۳۷۹ موفق به دریافت درجه دکتری در رشته هواشناسی با تخصص دینامیک جو از دانشگاه کمبریج شد. زمینه‌های مورد علاقه ایشان دینامیک بزرگ مقیاس جو و روش‌های عددی برای شاره‌های ژئوفیزیکی است.

نشانی رایانمۀ ایشان عبارت است از:
amoheb@ut.ac.ir

WMO, (2010). Guide to Agricultural Meteo-rological Practices, WMO-No 134, Geneva 2, Switzerland, part 7.4, [World Meteorological Organization's handbook for Agricultural Meteorology].

Woods, A. (2006). Medium-Range Weather Prediction: The European Approach, Springer.

Wu, J., Zhou, J. (2009), "Support Vector Regression Based on Particle Swarm Optimization and Projection Pursuit Technology for Rainfall Forecasting", International Conference on Computational Intelligence and Security, pp. 227-233.

Zhang, Kui, Li, Yuhua, Scarf, Philip, & Ball, Andrew. (2011). Feature selection for high-dimensional machinery fault diagnosis data using multiple models and Radial Basis Function networks. Neurocomputing, 74(17), 2941-2952.



محسن مشکی دوره کارشناسی خود را در سال ۱۳۸۵ از دانشگاه پیام نور مشهد در رشته مهندسی کامپیوتر (نرم افزار) پشت سر گذاشت و مدرک کارشناسی ارشد خود در همین رشته (گرایش هوش مصنوعی) در سال ۱۳۸۸ از دانشگاه علم و صنعت ایران دریافت کرد. وی از سال ۱۳۸۸ مشغول به تحصیل در دوره دکترای مهندسی کامپیوتر (گرایش هوش مصنوعی) در دانشگاه علم و صنعت است. زمینه‌های مورد علاقه وی شناسایی الگو، داده‌کاوی، انتخاب ویژگی، رگرسیون ترکیبی و متن‌کاوی است.

نشانی رایانمۀ ایشان عبارت است از:

moshki@iust.ac.ir



پیمان کبیری دوره کارشناسی خود را در سال ۱۳۷۱ از دانشگاه علم و صنعت ایران در رشته مهندسی کامپیوتر (سخت افزار) پشت سر گذاشت و مدرک کارشناسی ارشد خود در سامانه‌های بلاذرنگ در سال ۱۳۷۵ از دانشگاه ناتینگهام ترن特 دریافت کرد. ایشان در سال ۱۳۷۹ موفق به دریافت درجه دکتری در رشته علوم کامپیوتر از دانشگاه ترن特 ناتینگهام شد. موضوع پروژه دوره دکترای ایشان جبران خطای ناشی از عدم دقیقت در ساخت یک ربات بود. زمینه‌های مورد علاقه ایشان سنجش از دور، امنیت شبکه، رباتیک، کاهش ابعاد داده، فرآگیری ماشینی، اسکن سه بعدی، مصورسازی و پایش انتشارات صوتی است.

سال ۱۳۹۲ شماره ۲ پیاپی ۲۰



فصلنامه

۸- پیوست

در جدول (۱) پیوست، مقدار صفر برای درایه سطر ۱ و ستون GMT DD Arak به این معنی است که پاسخ اجرای اول الگوریتم ژنتیک نشان می‌دهد که نیازی به استفاده از پارامتر جهت باد (GMT DD) نیست. ستون‌هایی که در

جدول‌های (۱ تا ۴) پیوست با رنگ زرد (در نسخه سیاه و سفید تیره‌ترند) مشخص شده‌اند مربوط به پارامترهایی است که در هر هفت اجرای دو الگوریتم انتخاب ویژگی برگزیده شده‌اند. سطر آخر جدول‌های (۱ تا ۴)، تعداد بارهای انتخاب هر یک از پارامترها در هفت اجرای نشان می‌دهد.

جدول (۱ پیوست)- ویژگی‌های انتخاب شده در اجرای هفت‌گانه الگوریتم ژنتیک و الگوریتم RELIEF

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Arak	0	1	2	0	0	5	3	7	5	1	7	7	7	7	4	1	6	5	0	2	4	6	4	7	
Sum	0	7	2	0	0	5	3	7	5	1	7	7	7	7	4	1	6	5	0	2	4	6	4	7	
RELIEF 50	0	1	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	0	0	0	1	1	0	1
RELIEF 55	0	1	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	0	0	1	0	0	1	0
GA 1	0	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
GA 2	0	1	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	0	1	0
GA 3	0	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1
GA 4	0	1	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	0	1	0
GA 5	0	1	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
Kangavar	0	1	2	0	0	5	3	7	5	1	7	7	7	7	4	1	6	5	0	2	4	6	4	7	
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kangavar	0	1	0	0	0	1	1	1																	

جدول (۳ پیوست)- ویژگی های انتخاب شده در اجراهای هفت گانه الگوریتم زنیک و الگوریتم RELIEF (ادامه)

		Tehran	GMT DRY	1	72	73	74	75
	Tehran		GMT DEW	0	0	0	0	1
	Tehran		GMT DD	0	0	0	0	1
	SarpolZohab	GMT WET	1	0	0	0	0	1
	SarpolZohab	GMT VV	0	1	0	1	0	1
	SarpolZohab	GMT VP	1	0	0	0	1	0
	SarpolZohab	GMT QFF	1	0	0	1	0	0
	SarpolZohab	GMT QFE	1	0	0	0	1	0
	SarpolZohab	GMT N	0	0	0	0	1	0
	SarpolZohab	GMT HUM	0	0	0	0	0	1
	SarpolZohab	GMT FF	0	1	0	0	0	1
	SarpolZohab	GMT DRY	1	1	0	0	0	1
	SarpolZohab	GMT DEW	1	1	0	0	0	1
	SarpolZohab	GMT DD	0	1	0	0	0	1
	Sanandaj	GMT WET	1	1	0	0	0	1
	Sanandaj	GMT HUM	1	1	0	0	0	1
	Sanandaj	GMT DRY	1	1	0	0	0	1
	Sanandaj	GMT DEW	0	0	1	0	0	1
	Sanandaj	GMT VV	1	1	0	0	0	1
	Malayer	GMT VP	1	0	1	0	0	1
	Malayer	GMT QFF	1	1	0	0	0	1
	Malayer	GMT QFE	1	1	0	0	0	1
	Malayer	GMT N	0	1	0	0	0	1
	Malayer	GMT HUM	0	1	0	0	0	1
	Malayer	GMT FF	0	1	0	0	0	1
GA 1	0	0	0	0	1	0	0	1
GA 2	0	0	0	1	1	0	0	1
GA 3	0	0	0	1	1	0	0	1
GA 4	0	1	0	1	0	0	0	1
GA 5	0	1	0	1	1	0	1	0
RELIEF 55	0	1	1	1	1	0	1	0
RELIEF 50	1	1	1	1	1	0	1	0
Sum	1	4	3	7	6	2	4	5

جدول (٤ پیوست) - وزیری‌های انتخاب شده در اجراهای هفت‌گانه الگوریتم ژنتیک و الگوریتم RELIEF (ادامه)

	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92
	Tehran	Zanjan	GMT WET														
															GMT VV		
															GMT QFF		
															GMT N		
															GMT HUM		
															GMT FF		
																GMT DEW	
																GMT DD	
																GMT WET	
																GMT VV	
																GMT QFF	
																GMT QFE	
GA 1	1	0	0	1	1	1	1	1	1	1	0	0	1	0	1	1	0
GA 2	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	0	1
GA 3	0	0	0	1	0	1	1	1	1	0	0	0	1	1	1	0	0
GA 4	0	1	0	1	1	1	0	1	1	0	0	0	1	1	1	1	0
GA 5	0	0	0	1	1	1	1	1	0	0	0	0	0	0	1	0	1
RELIEF 55	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0
RELIEF 50	0	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1	0
Sum	2	3	2	6	4	6	5	7	6	1	0	6	5	7	6	2	3

الگوریتم RELIEF که در آن ۵۵ پارامتر انتخاب گردید) انتخاب شده‌اند.

جدول (۵ پیوست)، فهرستی از پارامترها را نشان می‌دهد که به طور مشترک در دو اجرای GA5 و RELIEF55 (اجرا یک) از



جدول (۵ پیوست) - شباهت زیر مجموعه انتخابی توسط GA1 و RELIEF55

آب و هوایی و سطح پنجم وزن داده شده به ویژگی توسعه الگوریتم RELIEF را نشان می دهد. ستون هایی از جدول که تیره شده اند، متناظر با پارامتر هایی هستند که الگوریتم RELIEF بیشترین وزن را به آنها اختصاص داده است.

جدول های (۶ تا ۸) پیوست مربوط به اجرای دوم الگوریتم RELIEF است که در آن ۵۵ ویژگی انتخاب شده است. سطر اول، شماره ویژگی انتخاب شده به ترتیب اولویت، سطر دوم شماره ویژگی، سطر سوم شهر، سطر چهارم پارامتر

جدول (۶ پیوست) - وزن ویژگی های انتخاب شده توسط الگوریتم RELIEF در اجرای دوم 55

جدول (۷ پیوست) - وزن ویژگی‌های انتخاب شده توسط الگوریتم RELIEF در اجرای دوم ۵۵ (ادامه)

۶۰	SarpolZohab	GMT QFF	0.0254	۹۲	Kashan	GMT DD	-0.0235
۵۹	Hamedan	GMT WET	0.0254	۹۱	Hamedan	GMT DD	-0.0235
۵۸	Tehran	GMT DRY	0.0264	۹۰	Zanjan	GMT FF	-0.0023
۵۷	Kangavar	GMT QFF	0.0279	۸۹	Kashan	GMT FF	-0.0014
۵۶	Arak	GMT HUM	0.0284	۸۸	Kangavar	GMT DRY	-0.0009
۵۵	Kermanshah	GMT QFF	0.0291	۸۷	Kermanshah	GMT FF	-0.0007
۵۴	Kashan	GMT WET	0.0302	۸۶	Kermanshah	GMT DD	0.0023
۵۳	Sanandaj	GMT HUM	0.0313	۸۵	Kangavar	GMT DD	0.0026
۵۲	Tehran	GMT FF	0.032	۸۴	Malayer	GMT DRY	0.0027
۵۱	Tehran	GMT QFF	0.0347	۸۳	Kermanshah	GMT DRY	0.0027
۵۰	Zanjan	GMT DD	0.0369	۸۲	Sanandaj	GMT DRY	0.0036
۴۹	Malayer	GMT HUM	0.0378	۸۱	Arak	GMT DD	0.0041
۴۸	Hamedan	GMT QFF	0.0401	۸۰	Kangavar	GMT FF	0.0052
۴۷	Malayer	GMT DD	0.0411	۷۹	Arak	GMT DRY	0.0063
۴۶	Malayer	GMT VV	0.0419	۷۸	Kangavar	GMT WET	0.0089
۴۵	Kashan	GMT QFF	0.0439	۷۷	Hamedan	GMT FF	0.0091
۴۴	Tehran	GMT WET	0.044	۷۶	Malayer	GMT WET	0.0097
۴۳	Kermanshah	GMT HUM	0.0446	۷۵	Sanandaj	GMT WET	0.0101
۴۲	Kashan	GMT HUM	0.0447	۷۴	Kashan	GMT DRY	0.0121
۴۱	Zanjan	GMT DEW	0.0471	۷۳	Kermanshah	GMT WET	0.0127
۴۰	Zanjan	GMT QFF	0.0478	۷۲	SarpolZohab	GMT DD	0.0135
۳۹	Kangavar	GMT DEW	0.048	۷۱	Hamedan	GMT DRY	0.0141
۳۸	Kermanshah	GMT VV	0.0507	۷۰	Zanjan	GMT DRY	0.0145
۳۷	Arak	GMT DEW	0.054	۶۹	Arak	GMT WET	0.0167
۳۶	Hamedan	GMT HUM	0.055	۶۸	Kangavar	GMT HUM	0.0168
۳۵	Zanjan	GMT HUM	0.0552	۶۷	SarpolZohab	GMT DRY	0.0171
۳۴	Kashan	GMT N	0.0569	۶۶	Malayer	GMT FF	0.0175
۳۳	Malayer	GMT N	0.0576	۶۵	SarpolZohab	GMT FF	0.0176
۳۲	SarpolZohab	GMT HUM	0.0603	۶۴	SarpolZohab	GMT WET	0.0177
۳۱	Kermanshah	GMT QFF	0.0607	۶۳	Zanjan	GMT WET	0.0206
۶۱	Arak	GMT FF	0.0241	۶۲	Tehran	GMT DD	0.0221

جدول (۸ پیوست) - وزن ویژگی‌های انتخاب شده توسط الگوریتم RELIEF در اجرای دوم ۵۵ (ادامه)

۹	۲	Kashan	GMT DD	-0.0235
۹۱	۱	Hamedan	GMT DD	-0.0235
۹۰	۰	Zanjan	GMT FF	-0.0023
۸۹	۹	Kashan	GMT FF	-0.0014
۸۸	۸	Kangavar	GMT DRY	-0.0009
۸۷	۷	Kermanshah	GMT FF	-0.0007
۸۶	۶	Kermanshah	GMT DD	0.0023
۸۵	۵	Kangavar	GMT DD	0.0026
۸۴	۴	Malayer	GMT DRY	0.0027
۸۳	۳	Kermanshah	GMT DRY	0.0027
۸۲	۲	Sanandaj	GMT DRY	0.0036
۸۱	۱	Arak	GMT DD	0.0041
۸۰	۰	Kangavar	GMT FF	0.0052
۷۹	۹	Arak	GMT DRY	0.0063
۷۸	۸	Kangavar	GMT WET	0.0089
۷۷	۷	Hamedan	GMT FF	0.0091
۷۶	۶	Malayer	GMT WET	0.0097
۷۵	۵	Sanandaj	GMT WET	0.0101
۷۴	۴	Kashan	GMT DRY	0.0121
۷۳	۳	Kermanshah	GMT WET	0.0127
۷۲	۲	SarpolZohab	GMT DD	0.0135
۷۱	۱	Hamedan	GMT DRY	0.0141
۷۰	۰	Zanjan	GMT DRY	0.0145
۶۹	۹	Arak	GMT WET	0.0167
۶۸	۸	Kangavar	GMT HUM	0.0168
۶۷	۷	SarpolZohab	GMT DRY	0.0171
۶۶	۶	Malayer	GMT FF	0.0175
۶۵	۵	SarpolZohab	GMT FF	0.0176
۶۴	۴	SarpolZohab	GMT WET	0.0177
۶۳	۳	Zanjan	GMT WET	0.0206
۶۲	۲	Tehran	GMT DD	0.0221
۶۱	۱	Arak	GMT FF	0.0241

فصل نهم



سال ۱۳۹۲ شماره ۲ پیاپی ۲۰