



پیش‌بینی دمای بحرانی برخی از ترکیبات آلی با استفاده از توصیف‌کننده‌های مکانیک کوانتومی و شبکه‌های عصبی مصنوعی

مهدی نکوئی*، مجید محمدحسینی، حمیدرضا برزگر، مهدی رحیمی

دانشگاه آزاد اسلامی، واحد شاهرود، گروه شیمی، شاهرود، ایران

تاریخ ثبت اولیه: ۱۳۹۰/۳/۱۴، تاریخ دریافت نسخه اصلاح شده: ۱۳۹۰/۵/۲۰، تاریخ پذیرش قطعی: ۱۳۹۰/۵/۲۷

چکیده

روشهای خطی و غیر خطی ارتباط کمی ساختار- خاصیت جهت پیش‌بینی دمای بحرانی برخی از ترکیبات آلی مطالعه شد. گروه مناسبی از توصیف‌کننده‌های مکانیک کوانتومی محاسبه شدند و از روش انتخاب مرحله‌ای برای بدست آوردن بهترین توصیف‌کننده‌ها که بیشترین ارتباط را با خاصیت مورد نظر دارند استفاده شد. در ابتدا مدل خطی رگرسیون خطی چندگانه ساخته شد. سپس برای به دست آوردن مدل بهتر، از شبکه عصبی مصنوعی استفاده گردید. به همین خاطر، توصیف‌کننده‌های انتخاب شده بوسیله روش مرحله‌ای به عنوان ورودی شبکه عصبی مصنوعی مورد استفاده قرار گرفتند. نتایج، برتری روش غیرخطی شبکه عصبی مصنوعی را نسبت به روش خطی رگرسیون خطی چندگانه نشان می‌دهد.

واژه‌های کلیدی: ارتباط کمی ساختار- خاصیت، دمای بحرانی، رگرسیون خطی چندگانه، شبکه عصبی مصنوعی، توصیف‌کننده‌های مکانیک کوانتومی.

۱. مقدمه

نسبتاً ضعیف بین مولکولی، مولکولها را در حالت مایع نگه دارد. اندازه‌گیری دمای بحرانی ترکیبات از اهمیت بالایی برخوردار است. روشهای آزمایشگاهی جهت اندازه‌گیری این پارامتر، وقت گیر، هزینه بر، و نسبتاً مشکل است. بنابراین اندازه‌گیری آن با روشهای محاسباتی و کمومتریکس از اهمیت بالایی برخوردار است. یکی از این روشهاکه امروزه از آن جهت پیش‌بینی خواص

برای هر گاز دمایی وجود دارد که بالاتر از آن، هر چند هم فشار را زیاد کنیم، مایع شدن آن گاز ناممکن است. این دما را دمای بحرانی آن گاز می‌نامند. دمای بحرانی یک گاز نشانه‌ای از قدرت نیروهای جاذبه بین مولکولی آن گاز است. ماده‌ای با نیروهای جاذبه ضعیف، دمای بحرانی پایین دارد؛ بالاتر از این دما، حرکت مولکولی آن قدر شدید است که نمی‌گذارد نیروهای

*عهده‌دار مکاتبات: مهدی نکوئی

نشانی: شاهرود، دانشگاه آزاد اسلامی، گروه شیمی

تلفن: ۰۲۷۳-۳۳۹۴۵۳۰ پست الکترونیک: E-Mail: m_nekoei1356@yahoo.com

لایه میانی در کل بسیار پیچیده است. اما می توان گفت که تعداد نرون های لایه میانی تابعی از تعداد عناصر بردار ورودی و همچنین حداکثر تعداد نواحی از فضای ورودی، به طور خطی جداپذیر، است. از این رو تعداد نرون های لایه مخفی عموماً به طور تجربی به دست می آید. هر نرون توسط خروجی اش به نرون های لایه بعد متصل بوده ولی با نرون های لایه خودش ارتباط ندارد. خروجی هر نرون توسط رابطه زیر تعریف می شود [۲]:

$$\alpha = f(\sum P_i W_{ij} + b_j)$$

که در این رابطه $W_{j,i}$ مقدار وزن اتصال بین نرون i لایه مذکور با نرون i لایه قبل است که بیانگر اهمیت ارتباط بین دو نرون در دو لایه متوالی است، b_j وزن مربوط به بایاس (Bias) برای نرون i لایه قبل، P_i مقدار خروجی از نرون i لایه قبل، α مقدار خروجی از نرون i لایه بعد است.

توابع زیادی وجود دارند که در انتقال اعداد از لایه قبل به لایه بعد ممکن است مورد استفاده قرار گیرند. از جمله میتوان توابع سیگموئید، گوس، تانژانت هیپربولیک و سکانت هیپربولیک را نام برد. ولی تابع سیگموئید بیشترین استفاده را در علوم مختلف دارد. این تابع به صورت زیر است:

$$f(z) = 1 / (1 + \exp(-z))$$

شبکه های عصبی مصنوعی شامل دو مرحله هستند. مرحله اول آموزش شبکه است که بر اساس تعدادی از داده های اندازه گیری شده (حدود ۷۰ درصد کل داده ها) انجام می پذیرد. در مرحله دوم مدل با حدود ۲۰ تا ۳۰ درصد مقادیر اندازه گیری شده صحت سنجی می گردد. در مدل سازی به کمک شبکه های عصبی مصنوعی بسته به متغیرهای ورودی و ارتباط آن ها با پارامتر مورد نظر که مدل می گردد، حدود ۱۰ درصد اطلاعات را برای تست و ارزیابی مدل به کار می برند. البته این داده ها نبایستی در مراحل آموزش و صحت سنجی استفاده شده باشند [۱۷-۱۳].

۲. بخش محاسباتی

۲-۱. دستگاه های مورد استفاده

یک کامپیوتر پنتیوم IV (CPU=3.06 GHz) با ویندوز XP برای انجام محاسبات مورد استفاده قرار گرفت. برای رسم و بهینه سازی

ترکیبات استفاده می شود روش ارتباط کمی ساختار- خاصیت (QSPR) است. هدف از مطالعات QSPR پیدا کردن رابطه ای است که بین رفتار فیزیکیو شیمیایی یک مولکول با پارامترهای ساختاری آن وجود دارد [۹-۱]. نتایج این مطالعات علاوه بر شفاف سازی نحوه ارتباط بین خواص مولکولها و ویژگی های ساختاری آنها به پژوهشگران در پیش بینی رفتار مولکولهای جدید براساس رفتار مولکولهای مشابه کمک می کند.

هدف از مطالعه حاضر، پیش بینی دمای بحرانی برخی از ترکیبات آلی با استفاده از توصیف کننده های مکانیک کوانتومی و شبکه های عصبی مصنوعی است.

شبکه های عصبی مصنوعی (Artificial Neural Networks)

ایده استفاده از شبکه های عصبی مصنوعی خیلی تازه نیست. ولی استفاده کاربردی از آن تقریباً از سال ۱۹۴۶ شروع شد. هو (HU) در تز خود برای پیش بینی وضع هوا از آن استفاده نمود. ولی به دلیل محاسبات پیچیده و عدم استفاده از کامپیوتر موفقیت چندانی نداشت. استفاده علمی و گسترده از این روش در علوم مختلف مهندسی از سال ۱۹۸۶ با معرفی روش پس انتشار خطا توسط روملهارت و همکاران سرعت بسیار بالاتری به خود گرفت [۱۰-۱۲].

شبکه عصبی مصنوعی از مجموعه ای از نرون ها با ارتباطات داخلی بین یکدیگر تشکیل می شود که قادر است بر اساس اطلاعات و داده های ورودی، جوابهای خروجی را تولید نماید. شبکه های عصبی معمولاً به صورت لایه لایه و منظم ایجاد میگردند. اولین لایه که اطلاعات و داده های ورودی به آن وارد میگردند لایه ورودی است. لایه های میانی، لایه های مخفی و آخرین لایه که جواب های خروجی از مدل را فراهم می نماید لایه خروجی است [۱۰]. ساده ترین و معمولترین نوع شبکه عصبی که در بسیاری از علوم از آن استفاده می گردد، شبکه های عصبی چند لایه پیشخورند (MLP, Multi-Layer Perceptron) که همراه با ناظر از روش پس انتشار خطا برای آموزش کمک می گیرند.

در این شبکه تعداد نرون های لایه ورودی برابر با تعداد عناصر بردار ورودی و تعداد نرون های لایه خروجی برابر با تعداد عناصر بردار خروجی است. آنالیز دقیق و واقعی برای پیدا کردن تعداد نرون های

پیش‌بینی شامل ۲۹ ترکیب است. مقادیر دمای بحرانی به عنوان متغیر وابسته و توصیف کننده‌ها به عنوان متغیر مستقل انتخاب شدند.

۲-۳. محاسبه توصیف کننده‌ها

در اولین قدم باید مجموعه آزمایشی (دمای بحرانی ترکیبات) و ساختارهای متناظر جمع‌آوری شده و به صورت قابل پردازش برای کامپیوتر درآید. در مورد داده‌های آزمایشی که مقادیر عددیند مشکلی نداریم، مساله مهم بیان ساختارهای شیمیایی به شکل قابل پذیرش برای کامپیوتر است. ابتدا ساختارهای مولکولی به کمک نرم افزار هایپرکم رسم و به وسیله روش AM1 بهینه گردید سپس محاسبه توصیف کننده‌ها به کمک این نرم‌افزار و نرم‌افزار گوسین با خط دستور B3LYP/6-31+G** انجام شد.

شکل مولکولها از نرم افزار هایپرکم (hyperchem) (نسخه ۷/۰) استفاده گردید. برای محاسبه توصیف کننده‌های مکانیک کوانتومی و شیمیایی از نرم افزار گوسین در سطح B3LYP/6-31+G** و هایپرکم استفاده شد. از نرم افزار SPSS برای انتخاب بهترین توصیف کننده‌ها و از برنامه ANN در محیط MATLAB برای پیش‌بینی دمای بحرانی و به دست آوردن سایر پارامترهای آماری استفاده گردید.

۲-۲. سری داده‌ها

سری داده‌ها مربوط به دمای بحرانی ۱۴۶ ترکیب از آلکانهای خطی است. ترکیبات به صورت تصادفی به سه سری شامل سری آموزش، سری ارزیابی و سری پیش‌بینی تقسیم شدند (جدول ۱). سری آموزش شامل ۸۸ ترکیب، سری ارزیابی شامل ۳۱ ترکیب و سری

جدول ۱- مقادیر تجربی و محاسبه شده دمای بحرانی برای ترکیبات مختلف برای مجموعه‌های آموزشی، ارزیابی و پیش‌بینی در مدل SW-ANN همراه با مقادیر خطای نسبی.

No.	Compound	Exp ^a	ANN ^b	E (%) ^c
	سری آموزش			
۱	Ethane	۳۰۵/۴۲	۳۲۳/۵۸	۵/۹۵
۲	Propane	۳۶۹/۸۰	۳۵۵/۱۱	-۳/۹۷
۳	n-Butane	۴۲۵/۱۴	۴۱۴/۵۰	-۲/۵۰
۴	N-Pentane	۴۶۹/۶۹	۴۷۵/۷۰	۱/۲۸
۵	2,2-Dimethylpropane	۴۳۳/۷۵	۴۴۹/۲۶	۳/۵۸
۶	2-Methylpentane	۴۹۷/۵۰	۴۹۷/۳۱	-۰/۰۴
۷	2,2-Dimethyl butane	۴۸۸/۷۰	۴۸۸/۱۷	-۰/۱۱
۸	N-Heptane	۵۴۰/۳۰	۵۴۱/۱۰	۰/۱۵
۹	2-Methyl hexane	۵۳۰/۱۰	۵۲۰/۱۷	-۱/۸۷
۱۰	3-Ethylpentane	۵۳۵/۲۰	۵۳۰/۵۰	-۰/۸۸
۱۱	2,3- Dimethylpentane	۵۳۷/۳۰	۵۳۴/۸۴	-۰/۴۶
۱۲	3,3- Dimethylpentane	۵۳۶/۳۰	۵۳۲/۷۶	-۰/۶۶
۱۳	2,2,3-Trimethylbutane	۵۳۴/۱۰	۵۲۸/۲۴	-۱/۱۰
۱۴	N-Octane	۵۶۸/۸۳	۵۶۸/۵۹	-۰/۰۴
۱۵	3- Methylheptane	۵۶۳/۶۰	۵۷۱/۱۳	۱/۳۴
۱۶	3-Ethylehexane	۵۶۵/۴۰	۵۷۳/۹۹	۱/۵۲
۱۷	2,3- Dimethylhexane	۵۶۳/۴۰	۵۷۰/۸۰	۱/۳۱
۱۸	2,4- Dimethylhexane	۵۵۳/۵۰	۵۵۳/۵۴	۰/۰۱
۱۹	2,5- Dimethylhexane	۵۵۰/۰۰	۵۴۶/۹	-۰/۵۶
۲۰	3,4- Dimethylhexane	۵۶۸/۸۰	۵۷۹/۱۶	۱/۸۲

ادامه جدول ۱.

۲۱	3- Methyl-3-ethylpentane	۵۷۶/۵۰	۵۸۹/۷۶	۲/۳۰
۲۲	2,2,4- Trimethylpentane	۵۴۳/۹۰	۵۳۴/۷۹	-۱/۶۷
۲۳	2,3,3- Trimethylpentane	۵۷۳/۵۰	۵۸۵/۸۲	۲/۱۵
۲۴	n-Nonane	۵۹۴/۶۰	۵۹۱/۰۳	-۰/۶۰
۲۵	3- Methyl-octane	۵۸۹/۸۰	۵۸۷/۱۹	-۰/۴۴
۲۶	4- Methyl-octane	۵۸۹/۸۰	۵۸۷/۱۹	-۰/۴۴
۲۷	3- Ethylheptane	۵۹۴/۳۰	۵۹۰/۸۰	-۰/۵۹
۲۸	2,3-Dimethylheptane	۵۱۹/۴۰	۵۸۸/۴۸	-۰/۴۹
۲۹	2,5- Dimethylheptane	۵۸۱/۰۰	۵۷۹/۸۶	-۰/۲۰
۳۰	3,3- Dimethylheptane	۵۹۱/۰۰	۵۸۸/۱۶	-۰/۴۸
۳۱	3,5- Dimethylheptane	۵۸۵/۵۰	۵۸۳/۶۷	-۰/۳۱
۳۲	3-Ethyl-2-methylhexane	۵۹۵/۸۰	۵۹۱/۹۸	-۰/۶۴
۳۳	3-Ethyl-3-methylhexane	۶۰۳/۵۰	۵۹۸/۰۵	-۰/۹۰
۳۴	3-Ethyl-4-methylhexane	۶۰۰/۳۰	۵۹۵/۵۳	-۰/۷۹
۳۵	2,2,3-Tri methylhexane	۵۹۱/۳۰	۵۸۸/۴	-۰/۴۹
۳۶	2,2,5- Tri methylhexane	۵۶۹/۷۰	۵۶۹/۴۶	-۰/۰۴
۳۷	2,3,4- Tri methylhexane	۵۹۷/۴۰	۵۹۳/۲۵	-۰/۶۹
۳۸	2,3,5- Tri methylhexane	۵۸۲/۶۰	۵۸۱/۲۳	-۰/۲۳
۳۹	3,3,4- Tri methylhexane	۶۰۳/۷۰	۵۹۸/۲۱	-۰/۹۱
۴۰	3-Ethyl-2,2-dimethylpentane	۵۹۵/۷۰	۵۹۱/۹۱	-۰/۶۴
۴۱	3-Ethyl-2,4-dimethylpentane	۵۹۷/۴۰	۵۹۳/۲۵	-۰/۶۹
۴۲	2,2,3,4- Tetramethylpentane	۵۹۲/۷۰	۵۸۹/۵۲	-۰/۵۴
۴۳	2,2,4,4- Tetramethylpentane	۵۷۴/۷۰	۵۷۴/۲۴	-۰/۰۸
۴۴	n-Decane	۶۱۷/۹۰	۶۱۸/۱۸	۰/۰۵
۴۵	3-Methylnonane	۶۱۳/۷۰	۶۱۴/۷۰	۰/۱۶
۴۶	4-Methylnonane	۶۱۳/۷۰	۶۱۴/۷	۰/۱۶
۴۷	3-Ethyl-octane	۶۱۸/۲۰	۶۱۸/۴۳	۰/۰۴
۴۸	2,3-Dimethyloctane	۶۱۵/۳۰	۶۱۶/۰۳	۰/۱۲
۴۹	2,5-Dimethyloctane	۶۰۴/۹۰	۶۰۷/۲۷	۰/۳۹
۵۰	2,7-Dimethyloctane	۶۰۰/۵۰	۶۰۳/۵۲	۰/۵۰
۵۱	3,4-Dimethyloctane	۶۱۹/۸۰	۶۱۹/۷۴	-۰/۰۱
۵۲	3,5-Dimethyloctane	۶۰۹/۴۰	۶۱۱/۰۹	۰/۲۸
۵۳	3,6-Dimethyloctane	۶۰۹/۴۰	۶۱۱/۰۹	۰/۲۸
۵۴	4,5-Dimethyloctane	۶۱۹/۸۰	۶۱۹/۷۴	-۰/۰۱
۵۵	3-Ethyl-2-methylheptane	۶۱۹/۸۰	۶۱۹/۷۴	-۰/۰۱
۵۶	3-Ethyl-4-methylheptane	۶۲۴/۲۰	۶۲۳/۲۸	-۰/۱۵
۵۷	3-Ethyl-5-methylheptane	۶۱۳/۹۰	۶۱۴/۸۷	۰/۱۶
۵۸	4-Ethyl-2-methylheptane	۶۰۹/۴۰	۶۱۱/۰۹	۰/۲۸
۵۹	4-Ethyl-4-methylheptane	۶۲۷/۴۰	۶۲۵/۸۰	-۰/۲۵

ادامه جدول ۱.

۶۰	2,2,3-Tri methylheptane	۶۱۵/۲۰	۶۱۵/۹۵	۰/۱۲
۶۱	2,2,5-Tri methylheptane	۵۹۸/۱۰	۶۰۱/۴۸	۰/۵۶
۶۲	2,2,6-Tri methylheptane	۵۹۳/۶۰	۵۹۷/۶۴	۰/۶۸
۶۳	2,3,3-Tri methylheptane	۶۲۳/۲۰	۶۲۲/۴۸	-۰/۱۲
۶۴	2,3,5-Tri methylheptane	۶۱۱/۰۰	۶۱۲/۴۴	۰/۲۳
۶۵	2,3,6-Tri methylheptane	۶۰۶/۵۰	۶۰۸/۶۳	۰/۳۵
۶۶	2,4,5-Tri methylheptane	۶۱۱/۰۰	۶۱۲/۴۴	۰/۲۳
۶۷	2,5,5-Tri methylheptane	۶۰۶/۱۰	۶۰۸/۲۹	۰/۳۶
۶۸	3,3,4-Tri methylheptane	۶۲۷/۷۰	۶۲۶/۰۴	-۰/۲۷
۶۹	3,4,4-Tri methylheptane	۶۲۷/۷۰	۶۲۶/۰۴	-۰/۲۷
۷۰	3,4-Diethylhexane	۶۳۹/۹۰	۶۴۰/۹۷	۰/۱۷
۷۱	3,3-Diethylhexane	۶۲۸/۷۰	۶۳۴/۰۶	۰/۸۵
۷۲	3-Ethyl-2,3-dimethylhexane	۶۳۵/۷۰	۶۳۲/۰۷	-۰/۵۷
۷۳	3-Ethyl-2,5-dimethylhexane	۶۱۱/۰۰	۶۱۲/۴۴	۰/۲۳
۷۴	4-Ethyl-2,2-dimethylhexane	۶۰۲/۶۰	۶۰۵/۳۱	۰/۴۵
۷۵	4-Ethyl-2,4-dimethylhexane	۶۱۸/۶۰	۶۱۸/۷۵	۰/۰۳
۷۶	2,2,3,3,-Tetramethylhexane	۶۲۳/۰۰	۶۲۲/۳۲	-۰/۱۱
۷۷	2,2,3,4,-Tetramethylhexane	۶۲۱/۲۰	۶۲۰/۸۷	-۰/۰۵
۷۸	2,2,4,4,-Tetramethylhexane	۵۹۹/۳۰	۶۰۲/۵۰	۰/۵۳
۷۹	2,2,4,5,-Tetramethylhexane	۵۹۹/۷۰	۶۰۲/۸۴	۰/۵۲
۸۰	2,3,3,5,-Tetramethylhexane	۶۱۴/۴۰	۶۱۵/۲۸	۰/۱۴
۸۱	2,3,4,4,-Tetramethylhexane	۶۲۹/۲۰	۶۲۷/۲۰	-۰/۳۲
۸۲	2,3,4,5,-Tetramethylhexane	۶۲۲/۹۰	۶۲۲/۲۴	-۰/۱۱
۸۳	3-Ethyl-2,2,3-Trimethylpentane	۶۴۰/۹۰	۶۳۵/۷۶	-۰/۰۸
۸۴	3-Ethyl-2,2,4-Trimethylpentane	۶۲۱/۲۰	۶۲۰/۸۷	-۰/۰۵
۸۵	2,2,3,4,4-pentamethylpentane	۶۱۶/۶۰	۶۱۷/۱۱	۰/۰۸
۸۶	n-Dodecane	۶۵۸/۲۰	۶۵۹/۱۳	۰/۱۴
سری ارزیابی				
۱	2-Methyl Propane	۴۰۷/۸۵	۴۰۵/۵۲	-۰/۵۷
۲	n-Hexane	۵۰۷/۵۰	۵۰۸/۴۵	۰/۱۹
۳	2,3 -Dimethyl butane	۴۹۹/۹۰	۴۹۹/۹۰	۰/۰۰
۴	3- Methyl hexane	۵۳۵/۲۰	۵۳۰/۵۰	-۰/۸۸
۵	2,4- Dimethylpentane	۵۱۹/۷۰	۵۰۰/۴۷	-۳/۷۰
۶	2-Methylheptane	۵۵۹/۶۰	۵۶۴/۴۶	۰/۸۷
۷	2,2-Dimethylhexane	۵۴۹/۸۰	۵۴۶/۵۱	-۰/۶۰
۸	3,3- Dimethylhexane	۵۶۲/۰۰	۵۶۸/۵۱	۱/۱۶
۹	2,2,3-Trimethylpentane	۵۶۳/۴۰	۵۷۰/۸۰	۱/۳۱
۱۰	2-Methyloctane	۸۵۸/۳۰	۸۸۳/۵۰	-۰/۳۱
۱۱	2-Ethylheptane	۵۹۴/۳۰	۵۹۰/۸۰	-۰/۵۹

ادامه جدول ۱.

۱۲	2,2-Dimethylheptane	۵۷۸/۵۰	۵۷۷/۶۸	-۰/۱۴
۱۳	2,6-Dimethylheptane	۵۷۶/۶۰	۵۷۵/۹۸	-۰/۱۱
۱۴	4,4-Dimethylheptane	۵۹۱/۴۰	۵۸۸/۴۸	-۰/۴۹
۱۵	4-Ethyl-2-methylhexane	۵۸۵/۵۰	۵۸۳/۶۷	-۰/۳۱
۱۶	2,3,3-Tri methylhexane	۵۹۹/۳۰	۵۹۴/۷۵	-۰/۷۶
۱۷	2,4,4-Tri methylhexane	۵۸۲/۲۰	۵۸۰/۸۹	-۰/۲۲
۱۸	3-Ethyl-2,3-dimethylpentane	۶۱۱/۸۰	۶۰۴/۶۱	-۱/۱۷
۱۹	2,3,3,4-Tetramethylpentane	۶۰۷/۶۰	۶۰۱/۲۸	-۱/۰۴
۲۰	5-Methylnonane	۶۱۳/۷۰	۶۱۴/۷۰	۰/۱۶
۲۱	2,4-Dimethyloctane	۶۰۴/۹۰	۶۰۷/۲۷	۰/۳۹
۲۲	3,3-Dimethyloctane	۶۱۴/۹۰	۶۱۵/۷۰	۰/۱۳
۲۳	4,4-Dimethyloctane	۶۱۴/۹۰	۶۱۵/۷۰	۰/۱۳
۲۴	3-Ethyl-3-methylheptane	۶۲۷/۴۰	۶۲۵/۸۰	-۰/۲۵
۲۵	4-Ethyl-3-methylheptane	۶۲۴/۲۰	۶۲۳/۲۸	-۰/۱۵
۲۶	2,2,4-Tri methylheptane	۵۹۸/۱۰	۶۰۱/۴۸	۰/۵۶
۲۷	2,4,6-Tri methylheptane	۵۹۶/۲۰	۵۹۹/۸۶	۰/۶۱
۲۸	3-Ethyl-2,2-dimethylhexane	۶۱۹/۶۰	۶۱۹/۵۷	۰/۰۰
۲۹	4-Ethyl-3,3-dimethylhexane	۶۳۲/۱۰	۶۴۴/۹۵	۲/۰۳
۳۰	3,3,4,4-Tetramethylhexane	۶۴۰/۹۰	۶۳۵/۷۶	-۰/۸۰
۳۱	2,2,3,3,4-pentamethylpentane	۶۳۶/۷۰	۶۳۲/۷۹	-۰/۶۱
سری پیش بینی				
۱	2-Methyl butane	۴۶۰/۳۹	۴۶۹/۰۵	۱/۸۸
۲	3-Methylpentane	۵۰۴/۴۰	۵۰۴/۹۰	۰/۱۰
۳	2,2-Dimethylpentane	۵۲۰/۴۰	۵۰۱/۷۲	-۳/۵۹
۴	4-Methylheptane	۵۶۱/۷۰	۵۶۸/۰۱	۱/۱۲
۵	2-Methyl-3-ethylpentane	۵۶۷/۰۰	۵۷۶/۴۶	۱/۶۷
۶	2,3,4-Trimethylpentane	۵۶۶/۳۰	۵۷۵/۳۹	۱/۶۱
۷	4-Ethylheptane	۵۹۴/۳۰	۵۹۰/۸	-۰/۵۹
۸	2,4-Dimethylheptane	۵۸۱/۰۰	۵۷۹/۸۶	-۰/۲۰
۹	3,4-Dimethylheptane	۵۹۵/۸۰	۵۹۱/۹۸	-۰/۶۴
۱۰	2,2,4-Tri methylhexane	۵۷۴/۲۰	۵۷۳/۷۸	-۰/۰۷
۱۱	3,3-Diethylpentane	۶۱۶/۰۰	۶۰۷/۹۸	-۱/۳۰
۱۲	2,2,3,3-Tetramethylpentane	۶۰۷/۶۰	۶۰۱/۲۸	-۱/۰۴
۱۳	2-Methylnonane	۶۰۹/۳۰	۶۱۱/۰۰	۰/۲۸
۱۴	4-Ethyloctane	۶۱۸/۲۰	۶۱۸/۴۳	۰/۰۴
۱۵	2,6-Dimethyloctane	۶۰۴/۹۰	۶۰۷/۲۷	۰/۳۹
۱۶	4-Propylheptane	۶۱۸/۲۰	۶۱۸/۴۳	۰/۰۴
۱۷	5-Ethyl-2-methylheptane	۶۰۹/۴۰	۶۱۱/۰۹	۰/۲۸
۱۸	2,3,4-Tri methylheptane	۶۲۱/۳۰	۶۲۰/۹۵	-۰/۰۶

جدول ۱. ۱

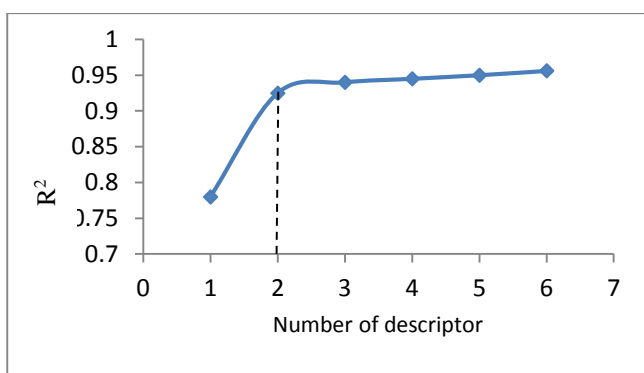
۱۹	2,4,4-Tri methylheptane	۶۰۶/۱۰	۶۰۸/۲۹	۰/۳۶
۲۰	3,3,5-Tri methylheptane	۶۰۹/۶۰	۶۱۱/۲۶	۰/۲۷
۲۱	3,4,5-Tri methylheptane	۶۲۵/۸۰	۶۲۴/۵۵	-۰/۲۰
۲۲	3-Ethyl-2,4-dimethylhexane	۶۲۵/۸۰	۶۲۴/۵۵	-۰/۲۰
۲۳	3-Ethyl-3,4-dimethylhexane	۶۴۰/۱۰	۶۳۵۰/۲	-۰/۷۷
۲۴	4-Ethyl-2,3-dimethylhexane	۶۲۵/۸۰	۶۲۴/۵۵	-۰/۲۰
۲۵	2,2,3,5,-Tetramethylhexane	۶۰۶/۴۰	۶۰۸/۵۵	۰/۳۵
۲۶	2,3,3,4,-Tetramethylhexane	۶۳۵/۹۰	۶۳۲/۲۱	-۰/۵۸
۲۷	3,3-Diethyl-2-Methylpentane	۶۴۸/۲۰	۶۴۰/۶	-۱/۱۷
۲۸	3-Ethyl-2,3,4-Trimethylpentane	۶۴۳/۹۰	۶۳۷/۸	-۰/۹۵
۲۹	n-Undecane	۶۳۹/۰۰	۶۴۷/۸۹	۱/۳۹

(a) مقادیر تجربی دمای بحرانی

(b) مقادیر پیش بینی شده توسط شبکه عصبی مصنوعی

(c) درصد خطای نسبی

استاندارد و پس از رسم مقادیر R و R² و SE بر حسب تعداد توصیف کننده‌ها بهترین مدل که دارای بیشترین مقدار R و F و کمترین مقدار خطای استاندارد و شامل توصیف کننده‌های تا حد امکان قابل توجیه باشد، به عنوان مدل نهایی برای ارتباط دمای بحرانی مولکول‌ها با ساختار آن‌ها انتخاب شد. با این روش مدل دوم با تعداد ۲ توصیف کننده به عنوان مناسبترین آنها انتخاب و توسط روش MLR مدل سازی و مورد ارزیابی قرار گرفت. فهرست توصیف کننده های انتخاب شده توسط نرم افزار SPSS به همراه توصیف مختصری از آنها در جدول ۲ آورده شده است.



شکل ۱- تاثیر تعداد توصیف کننده ها بر مقدار R².

۳. نتایج و بحث

۳-۱. مدل سازی با روش رگرسیون خطی چندگانه

برای ساختن مدلی که بیانگر ارتباط ساختاری ترکیبات مورد بررسی با دمای بحرانی آنها باشد، از روش رگرسیون خطی چندگانه (MLR) استفاده شد. در ابتدا توصیف کننده‌های محاسبه شده، به عنوان متغیرهای مستقل و مقادیر دمای بحرانی مولکول‌های مورد نظر به عنوان متغیرهای وابسته، به عنوان ورودی به نرم افزار SPSS وارد شدند. در نهایت با استفاده از منوی آنالیز، گزینه‌ی رگرسیون خطی و روش مرحله‌ای انتخاب و نهایتاً چندین مدل مختلف بطور جداگانه به دست آمد، که با توجه به خصوصیات آماری آنها از جمله ضریب رگرسیون (R)، آماره F و خطای

جدول ۲- توصیف کننده های انتخاب شده با SPSS و توصیف آنها.

توصیف کننده	نوع توصیف کنند	ضریب	MF	VIF
Polarizability	Electronic descriptors	۱۱/۶۹	-۱/۱۹۹	۳/۴۳
E GAP	Electronic descriptors	-۶۲/۹۹	۱/۱۹۹	۳/۴۰
Constant		۱۵۹۰/۱۸		

شد. سپس سرعت آموزش، مومتم و تعداد دورها نیز بهینه گردید. جدول ۳ مشخصات شبکه عصبی مصنوعی بهینه شده را نشان می‌دهد.

جدول ۳- ساختار و مشخصات ANN بهینه شده.

تعداد نرون‌ها در لایه مخفی	۵
سرعت آموزش	۰/۵
مومتم	۰/۳
تعداد دورها	۱۰۰۰

شبکه ساخته شده دارای دو نرون در لایه ورودی، ۵ نرون در لایه مخفی و یک نرون در لایه خروجی می‌باشد. بنابراین ساختار شبکه عصبی مصنوعی بهینه شده بصورت ۱-۵-۲ می‌باشد. باید توجه نمود که آموزش شبکه زمانی متوقف می‌شود که RMSE برای سری ارزیابی شروع به افزایش نماید زیرا Overfitting اتفاق می‌افتد و این اتفاق، قدرت پیشگویی شبکه را کاهش خواهد داد. برای جلوگیری از Overfitting در طول آموزش مقادیر RMSE بعد از هر ۵۰ سیکل برای سری آموزش و ارزیابی محاسبه و ثبت گردید. شکل ۲ مقادیر RMSE برای مجموعه‌های آموزشی و ارزیابی برحسب تعداد دورهای آموزش را نشان می‌دهد که بعد از ۱۰۰۰ سیکل مقادیر RMSE برای دو سری آموزشی و ارزیابی تغییر قابل ملاحظه‌ای نداشته است بنابراین آموزش شبکه در این نقطه متوقف می‌شود.

با استفاده از مدل ANN بهینه شده، مقادیر دمای بحرانی ترکیبات مورد نظر در مجموعه آموزش، ارزیابی و پیش‌بینی مورد محاسبه قرار گرفت که نتایج آن در جدول ۱ نشان داده شده است. شکل ۳ مقادیر محاسبه شده دمای بحرانی ترکیبات مورد نظر توسط شبکه عصبی مصنوعی در مجموعه‌های آموزش، ارزیابی و پیش‌بینی را برحسب مقادیر تجربی نشان می‌دهد. در این شکل، میزان نزدیکی داده‌ها به خط راست قدرت پیشگویی مدل را نشان می‌دهد. شکل ۴ مقادیر باقیمانده خطاها را نسبت به مقادیر تجربی نشان می‌دهد. میزان پراکندگی خطاها در اطراف محور، نشان دهنده این است که خطای سیستماتیک در مدل وجود ندارد.

شکل ۱ تاثیر تعداد توصیف کننده‌ها را بر مقدار R^2 نشان می‌دهد. همانطور که ملاحظه می‌شود، تغییرات R^2 بعد از ۲ توصیف کننده خیلی کم است. بنابراین ۲ توصیف کننده به عنوان تعداد بهینه توصیف کننده‌ها انتخاب گردید.

پس از انتخاب مناسب ترین توصیف کننده‌ها توسط روش مرحله ای با استفاده از SPSS، مرحله بعد، ایجاد مدل بین توصیف کننده های انتخاب شده و دمای بحرانی ترکیبات است. از نرم افزار SPSS برای این منظور استفاده گردید و بین توصیف کننده‌ها و دمای بحرانی ترکیبات سری آموزش با استفاده از روش MLR رابطه زیر به دست آمد:

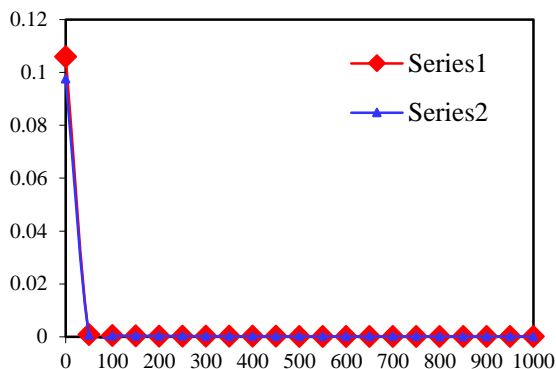
$$T_c = 1590.18 + 11.69(\text{Polarizability}) - 82.99(E_{\text{GAP}})$$

$$R^2_{\text{training}} = 0.936, R^2_{\text{test}} = 0.949$$

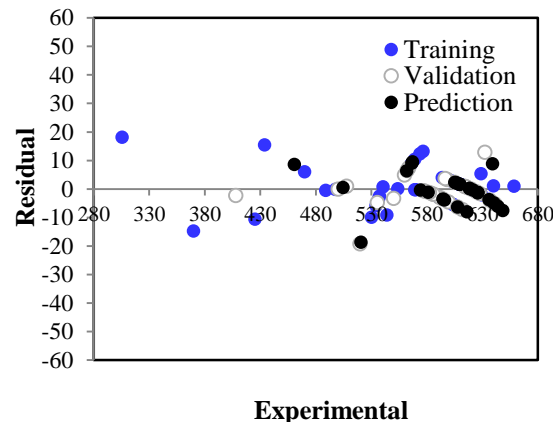
سپس از معادله به دست آمده برای پیش بینی دمای بحرانی سری پیش بینی استفاده گردید.

سپس برای حصول نتایج بهتر، از شبکه عصبی مصنوعی برای پیش بینی دمای بحرانی ترکیبات استفاده شد. برای این کار، ابتدا یک شبکه عصبی مصنوعی سه لایه با تابع انتقال سیگموئیدی و الگوریتم آموزشی لونیبرگ-مارکوات طراحی شد. مقادیر اولیه وزن ها بطور تصادفی از بازه [۰-۱] بوده و قبل از عمل آموزش مقادیر ورودی و خروجی در فاصله [۰/۹-۰/۱] نرمالیز گردید. مجموعه داده‌ها بطور تصادفی به سه سری آموزش، ارزیابی و پیشگویی تقسیم گردید. برای آموزش دادن شبکه از سری آموزش، برای ارزیابی شبکه، از سری ارزیابی و برای پیشگویی نتایج از سری پیشگویی استفاده شد. ترکیبات سری آموزش، ارزیابی و پیشگویی در جدول ۱ آورده شده است.

برای شبکه عصبی مصنوعی، ابتدا پارامترهایی مانند تعداد نرون‌ها در لایه مخفی، سرعت آموزش، مومتم و تعداد دورها بهینه گردید. تعداد نرون‌ها در لایه ورودی نیز برابر تعداد توصیف کننده‌های وارد شده به شبکه عصبی مصنوعی می‌باشد. برای بهینه سازی تعداد نرون‌ها در لایه مخفی، تعداد نرون‌ها از ۱ تا ۱۰ تغییر داده شد و مقدار متوسط مربعات خطا (RMSE) برای مجموعه‌های آموزش و ارزیابی محاسبه گردید و با توجه به مقدار RMSE برای سری ارزیابی برای ۱۰۰۰ دور، تعداد نرون‌ها در لایه مخفی بهینه



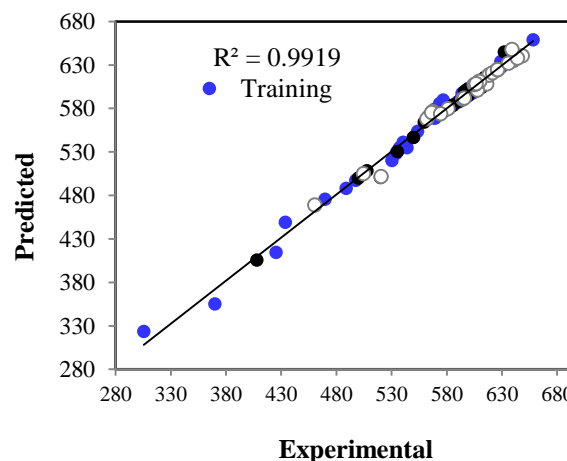
شکل ۴- نمودار تغییرات باقیمانده‌ها بر حسب مقادیر تجربی برای مقادیر Tc محاسبه شده بر اساس مدل SW-ANN در سه مجموعه آموزشی، ارزیابی و پیش‌بینی.



شکل ۵- مقادیر RMSE برای مجموعه‌های آموزشی و ارزیابی بر حسب تعداد چرخه‌های آموزش.

۴. نتیجه گیری

ارتباط کمی بین ویژگی و ساختار ترکیباتی از آلکانهای خطی با استفاده از روش انتخاب متغیر به صورت مرحله‌ای- رگرسیون خطی چندگانه (SW-MLR) و شبکه عصبی مصنوعی جهت پیش‌بینی دمای بحرانی مورد مطالعه قرار گرفت. در ابتدا توصیف‌کننده‌های مولکولی با کمک نرم افزار گوسین و هایپرکم محاسبه و توصیف‌کننده‌های مهم با استفاده از روش مرحله‌ای انتخاب و با شبکه عصبی مصنوعی مدل‌سازی و پیش‌بینی انجام شد. مدل SW-ANN با ۲ توصیف‌کننده انتخاب شده ساخته شد. پارامترهای آماری نشان می‌دهد که مدل SW-ANN توانایی پیش‌بینی دمای بحرانی مولکولها را با قدرت و توانایی بالا دارد.



شکل ۶- نمودار مقادیر دمای بحرانی محاسبه شده با کمک مدل SW-ANN برای مجموعه‌های آموزشی، ارزیابی و پیش‌بینی بر حسب مقادیر تجربی.

جدول ۲- انواع توصیف‌کننده‌های محاسبه شده برای ترکیبات.

توصیف‌کننده‌ها	نماد	اختصار	توصیف‌کننده‌ها	نماد	اختصار
توصیف‌کننده‌های مکانیک کوانتومی	Highest Occupied Molecular Orbital	HOMO	توصیف‌کننده‌های مکانیک کوانتومی	Hardness [$\eta=1/2 (HOMO+LUMO)$]	H
	Lowest Unoccupied Molecular Orbital	LUMO		Softness ($S=1/\eta$)	S
	difference between LUMO and HOMO	E GAP		Electro negativity [$\chi=-1/2 (HOMO-LUMO)$]	χ
	Molecular Polarizability	MP		El Electro philicity ($\omega=\chi^2 / 2 \eta$)	Ω
ویژگی‌های شیمیایی	Molecule surface area (Approx)	SA (A)	ویژگی‌های شیمیایی	Partition Coefficient	Log P
	Molecule surface area (Grid)	SA (G)		Hydration Energy	HE
	Mass	M		Refractivity	REF
	Molecule volume	V			

۵. مراجع

- [9] M. Shamsipur, A. Siroueinejad, B. Hemmateenejad, A. Abbaspour, H. Sharghi, K. Alizadeh, S. Arshadi, *J Electranal Chem.*, 600(2007) 345.
- [۱۰] م. منهاج، مبانی شبکه‌های عصبی، انتشارات دانشگاه صنعتی امیرکبیر، ایران، ۷۱۵ (۱۳۷۹).
- [11] R. Beals and T. Jackson, *Neural computing :An introduction*, IOP Publishing Ltd, NewYork, (1998).
- [12] D.T. Pham and X. Liu, *Neural networks for identification. predication and control*, Springer-Verlag London Limited. (1999).
- [13] J. Meiler, R. Meusinger, M. Will, *J Chem Inf Comput Sci.*, 40(2000) 1169.
- [14] A. Habibi-Yangjeh, M. Nooshyar, *Phys Chem Liq.*, 43(2005) 239.
- [15] A. Habibi-Yangjeh, M. Nooshyar, *Bull Korean Chem Soc.*, 26(2005) 139.
- [16] A. Habibi-Yangjeh, M. Danandeh-Jenagharad, M. Nooshyar, *Bull Korean Chem Soc.*, 26(2005) 2007.
- [17] A. Habibi-Yangjeh, *Phys Chem Liq.*, 45(2007) 471.
- [1] X.J. Yao, Y.W. Wang, X.Y. Zhang, R.S. Zhang, M.C. Liu, Z.D. Hu, B.T. Fan, *Chemom Intell Lab Syst.*, 62(2002) 217.
- [2] R. Guha, J.R. Serra, P.C. Jurs, *J Mol Graph Model.*, 23(2004) 1.
- [3] P. Krogsgaard-Larsen, T. Liljefors, U. Madsen, *Textbook of Drug Design and Discovery*. Taylor & Francis, London, (2002).
- [4] V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, *J Chem Inf Comput Sci.*, 42(2002) 693.
- [5] M. Karthikeyan, R.C. Glen, A. Bender, *J Chem Inf Model.*, 45(2005) 581.
- [6] A.A. Melnikov, V.A. Palyulin, N.S. Zefirov, *J Chem Inf Model.*, 47(2007) 2077.
- [7] S. Ajmani, S.C. Rogers, M.H. Barley, D.J. Livingstone, *J Chem Inf Model.*, 46(2006) 2043.
- [8] A.R. Katritzky, I.B. Stoyanova-Slavova, D.A. Dobchev, *J Mol Graph Model.*, 26(2007) 529.