مجله شیمی کوانتومی و اسپکتروسکوپی
۵:۱۳ (۱۳۹٤) ٤٥-٥٣
Journal of Quantum Chemistry and Spectroscopy (JQCS)

پیش بینی ضرایب توزیع برخی از ترکیبات شیمیایی بوسیله الگوریتم ژنتیک- رگرسیون خطی چند گانه

مهدی نکوئی*، بهزاد چهکندی، سید محسن بابازاده
دانشگاه آزاد اسلامی، واحد شاهرود، دانشکده علوم پایه، گروه شیمی، شاهرود، ایران

تاریخ ثبت اولیه: ۱۳۹۴/۱/۲۰، تاریخ دریافت نسخه اصلاح شده: ۱۳۹۴/۲/۱۱ ، تاریخ پذیرش قطعی: ۱۳۹۴/۲/۲۷

چکیده
پیش بینی ضرایب توزیع برخی از ترکیبات شیمیایی با استفاده از الگوریتم ژنتیک-رگرسیون خطی چندگانه مورد مطالعه قرار گرفت. با استفاده از نرم افزار دراگون تعداد زیادی توصیف کننده محاسبه و مناسب ترین و بهترین آنها که بیشترین ارتباط را با پارامتر مورد نظر(ضریب توزیع) داشت توسط الگوریتم ژنتیک انتخاب گردید. سپس با استفاده از روش خطی رگرسیون خطی چندگانه مدل‌سازی و پیش بینی ضرایب توزیع ترکیبات مختلف انجام شد. نتایج حاصل نشان می‌دهد که از این روش با درصد خطای کم می‌توان ضرایب توزیع ترکیبات را پیش بینی نمود.

واژه های کلیدی: ارتباط کمی ساختار خاصیت، ضریب توزیع، الگوریتم ژنتیک، رگرسیون خطی چندگانه.

۱. مقدمه
با افزایش فرآیندهای شیمیایی و کشف مواد شیمیایی جدید و در نتیجه افزایش روز افزون دامنه‌ی علم شیمی، نیاز به سیستم‌های پیچیده‌ای جهت جمع‌آوری و پردازش داده‌های شیمیایی می‌باشد که این کار در گرو صرف هزینه‌های بسیار و مطالعات گسترده است. برای رفع بخشی از این مشکلات روش‌های محاسباتی و کمومتریکس توسعه داده شده‌اند.

کمومتریکس یا شیمی سنجی اولین بار توسط والد دانشمند سوئدی در سال ۱۹۷۲ مطرح گردید و توسط کوالسکی توسعه یافت و این دو دانشمند برای نخستین بار در سال ۱۹۷۴ انجمن بین‌المللی کمومتریکس را بنا نهادند[۲-۱]. در حقیقت هدف از کمومتریکس، بهبود بخشیدن فرآیندهای اندازه‌گیری و استخراج اطلاعات شیمیایی مفیدتر از داده‌های اندازه‌گیری شده فیزیکی و شیمیایی می‌باشد. کمومتریکس در شاخه‌های مختلف علم شیمی به خصوص در علم شیمی تجزیه‌ای مورد استفاده قرار می‌گیرد و دارای کاربردهای بسیاری است که برخی از این کاربردها شامل

***عهده دار مکاتبات:** مهدی نکوئی
نشانی: شاهرود – دانشگاه آزاد اسلامی – دانشکده علوم – گروه شیمی
تلفن: ۰۲۳۳۲۳۹۴۲۸۹ **پست الکترونیک:** E-Mail: m_nekoei1356@yahoo.com

کنترل فرآیندها، تجزیه و تحلیل و شناخت الگوها، پردازش علائم، بهینه کردن شرایط، ارتباط کمی ساختار و فعالیت (QSAR) و ارتباط کمی ساختار ویژگی(QSPR) می‌باشد. یکی از زمینه‌های کاربرد کمومتریکس، مطالعات QSAR/QSPR می‌باشد که در این گونه بررسی‌ها، خواص مولکولها به ویژگی‌های ساختاری آنها نسبت داده می‌شوند. هنگامی که به ارتباط بین رفتار و ساختار مولکول دست یافتیم، می‌توان بطور هوشمند به طراحی مولکول‌های جدیدتر پرداخت و مکانیسم عملکرد مولکولها را مورد بررسی قرار داد. در واقع هدف از این گونه مطالعات، پیش‌بینی خواص بیولوژیکی، فیزیکی و شیمیایی یک ماده بر اساس ترکیب شیمیایی آن می باشد[۳].

QSAR/QSPR برای اولین بار در زمینه‌ی سم‌شناسی توسط کراس در سال ۱۸۶۳ ایجاد شد. پس از آن، محققین دیگری نیز به وجود روابط خطی بین فعالیت و ویژگی ترکیبات با ساختار آنها پی بردند. در این میان، هانش را که در سال ۱۹۶۲ توانست ارتباط بین فعالیت ترکیبات افزاینده‌ی سرعت رشد گیاهان با پارامترهای چربی دوستی آنها را نشان دهد، می‌توان پیشگام و پایه‌گذار روش QSAR نوین و به تبع آن QSPR دانست. روش‌های متعددی جهت ایجاد مدل‌های QSAR/QSPR وجود دارد. بطور کلی این روشها به دو دسته خطی و غیر خطی دسته‌بندی می‌شوند. در روش‌های خطی، ارتباط خطی بین توصیف کننده‌ها و فعالیت یا ویژگی را مورد بررسی قرار می‌دهیم، در حالی که در مواقعی که ویژگیهای مولکول‌ها از روابط غیرخطی پیروی کنند، از روشهای غیرخطی استفاده می‌کنیم. از متداولترین روشهای خطی می‌توان به روش رگرسیون خطی چند گانه(MLR) اشاره کرد. از روشهای غیر خطی مدلسازی نیز می‌توان به شبکه عصبی مصنوعی(ANN) و ماشین بردار پشتیبان (SVM) اشاره کرد[۱۰-۴].

به مجموعه ابزارها و روش‌های مورد استفاده در مطالعات QSPR/QSAR، روش‌های پارامتری گویند. به عبارتی دیگر روشهای پارامتری، مجموعه تکنیک‌هایی است که برای مدلسازی و پیش‌بینی فعالیت مولکولی بکار می‌رود. مراحل کلی مدلسازی به روش پارامتری به صورت زیر است:

۱- انتخاب سری داده‌ها

۲- محاسبه توصیف کننده‌های مولکولی

۳- تجزیه و تحلیل آماری توصیف کننده‌ها و حذف توصیف کننده‌های غیر ضروری

۴- روشهای مدلسازی

۵- تحلیل و ارزیابی اعتبار مدل‌های انتخاب شده.

هدف از این پروژه مدل‌سازی و پیش‌بینی ضرایب توزیع برخی از ترکیبات آلی با استفاده از الگوریتم ژنتیک به عنوان یک روش انتخاب متغیر و روش رگرسیون خطیMLR می‌باشد.

۲- محاسبات

۲-۱. انتخاب سری داده ها

در این کار پیش‌بینی ضرایب تقسیم تعداد ۶۱ ترکیب آلی توسط روشهای کمومتریکس مورد بررسی قرار گرفت. بدین منظور این ترکیبات به صورت تصادفی به دو گروه سری آموزش و سری پیش‌بینی تقسیم شدند، سری آموزش شامل ۴۹ مولکول (۸۰٪) و سری پیش‌بینی شامل ۱۲ مولکول (۲۰٪) می‌باشد. مقادیر ضرایب تقسیم به عنوان متغیر وابسته و توصیف کننده‌ها به عنوان متغیر مستقل انتخاب شدند. سری آموزش جهت ایجاد یک مدل مناسب و سری پیش‌بینی جهت ارزیابی مدل مورد استفاده قرار گرفت.

۲-۲. محاسبه توصیف کننده ها

در ابتدا برای محاسبه توصیف کننده ها، ساختار ترکیبات به کمک نرم افزار Hyperchem رسم شدند. سپس ساختارهای مولکولی رسم شده، به وسیله الگوریتم AM1 بهینه شدند. با استفاده از این نرم افزار می توان اطلاعات فراوانی نظیر زوایای پیوندی، طول پیوندها، زوایای پیچشی، بار اتم ها، انرژی تشکیل مولکول و... را بدست آورد. ساختارهای بهینه شده به نرم افزار دراگون منتقل و توصیف کننده ها به تعداد ۱۴۹۷ مورد به کمک این نرم افزار محاسبه شدند. نام گروه توصیف کننده‌های قابل محاسبه به همراه توصیف کننده و تعداد آنها که توسط نرم‌افزار Dragon محاسبه می‌شوند بطور کامل در جدول ۱ نشان داده شده‌اند.

جدول ۱. انواع توصیف کننده‌های قابل محاسبه توسط بسته نرم‌افزاری *Dragon* ویرایش ½

شماره	نام گروه توصیف کننده	بعد توصیف کننده	تعداد توصیف کننده‌ها
۱	توصیف کننده‌های ساختاری	۰	۴۷
۲	توصیف‌کننده‌های توپولوژیکی	۲	۲۶۶
۳	توصیف کننده‌های شمارنده گام مولکولی	۲	۲۱
۴	توصیف کنندههای BCUT	۲	۶۴
۵	اندیس‌های بار توپولوژیکی Galvez	۲	۲۱
۶	توصیف کننده‌های خود ارتباطی دو بعدی	۲	۹۶
۷	توصیف کننده‌های مربوط به بار	۳	۱۴
۸	اندیس‌های آروماتیکی	۳	۴
۹	پروفایل‌های مولکولی Randic	۳	۴۱
۱۰	توصیف کننده‌های هندسی	۳	۷۰
۱۱	توصیف کننده‌های RDF	۳	۱۵۰
۱۲	توصیف کننده‌های 3D-MORSE	۳	۱۶۰
۱۳	توصیف کننده‌های WHIM	۳	۹۹
۱۴	توصیف کننده‌های GETAWAY	۳	۱۹۷
۱۵	گروه‌های عاملی	۱	۱۲۱
۱۶	اجزای متمرکز اتمی	۱	۱۲۰
۱۷	توصیف‌کننده‌های تجربی	۱	۳
۱۸	ویژگی‌ها	۱	۳
مجموع توصیف کننده‌ها		۱۴۹۷	

۲-۳.کاهش تعداد توصیف کننده های نظری

یکی از مشکلاتی که در هنگام ایجاد مدلهای QSPR با آن مواجه می شویم، تعداد زیاد متغیرهای مستقل می باشد. در اغلب موارد تعداد توصیف کننده ها از تعداد مولکول ها بسیار بیشتر است. در این صورت استفاده از روشهای حداقل مربعات باعث ایجاد مشکلاتی نظیر انتخاب شانسی و همبستگی تصادفی می گردد. با توجه به این که بعضی از متغیرهای مستقل ثابت بوده و همچنین برخی دیگر با یکدیگر همبستگی نشان می دهند، لذا به روش زیر بعضی از متغیرها حذف شدند.

۱- توصیف کننده هایی که مقادیر ثابت و یا تقریبا ثابت دارند (بیش از ۹۰٪ داده های ثابت دارند)، حذف شدند. در این مرحله تعداد ۴۱۵ توصیف کننده حذف و بدین ترتیب ۱۰۸۲ توصیف کننده باقی ماند.

۲- توصیف کننده هایی که همبستگی بالای ۰/۹ با یکدیگر دارند مورد بررسی قرار گرفتند و بین آنها، توصیف کننده ای که همبستگی کمتری با متغیر مستقل داشت حذف گردید. بدین ترتیب تعداد ۶۵۷ توصیف کننده، کنار گذاشته شد و در نهایت تعداد ۴۲۵ توصیف کننده باقی ماند.

۲-٤. انتخاب توصیف‌کننده‌های موثر

مهم ترین بخش در ایجاد یک مدل کار آمد، انتخاب توصیف کننده های مناسب است. پس از محاسبه توصیف کننده های مختلف، تعدادی از آنها به عنوان توصیف کننده های مناسب برای ساخت مدل انتخاب می شوند. این مرحله شامل یافتن توصیف کننده های حاوی اطلاعات مفید است به طوری که قدرت پیش بینی مدل در سطح قابل قبولی باشد.

در این کار از روش الگوریتم ژنتیک جهت انتخاب مناسب ترین توصیف کننده ها استفاده شد. این الگوریتم از اصول انتخاب طبیعی داروین برای یافتن فرمول بهینه جهت پیش بینی یا تطبیق الگو استفاده می‌کند و اغلب گزینه خوبی برای تکنیک های پیش بینی بر مبنای رگرسیون است. الگوریتم ژنتیک برای حل مسائل بهینه کردن عددی، که حل کلاسیک آن مشکل است، مفید می‌باشد. برای حل مسئله به روش ژنتیک، ابتدا باید پاسخ مسئله را کدگذاری کرده، به گونه‌ای که در ادامه اجرای الگوریتم بتوان این پاسخ را مورد ارزیابی قرار داد و عملگرهای مختلف را بر آن عمل کرد. اجرای الگوریتم با ایجاد جمعیت اولیه شروع می‌گردد. هر عضو در جمعیت یک کروموزوم نامیده می‌شود که نمایانگر یک حل برای مساله موجود است. طی هر تکرار الگوریتم ژنتیک، مجموعه جدیدی از کروموزوم‌ها تولید می‌شوند. جمعیت در زمان معلوم را نسل می‌نامند. طی هر نسل، میزان برازش کروموزوم‌ها یا تابع برازش یک کروموزوم که با توجه به تابع هدف مسئله برآورد شده، تعیین می‌شود. طی فرایند بازتولید، عملگرهای ژنتیک یعنی عملگر ترکیب و عملگر جهش بر روی کروموزوم‌ها اعمال می‌شود. به کروموزوم‌هایی که از این طریق تولید می‌شوند نوزاد اطلاق می‌شود، سپس برازندگی نوزادان ارزیابی شده و به وسیله یکی از روش‌های انتخاب، کروموزوم‌های بهتر انتخاب و به نسل بعد منتقل می‌شوند. در این فرایند، الگوریتم به بهترین کروموزوم همگرا می‌شود که نمایانگر جواب بهینه یا زیر بهینه مساله است. کلیه پارامترهای استفاده شده برای الگوریتم ژنتیک در جدول ۲ آورده شده است. کاربرد الگوریتم ژنتیک در مدلسازی QSPR جستجو در میان توصیف کننده ها به منظور یافتن مؤثرترین توصیف کننده ها جهت انجام مدلسازی است. الگوریتم ژنتیک، یک روش جستجوی هوشمند و تصادفی است که با بکارگیری عملگرهای ژنتیک از یک فرایند تکامل تدریجی تبعیت می کند. از الگوریتم ژنتیک در شیمی بطور گسترده ای در زمینه مدل‌سازی QSPR و کمومتریکس استفاده شده است[۱۴-۱۲].

جدول ۲. پارامترهای بهینه شده برای الگوریتم ژنتیک.

مشخصات الگوریتم ژنتیک	
اندازه جمعیت	۱۵۰
(٪) جمعیت اولیه	۱۰
حداکثر تعداد نسل	۱۰۰
نرخ جهش	۰/۰۵
همگذری	دو تایی

۳-نتایج و بحث

۳-۱. ایجاد مدل با استفاده از MLR

پس از انتخاب مناسب‌ترین توصیف کننده‌ها توسط الگوریتم ژنتیک، مرحله بعدی، ایجاد مدل میان توصیف‌کننده‌های انتخاب شده و ضرایب توزیع ترکیبات می‌باشد. توصیف کننده های انتخاب شده بوسیله الگوریتم ژنتیک به همراه توصیف مختصری از آنها در جدول ۳ آورده شده است.

جدول۳.توصیف کننده های انتخاب شده با الگوریتم ژنتیک و توصیف آنها.

Descriptor	Chemical meaning	ME	VIF
Constant	Intercept	–	–
P	Polarizability	4.230	1.173
Ms	Mean electrotopological state(Constitutional descriptors)	-3.667	1.104
X index	Balaban X index (topological descriptors)	0.437	1.236

بین توصیف‌کننده‌ها و لگاریتم ضرایب توزیع آب-اکتانول برای سری آموزش با استفاده از روش MLR رابطه زیر بدست آمد:

$$\log K_{ow} = 3.048(\pm 0.377) + 0.254\ (\pm\ 0.008)\ \mathbf{P} - 1.934\ (\pm 0.183)\ \mathbf{Ms} + 0.403\ (\pm 0.113)\ \mathbf{Xindex}$$

سپس از معادله بدست آمده برای پیش‌بینی ضرایب توزیع سری تست استفاده گردید. مقادیر واقعی و پیش‌بینی شده لگاریتم ضرایب توزیع برای کلیه ترکیبات مجموعه آموزش و تست در جدول ۳ آورده شده است.

جهت ارزیابی اهمیت و میزان تأثیر توصیف کننده‌هایی که در مدل وارد شدند، اثر متوسط هر توصیف کننده به صورت زیر محاسبه شد.

$$ME_j = \frac{\beta_j \sum_{i=1}^{i=n} d_{ij}}{\sum_{j}^{m}\sum_{i}^{n} \beta_j d_{ij}}$$

در این معادله ME_j اثر متوسط توصیف کننده $B_{j,j}$ ضرایب توصیف کننده در معادله MLR، d_{ij} مقدار توصیف کننده j برای مولکول m,i تعداد توصیف کننده‌های وارد شده در مدل و n تعداد مولکولهاست. مقادیر ME برای هر توصیف کننده در جدول ۳ آورده شده است.

۳-۲. ارزیابی توصیف کننده های انتخاب شده

به منظور ارزیابی توصیف کننده های انتخاب شده مبنی بر مستقل بودن از همدیگر در جدول ۴ ضرایب همبستگی توصیف کننده های انتخاب شده نسبت به یکدیگر آورده شده است. همانطور که در جدول زیر مشاهده می‌شود بیشترین ضریب همبستگی بین توصیف کننده Ms و توصیف کننده Xindex با مقدار ضریب همبستگی ۰/۳۴- می‌باشد و بقیه مقادیر کمتری را نشان می‌دهند. این نتایج نشان می‌دهد که بین توصیف کننده‌های انتخاب شده همبستگی چندانی وجود نداشته و توصیف کننده‌ها مستقل از هم هستند.

جدول٤. ماتریس ضرایب همبستگی توصیف کننده های انتخاب شده.

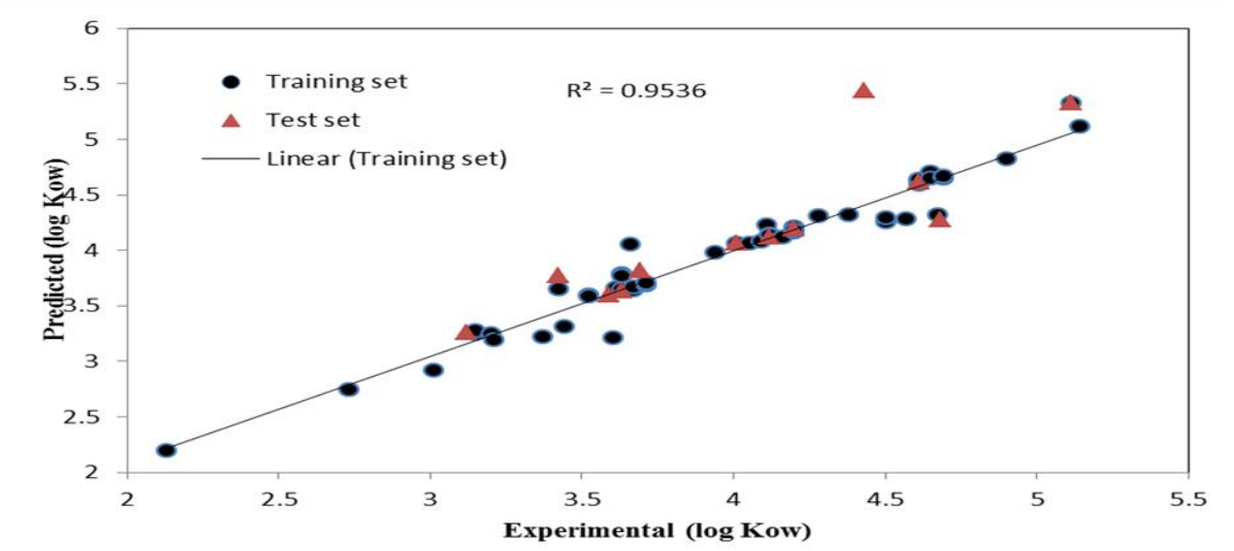
	p	Ms	Xindex
p	1		
Ms	0.26	1	
Xindex	-0.29	-0.34	1

جدول ۵. مقادیر تجربی و محاسبه شده ضرایب توزیع برای ترکیبات مختلف برای مجموعه های اموزشی و پیش بینی در مدلGA-MLR.

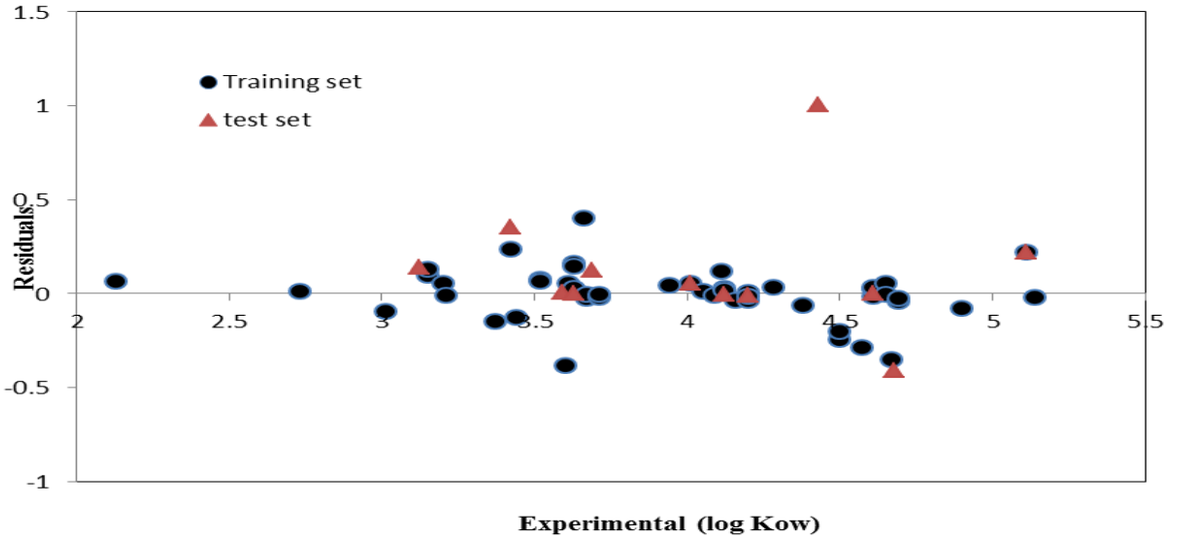
No	Compounds	Exp	GA-MLR
Train			
1	Benzene	2.13	2.2
2	Toluene	2.73	2.74
3	Cyclopentane	3.01	2.92
4	p-Xylen	3.15	3.25
5	Ethylbenzene	3.15	3.28
6	m-Xylen	3.2	3.26
7	2-Methylpentane	3.21	3.2
8	Methylcyclopentane	3.37	3.23
9	2,3-Dimethylbutane	3.42	3.66
10	Cyclohexane	3.44	3.31
11	trans-1,2-Dimethylcyclopentane	3.52	3.6
12	cis-1,3-Dimethylcyclopentane	3.52	3.59
13	3-Methylpentane	3.6	3.22
14	Methylcyclohexane	3.61	3.67
15	1-Methyl-4-Ethylbenzene	3.63	3.79
16	1,2,4-Trimethylbenzene	3.63	3.78
17	3-Methylhexane	3.63	3.66
18	Isopropylbenzene	3.66	4.06
19	1-Ethyl-1-methylcyclopentane	3.67	3.64
20	3,3-Dimethylpentane	3.67	3.67
21	2-Methylhexane	3.71	3.69
22	3,3-Dimethylpentane	3.71	3.71

23	cis-trans-cis-1,2,4-	3.94	3.99
24	2,2-Dimethylpentane	4.01	4.07
25	trans-1,2-Dimethylcyclohexane	4.05	4.06
26	1,3,5-Triethylbenzene	4.09	4.08
27	tert-Butylbenzene	4.11	4.23
28	3-Methylheptane	4.12	4.15
29	2,3-Dimethylhexane	4.12	4.14
30	2,5-Dimethylhexane	4.16	4.12
31	2,2-Dimethylhexane	4.2	4.16
32	3-Ethylhexane	4.2	4.21
33	3-Methylheptane	4.2	4.18
34	1,3-Dimethyl-2-Ethylbenzene	4.28	4.31
35	n-Butylbenzene	4.38	4.32
36	1-Methyl-3-Isopropylbenzene	4.5	4.26
37	1,2-Dimethyl-4-Ethylbenzene	4.5	4.3
38	sec-Butylbenzene	4.57	4.29
39	2,3-Dimethylheptane	4.61	4.59
40	3,5-Dimethylheptane	4.61	4.63
41	3,4-Dimethylheptane	4.61	4.65
42	3,3-Diethylpentane	4.65	4.71
43	3,3-Dimethylpentane	4.65	4.65
44	1-Methyl-3-n-Propylbenzene	4.67	4.32
45	2-Methyloctane	4.69	4.65
46	3-Methyloctane	4.69	4.67
47	n-Pentylbenzene	4.9	4.83
48	1,2,4-Triethylbenzene	5.11	5.33
49	3,3-Dimethyloctane	5.14	5.12
Test			
50	o-Xylene	3.12	3.26
51	1,3,5-Trimethylbenzene	3.42	3.77
52	2,2,3-Trimethylbutane	3.59	3.6
53	2,4-Dimethylpentane	3.63	3.63
54	n-Propylbenzene	3.69	3.83
55	trans-1,2-Dimethylcyclohexane	4.01	4.07
56	2,5-Dimethylhexane	4.12	4.12
57	4-Methylheptane	4.2	4.19
58	2-Methylbutylbenzene	4.43	5.44
59	1,3,5-Triethylbenzene	5.11	5.33
60	Isopropylbenzene	4.68	4.27
61	2,5-Dimethylheptane	4.61	4.61

همچنین مقادیر ضرایب توزیع محاسبه شده و تجربی برای ترکیبات براساس مدل MLR در دو مجموعه آموزشی و تست در شکل (۱) آورده شده است. همینطور در شکل (۲) 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. مقادیر ضرایب توزیع محاسبه شده براساس مدل MLR در دو مجموعه آموزشی و تست بر حسب مقادیر تجربی.



شکل ۲. نمودار تغییرات باقیمانده‌ها بر حسب مقادیر تجربی برای مقادیر ضرایب توزیع محاسبه شده براساس مدل *MLR* در دو مجموعه آموزشی و تست.

٤. نتیجه گیری

الگوریتم ژنتیک از اصول انتخاب طبیعی داروین برای یافتن فرمول بهینه جهت پیش‌بینی یا تطبیق الگو استفاده می کنند. الگوریتم‌های ژنتیک اغلب گزینه خوبی برای تکنیک‌های پیش‌بینی بر مبنای رگرسیون هستند و همینطور به عنوان یک الگوریتم بهینه جهت انتخاب بهترین توصیف

کننده ها در روشهای QSPR استفاده می شود. به همین جهت از این الگوریتم بدین منظور استفاده گردید. پارامترهای آماری مختلف نشان داد که روش GA-MLR قدرت پیش‌بینی خوبی برای ضرایب توزیع ترکیبات مورد مطالعه نشان می‌دهد. ($R^2_{test} = 0.953$, $RMSE_{test} = 0.449$)

۵. مراجع

[1] D.L. Massart, *Handbook of Chemmometrics*. Part A. (**1998**).
[2] P.p. Roy. S. Poai, I Miun, k. Roy, *molecuies.*, 14 (**2009**) 1000.
[3] A.R. Katritzky, V.S. Lobanov and M. Karelson, *Chem. Soc. Rev*., 24(4) (**1995**) 279.
[4] V.N. Vapnik, S. Golowich, A.J. Smola, *Adv. Neural inf. process. Syst*, 9 (**1997**) 281.
[5] F. Gharagheizi, *Computational materials science*, 40 (**2007**) 159.
[6] J. Ghasemi, S. Saaidpour and S.D. Brown, *Journal of Molecular Structure: THEOCHEM*, 805(1) (**2007**) 27.
[7] J. Xu, L. Wang, L. Wang, X. Shen and W. Xu, *Journal of computational chemistry*, 32(15) (**2011**) 3241.
[8] Y. Pan, J. Jiang, R. Wang and H. Cao, *Chemometrics and Intelligent Laboratory Systems*, 92(2), (**2008**) 169.
[9] X.J. Yao, A. Panaye, J.P. Doucet, R.S. Zhang, H.F. Chen, M.C. Liu, Z.D. Hu and B.T. Fan, *Journal of chemical information and computer sciences*, 44(4) (**2004**) 1257.
[10] M.H. Fatemi and S. Gharaghani, *Bioorganic & medicinal chemistry*, 15(24) (**2007**)7746.
[11] S. Riahi, M.R. Ganjali, P. Norouzi and F. Jafari, *Sensors and Actuators B: Chemical*, 132(1) (**2008**) 13.
[12] G. Liang, J. Xu and L. Liu, *Fluid Phase Equilibria*, 353 (**2013**)15.
[13] A.K. Saxena and P. Prathipati, *SAR and QSAR in Environmental Research*, *14*(5-6) (**2003**) 433.