

DOI: 10.13718/j.cnki.xdzk.2024.10.006

堵锡华, 陈艳, 李靖, 等. 新的原子指数用于松茸和姬松茸 QSRR 研究 [J]. 西南大学学报(自然科学版), 2024, 46(10): 80-89.

新的原子指数用于松茸和姬松茸 QSRR 研究

堵锡华, 陈艳, 李靖, 田林, 徐艳

徐州工程学院 材料与化学工程学院, 江苏 徐州 221018

摘要: 松茸和姬松茸含有多种活性成分, 不但具有较高的营养价值, 还具有较好的保健功效. 为构建松茸和姬松茸挥发性成分保留时间神经网络的定量结构-保留关系模型, 根据松茸和姬松茸挥发性化合物中原子的特性和连接状态, 提出了一种新的分子结构参数(原子指数 A), 并计算了松茸和姬松茸挥发性成分的电性拓扑状态指数(E_m)和电性距离矢量(M_m). 选取电性拓扑状态指数的 E_2 和电性距离矢量的 M_{15} 和 M_{21} 等 3 种指数, 结合原子指数 A 共 4 种结构参数, 作为 BP 神经网络法的输入层变量, 松茸挥发性成分的色谱保留时间作为输出层变量, 采用 4-8-1 的网络结构, 构建的模型总相关系数 r 为 0.991 9, 根据该模型计算得到的保留时间预测值与文献实验值的平均误差为 0.57; 同样采用这 4 种参数, 与姬松茸挥发性成分的保留时间进行研究, 采用 4-7-1 的网络结构, 所得模型总相关系数 r 为 0.992 9, 保留时间预测值与实验值的平均误差为 0.58. 结果表明: $-\text{CH}_3$, $-\text{CH}_2-$ 和氧氮杂原子等基团片段及连接方式是影响松茸和姬松茸挥发性成分色谱保留时间的主要因素.

关键词: 松茸; 姬松茸; 挥发性成分; 保留时间; 神经网络

中图分类号: TS262

文献标志码: A

文章编号: 1673-9868(2024)10-0080-10

开放科学(资源服务)标识码(OSID):



Quantitative Structure-Retention Relationship Study of Volatile Components in *Tricholoma matsutake* and *Agaricus blazei* Using New Atomic Index

DU Xihua, CHEN Yan, LI Jing, TIAN Lin, XU Yan

School of Materials and Chemical Engineering, Xuzhou University of Technology, Xuzhou Jiangsu 221018, China

Abstract: *Tricholoma matsutake* and *Agaricus blazei* contain a variety of active ingredients, which not only have high nutritional value, but also good for health. In order to establish neural network model of quantitative structure-retention relationship (QSRR) for the chromatographic retention time of volatile components of *T. matsutake* and *A. blazei*, a novel molecular structure parameter-atomic index A was derived based on characteristics and connectivity of the atoms in volatile components molecules. The elec-

收稿日期: 2023-02-12

基金项目: 国家自然科学基金项目(21472071); 江苏省自然科学基金项目(BK20171168); 江苏省高校自然科学基金重大项目(18KJA430015).

作者简介: 堵锡华, 教授, 主要从事化合物构效学研究.

trical topology state index (E_m) and the electric distance vector (M_m) of volatile components of the two substances were calculated. Three configuration parameters, including the electrical topology state indices E_2 and the electric distance vector M_{15} , M_{21} were selected. Then, the four parameters were used as input layer variables and the chromatographic retention time of *T. matsutake* was used as the output layer variable. Besides, the 4-8-1 network structure was adopted and BP neural network method was used to establish a neural network model. The total correlation coefficient r was 0.991 9. The mean error between the predicted values and the experimental values of retention time was 0.57. Using the same method, the five configuration parameters were selected. The 4-7-1 network structure was adopted and BP neural network method was used to establish a model to study chromatographic retention time of volatile components of *A. blazei*. The total correlation coefficient r was 0.992 9. The mean error between the predicted values and the experimental values of retention time was 0.58. The results showed that fragments of $-\text{CH}_3$, $-\text{CH}_2-$ and oxynitrogen heteroatoms groups and the way of connection were the main factors affecting the retention time of volatile components in *T. matsutake* and *A. blazei*. This study provides theoretical guidance for exploring the chromatographic retention mechanism of volatile components in *T. matsutake* and *A. blazei*.

Key words: *Tricholoma matsutake*; *Agaricus blazei*; volatile constituent; retention time; neural network

松茸是一种天然名贵的食药两用菌类, 因其营养丰富、味道鲜美, 历来被誉为“菌中之王”^[1], 由于它含有丰富的氨基酸和多种人体必需的矿物质, 还存在具有保健作用的活性成分, 因此具有降血脂、抗氧化、抗癌、改善心血管疾病等保健功能^[2-3], 还可起到提高免疫力和美白的作用^[4]. 姬松茸原产于巴西, 后传入我国, 它同样具有良好的食药两用营养价值, 也富含多种活性成分^[5], 具有增强免疫、抗氧化、抗突变和抗疲劳等功效^[6], 而且抗癌效果明显, 故被广泛应用于癌症和白血病等疾病的辅助治疗^[7]. 正是由于松茸与姬松茸具备的药食两用功效, 近年来受到国内外营养学家及药学研究工作者的广泛关注. 魏晓梅等^[8]研究了松茸多糖的提取方法及测定; 申秀娟等^[9]则利用醇解姬松茸中的提取物, 开展了姬松茸多糖对小鼠免疫和肠道菌群影响的研究; 石芳等^[10]开展了采用不同的干燥方式, 对松茸的营养成分、氨基酸及挥发性成分影响的研究; Zhai 等^[11]就松茸含有的赖氨酸等营养及抗氧化成分对人体作用展开了研究; 此外还有一些侧重松茸活性酶有关性质的研究^[12-13], 以及松茸挥发性成分的检测分析^[14]. 所有的这些研究虽然较多, 但普遍都是单纯通过实验进行的研究, 耗时长、过程复杂. 先理论研究再指导实验已成为科学研究中一种简便而高效的方法, 这一方法应用于松茸和姬松茸的研究目前还较为少见^[15]. 近年来人工神经网络(Artificial neural network)方法在众多领域得到广泛应用^[16-20], 本文采用人工神经网络法中常用的 BP 算法^[21-22], 对文献[23]所列松茸和姬松茸 145 种挥发性成分(其中 1 个分子重复, 实际 144 个分子), 通过定义一种新的分子结构指数(原子指数 A), 结合电性拓扑状态指数、电性距离矢量与色谱保留时间进行构效关系(QSRR)分析, 构建了 BP 神经网络模型, 预测所得保留时间与文献实验值较为吻合. 利用该神经网络模型可快速获取挥发性成分的保留值, 研究结果对提高定性分析能力、探索色谱保留机制具有积极的意义.

1 数据处理及研究方法

1.1 数据来源

松茸和姬松茸挥发性成分及色谱保留时间来源于张宪臣等^[23]的分析测试结果, 其中松茸包含了酯类、酸类等 9 类 76 种挥发性成分(表 1), 姬松茸包含了醛类、醇类等 9 类 68 种挥发性成分(表 2).

1.2 原子指数 A 的定义与建构

不同原子因电子层数、价电子数、电负性等性质及连接的氢原子数有所不同, 故导致构成分子的特性也各不相同. 通过考察文献[23]中 76 种松茸和 68 种姬松茸挥发性成分的保留时间与其分子结构之间的关系, 新定义一种用于表征原子特性的特征值(T_i)为:

$$T_i = (m_i - h_i) \cdot [(\chi_c + 1) / (\chi_i + 1)] \quad (1)$$

式中: m_i , χ_i 分别为原子 i 的价电子数、鲍林电负性; χ_c 为碳原子鲍林电负性; h_i 为与 i 原子直接连接的氢原子数, 如基团=O, 按式(1)计算氧原子 T_o 为 4.797; 如基团—OH, 则氧原子 T_o 为 3.997. 在邻接矩阵基础上, 建构化合物新的分子结构参数—原子指数 A 为:

$$A = \sum (T_i)^{0.5} \quad (2)$$

式中: \sum 表示每一个非氢原子对分子的贡献加和, 由于考虑的是每个原子对分子的贡献, 故将其命名为原子指数, 76 种松茸和 68 种姬松茸挥发性分子的原子指数 A 分别列于表 1 和表 2 中.

1.3 电性拓扑状态指数及电性距离矢量的计算及回归分析

根据文献[23]所列的 76 种松茸和 68 种姬松茸挥发性成分分子, 先用建构的原子指数 A 与这些分子的色谱保留时间进行回归分析, 对松茸所得一元方程的相关系数为 0.845, 对姬松茸所得一元方程的相关系数为 0.902, 基本能达到回归分析要求, 但并不理想. 考虑到这些挥发性分子的类型比较复杂, 只用一种指数显然不能满足准确预测的要求, 为此使用文献[24]的方法自编程序, 计算这 144 个分子的电性拓扑状态指数和电性距离矢量, 结合建构的原子指数 A , 先对松茸挥发性成分的色谱保留时间进行回归分析, 优化筛选出电性拓扑状态指数的 E_2 和电性距离矢量的 M_{15} 和 M_{21} 进行回归分析(表 1), 建立的四元线性方程为:

$$t_r = 1.108 A + 0.487 E_2 + 0.920 M_{15} - 0.162 M_{21} - 6.129 \quad (3)$$

$$n = 76 \quad r = 0.883 \quad r_{\text{adj}}^2 = 0.768 \quad S = 3.377 \quad F = 63.137$$

式中: n 为分子样本数; r , r_{adj}^2 , S , F 分别为所建模型的相关系数、调整的可决系数、标准误差、Fischer 检验值, 可以得出四元回归模型的相关系数为 0.883, 根据该模型计算得到的保留时间见表 1 中预测值 1.

同理用该 4 种参数对姬松茸挥发性成分的色谱保留时间进行回归分析(表 2), 建立的四元线性方程为:

$$t_r = 1.361 A + 0.509 E_2 + 0.450 M_{15} + 0.098 M_{21} - 9.949 \quad (4)$$

$$n = 68 \quad r = 0.916 \quad r_{\text{adj}}^2 = 0.829 \quad S = 3.256 \quad F = 82.005$$

可以得出四元回归模型的相关系数为 0.916, 根据该模型计算得到的保留时间见表 2 中预测值 1. 不难看出, 根据多元回归分析方程式(3)和式(4)得到的预测值与实验值还存在较大的误差.

表 1 松茸挥发性成分的分子结构参数和保留时间

序号	化合物	A	E_2	M_{15}	M_{21}	保留时间			
						实验值 ^[23]	预测值 1	预测值 2	误差
1	三甲胺	5.096	0	0	0	1.99	-0.48	2.43	0.44
2	乙酸甲酯	8.381	0	0	0	2.15	3.16	2.54	0.39
3	醋酸	7.159	0	0	0	2.50	1.80	1.74	-0.76
4	丙酸甲酯	9.795	0.469	-0.077	0.264	2.51	4.84	3.62	1.11
5	异戊醛	9.069	0.694	-0.369	6.907	2.76	2.80	3.11	0.35
6	2-甲基丁醛	9.069	0.962	-1.647	7.711	3.09	1.62	3.16	0.07
7	异丁酸甲酯	11.113	0	0	0	3.33	6.18	3.61	0.28
8	戊醛	9.165	2.902	0	8.042	3.48	4.14	3.09	-0.39
9	丁酸甲酯	11.209	1.406	-0.291	1.467	3.84	6.47	4.20	0.36
10	异戊醇	8.529	1.262	-0.432	2.961	4.07	3.06	4.71	0.64
11	2-甲基丁醇	8.529	1.410	-0.166	2.474	4.14	3.45	3.47	-0.67
12	2-己酮	10.433	2.936	-0.449	2.256	4.24	6.08	4.30	0.06
13	异丁酸	9.891	0	0	0	4.63	4.83	3.51	-1.12
14	异戊酸甲酯	12.527	0.524	-0.582	1.040	4.75	7.30	5.45	0.70

续表 1

序号	化合物	A	E_2	M_{15}	M_{21}	保留时间			
						实验值 ^[23]	预测值 1	预测值 2	误差
15	丁酸	9.987	1.023	-0.552	1.778	5.12	4.64	4.54	-0.58
16	正己醛	10.579	4.209	0	8.791	5.20	6.22	5.82	0.62
17	2-甲基庚酸	14.962	4.178	-1.712	2.970	5.70	10.43	6.12	0.42
18	戊酸甲酯	12.623	2.550	-0.365	2.286	5.71	8.39	5.79	0.08
19	2,4-二甲基-1-庚烯	12.389	3.842	-0.055	0	6.00	9.42	6.96	0.96
20	异戊酸	11.305	0.278	-0.742	1.213	6.49	5.65	6.51	0.02
21	2-甲基丁酸	11.305	0.718	-0.706	0.761	6.69	5.97	6.67	-0.02
22	正己醇	10.039	5.043	0	4.267	6.71	6.76	6.74	0.03
23	5-甲基-2-己酮	11.751	1.785	-0.939	1.925	7.24	6.59	6.59	-0.65
24	4-甲基戊酸甲酯	13.941	1.473	-0.875	1.947	7.24	8.91	7.50	0.26
25	2-乙基-2-己烯醛	13.897	3.017	4.359	9.593	7.31	13.19	9.30	1.99
26	正戊酸	11.401	2.082	-0.762	2.920	7.41	6.34	7.12	-0.29
27	环戊基醋酸酯	15.451	4.578	-0.357	2.075	7.50	12.56	6.00	-1.50
28	2,6-二甲基吡嗪	13.656	0	0.988	0	7.83	9.91	9.24	1.41
29	γ -丁内酯	10.623	2.178	-0.316	2.867	8.07	5.95	7.98	-0.09
30	己酸甲酯	14.037	3.805	-0.399	2.840	8.18	10.45	7.76	-0.42
31	呋喃酮	16.081	0	0	0	8.54	11.69	9.84	1.30
32	十四碳烯酸	24.765	12.726	-2.676	9.398	8.98	23.52	9.06	0.08
33	戊内酯	11.941	3.323	-0.330	3.535	9.22	7.84	10.72	1.50
34	苯甲醛	14.583	0	5.924	9.132	9.45	14.00	9.51	0.06
35	3-甲基壬烷	13.217	8.476	0.222	0	9.56	12.85	9.69	0.13
36	1-辛烯-3-酮	13.579	4.017	-0.136	4.717	9.92	9.98	11.60	1.68
37	1-辛烯-3-醇	13.357	4.413	-3.838	3.964	10.03	6.65	9.96	-0.07
38	3-辛酮	13.261	5.000	-0.048	3.590	10.18	10.37	10.56	0.38
39	乙酰丙酸甲酯	15.399	0.485	-0.448	1.626	10.35	10.49	9.42	-0.93
40	2-正戊基呋喃	16.043	4.937	2.524	0	10.35	16.37	10.27	-0.08
41	3-辛醇	12.771	5.590	-1.065	3.595	10.53	9.18	10.16	-0.37
42	己酸	12.815	3.277	-0.871	3.697	10.67	8.27	9.98	-0.69
43	2,3,5-三甲基吡嗪	14.924	0	0.843	0	10.76	11.18	9.87	-0.89
44	正辛醛	13.675	6.962	0	9.692	10.78	10.84	11.22	0.44
45	2-乙基-2-己烯醛	13.897	3.017	4.359	9.593	10.97	13.19	9.30	-1.67
46	庚酸甲酯	15.451	5.126	-0.418	3.232	11.56	12.58	10.90	-0.66
47	2-甲烯基丁内酯	12.623	1.233	0.630	0.969	12.11	8.88	9.78	-2.33
48	苯乙醛	15.997	0.529	4.001	9.104	12.40	14.06	12.49	0.09
49	羟基香茅醛	18.279	3.489	-0.597	11.572	12.71	13.40	12.64	-0.07
50	反-2-辛烯醛	14.043	4.769	0	9.665	12.91	10.19	12.77	-0.14
51	间甲酚	13.896	0	2.944	5.784	13.13	11.04	13.15	0.02
52	正辛醇	12.868	7.863	0	4.698	13.34	11.20	12.73	-0.61
53	反式-2-辛烯-1-醇	13.503	5.135	0	3.830	13.46	10.71	13.33	-0.13
54	3-乙基-2,5-二甲基吡嗪	16.339	0.971	1.543	0	13.72	13.87	13.90	0.18

续表 1

序号	化合物	A	E_2	M_{15}	M_{21}	保留时间			
						实验值 ^[23]	预测值 1	预测值 2	误差
55	2-壬酮	14.675	6.994	-0.536	3.460	14.21	12.48	14.91	0.70
56	十一烷	14.728	12.950	0	0	14.44	16.50	14.43	-0.01
57	芳樟醇	14.675	1.670	2.169	3.955	14.65	12.30	14.62	-0.03
58	壬醛	14.822	8.379	0	9.982	14.74	12.76	14.34	-0.40
59	辛酸甲酯	16.866	6.489	-0.430	3.521	15.57	14.75	16.42	0.85
60	苯乙酸甲酯	19.455	0.358	3.781	4.097	17.93	18.42	17.94	0.01
61	十一醇	17.110	12.210	0	5.024	18.34	17.96	18.39	0.05
62	α -松油醇	16.675	3.357	6.684	2.335	18.56	19.75	18.51	-0.06
63	癸醛	16.236	9.812	0	10.210	19.04	14.98	19.43	0.39
64	壬烯酸甲酯	18.916	5.203	-0.597	4.237	19.74	16.13	20.15	0.41
65	壬酸甲酯	18.280	7.881	-0.438	3.742	19.86	16.95	20.71	0.85
66	2-丁基-2-辛烯醛	19.554	8.089	6.501	11.883	20.34	23.53	23.43	3.09
67	乙酸芳樟酯	22.087	1.680	1.674	2.066	21.18	20.37	21.14	-0.04
68	3-苯丙酸甲酯	20.869	1.216	5.427	3.961	22.17	21.94	22.10	-0.07
69	甲基壬基甲酮	17.504	9.842	-0.555	3.845	22.88	16.92	22.46	-0.42
70	2,5-二甲基-3-(3-甲基丁基)吡嗪	20.485	2.245	1.971	0	23.84	19.48	24.05	0.21
71	癸酸甲酯	19.694	9.293	-0.444	3.916	24.15	19.17	24.28	0.13
72	椰子醛	17.598	6.349	-0.209	3.951	26.17	15.63	24.35	-1.82
73	2-丁基-2-辛烯醛	19.554	8.089	6.501	11.883	26.48	23.53	23.43	-3.05
74	肉桂酸甲酯	21.505	0	7.682	6.071	27.10	23.78	27.24	0.14
75	反式-肉桂酸甲酯	21.505	0	7.682	6.071	27.43	23.78	27.24	-0.19
76	雪松烯	22.681	7.093	6.134	0	28.16	28.10	28.12	-0.04

表 2 姬松茸挥发性成分的分子结构参数和保留时间

序号	化合物	A	E_2	M_{15}	M_{21}	保留时间			
						实验值 ^[23]	预测值 1	预测值 2	误差
1	乙酰肼	8.689	0	0	0	2.27	1.88	3.33	1.06
2	醋酸	7.159	0	0	0	2.51	-0.21	3.31	0.80
3	丙酸甲酯	9.795	0.469	-0.077	0.264	2.51	3.61	3.59	1.08
4	异丁醇	7.115	0.306	-0.208	1.765	2.71	-0.03	3.29	0.58
5	异戊醛	9.069	0.694	-0.369	6.907	2.76	3.26	3.29	0.53
6	2-甲基丁醛	9.069	0.962	-1.647	7.711	3.08	2.90	2.74	-0.34
7	异丁酸甲酯	11.113	0	0	0	3.33	5.18	3.83	0.50
8	丙酸	8.573	0.222	-0.054	0.130	3.61	1.82	3.48	-0.13
9	丁酸甲酯	11.209	1.406	-0.291	1.467	3.84	6.03	4.50	0.66
10	异戊醇	8.529	1.262	-0.432	2.961	4.02	2.40	5.12	1.10
11	异丁酸	9.891	0	0	0	4.55	3.51	3.34	-1.21
12	环庚三烯	11.807	1.076	0	0	4.59	6.67	4.25	-0.34
13	1-戊醇	8.625	3.682	0	3.899	4.63	4.05	3.66	-0.97

续表 2

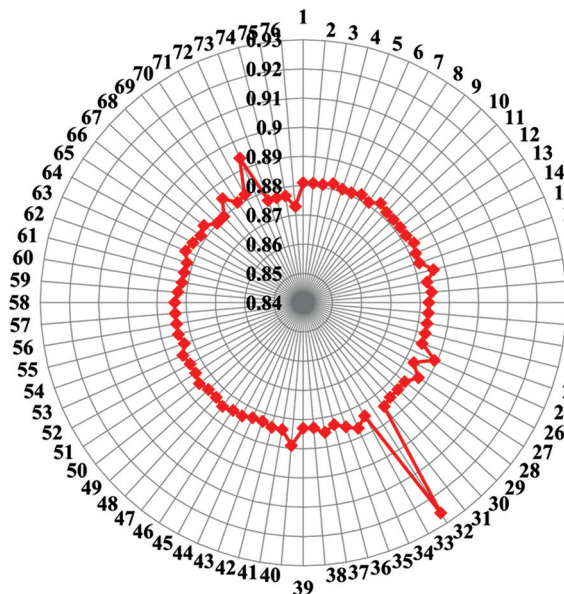
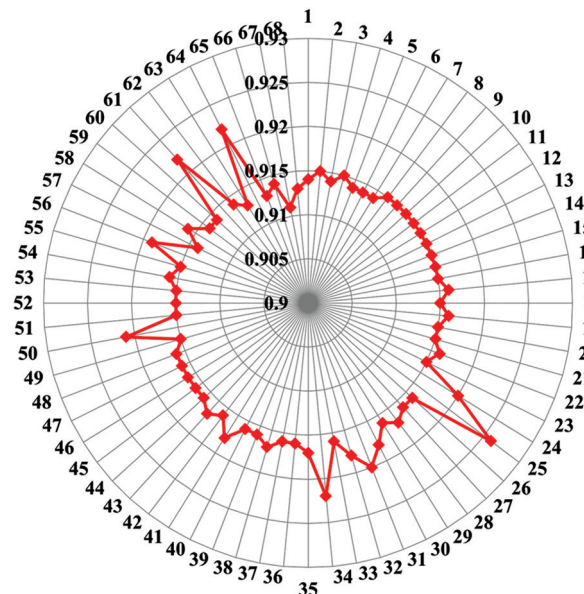
序号	化合物	A	E_2	M_{15}	M_{21}	保留时间			
						实验值 ^[23]	预测值 1	预测值 2	误差
14	异戊酸甲酯	12.527	0.524	-0.582	1.040	4.74	7.21	5.23	0.49
15	丁二醇	9.497	0	0	0	5.00	2.98	3.31	-1.69
16	正己醛	10.579	4.209	0	8.791	5.19	7.45	4.72	-0.47
17	戊酸甲酯	12.623	2.550	-0.365	2.286	5.70	8.59	6.41	0.71
18	2-甲基吡嗪	12.389	0	0.450	0	5.79	7.11	5.71	-0.08
19	2,4-二甲基-1-庚烯	12.389	3.842	-0.055	0	6.00	8.84	5.69	-0.31
20	异戊酸	11.305	0.278	-0.742	1.213	6.36	5.36	4.85	-1.51
21	正己醇	10.039	5.043	0	4.267	6.71	6.70	6.84	0.13
22	4-甲基戊酸甲酯	13.941	1.473	-0.875	1.947	7.24	9.57	7.89	0.65
23	正戊酸	11.401	2.082	-0.762	2.920	7.44	6.57	8.88	1.44
24	环戊基醋酸酯	15.451	4.578	-0.357	2.075	7.49	13.45	6.96	-0.53
25	辛酸甲酯	16.866	6.489	-0.430	3.521	7.65	16.46	11.60	3.95
26	2,6-二甲基吡嗪	13.656	0	0.988	0	7.83	9.08	7.70	-0.13
27	γ -丁内酯	10.623	2.178	-0.316	2.867	8.04	5.76	6.76	-1.28
28	己酸甲酯	14.037	3.805	-0.399	2.840	8.15	11.19	9.03	0.88
29	2-甲基戊酸	12.719	1.739	-1.279	1.729	8.56	7.84	7.36	-1.20
30	苯甲醛	14.583	0	5.924	9.132	9.40	13.46	9.61	0.21
31	2-正戊基呋喃	16.043	4.937	2.524	0	10.35	15.53	10.62	0.27
32	羧酸苯甲酯	16.773	0	1.893	6.896	10.67	14.41	10.94	0.27
33	己酸	12.815	3.277	-0.871	3.697	10.69	9.13	10.63	-0.06
34	8-壬炔酸	18.376	4.935	-0.927	4.237	10.96	17.57	12.01	1.05
35	庚酸甲酯	15.451	5.126	-0.418	3.232	11.53	13.82	11.03	-0.50
36	苯甲醇	14.043	0.140	4.675	3.288	12.16	11.66	12.79	0.63
37	3-辛烯-2-酮	13.897	3.419	-1.230	5.074	12.16	10.65	11.27	-0.90
38	己内酯	13.355	4.467	-0.337	4.209	12.90	10.76	13.02	0.12
39	苯乙酮	15.850	0	5.875	4.381	13.34	14.70	12.85	-0.49
40	3-乙基-2,5-二甲基吡嗪	16.339	0.973	1.486	0	13.72	13.45	13.15	-0.57
41	间甲酚	13.896	0	2.944	5.784	14.33	10.85	13.70	-0.63
42	辛酸	15.644	5.885	-0.979	4.654	14.33	14.35	14.18	-0.15
43	苯甲酸甲酯	18.041	0	6.174	4.540	14.45	17.83	14.43	-0.02
44	壬醛	14.822	8.379	0	9.982	14.69	15.47	15.78	1.09
45	可可醛	16.533	1.006	2.619	9.691	14.80	15.19	14.40	-0.40
46	辛酸甲酯	16.866	6.489	-0.430	3.521	15.56	16.46	11.60	-3.96
47	壬酸甲酯	18.280	7.881	-0.438	3.742	17.23	19.11	17.23	0
48	苯甲酸乙酯	19.455	0.424	6.676	4.883	17.66	20.23	17.68	0.02
49	苯乙酸甲酯	19.455	0.358	3.781	4.097	17.93	18.81	17.91	-0.02
50	呋喃酮	16.081	0	0	0	18.25	11.94	18.45	0.20
51	α -松油醇	16.675	3.357	6.684	2.335	18.56	17.69	18.60	0.04
52	癸醛	16.236	9.812	0	10.210	19.04	18.14	18.78	-0.27
53	壬烯酸甲酯	18.916	5.203	-0.597	4.237	19.75	18.59	19.70	-0.05
54	2-丁基-3-甲基吡嗪	17.899	3.492	2.407	0	21.03	17.27	20.80	-0.23

续表 2

序号	化合物	A	E_2	M_{15}	M_{21}	保留时间			
						实验值 ^[23]	预测值 1	预测值 2	误差
55	二正戊基酮	17.504	8.668	0.262	6.633	21.90	19.05	21.91	0.01
56	十六烯	23.706	10.104	0	0	22.11	27.46	22.22	0.11
57	3-苯丙酸甲酯	20.869	1.216	5.427	3.961	22.19	21.90	23.01	0.82
58	甲基壬基甲酮	17.504	9.842	-0.555	3.845	22.86	19.01	22.85	-0.01
59	2-丁基-2-辛烯醛	19.554	8.089	6.501	11.883	23.33	24.87	23.03	-0.30
60	4-癸烯酸甲酯	20.330	6.247	-0.509	4.014	23.53	21.06	23.10	-0.43
61	2-异丙基-5-甲基-1-庚醇	16.821	4.063	0.513	3.712	23.57	15.61	23.51	-0.06
62	癸酸甲酯	19.694	9.293	-0.444	3.916	24.16	21.77	24.15	-0.01
63	苯甲酸异丁酯	22.187	0.480	6.403	5.313	24.43	23.89	24.64	0.21
64	椰子醛	17.598	6.349	-0.209	3.951	26.18	17.53	26.19	0.01
65	肉桂酸甲酯	21.505	0	7.682	6.071	27.10	23.37	27.18	0.08
66	反式-肉桂酸甲酯	21.505	0	7.682	6.071	27.43	23.37	27.18	-0.25
67	雪松烯	22.681	7.093	6.134	0	28.16	27.29	27.92	-0.24
68	十一酸甲酯	21.108	10.721	-0.448	4.057	28.31	24.43	28.29	-0.02

1.4 模型稳健性检验

为检验选用的分子结构参数作为神经网络模型变量的可靠性,用逐一剔除法对选择的 4 个结构参数所建的 2 个方程稳健性进行检验.先对式(3)每次剔除 1 个分子,用余下的分子进行建模,这样共得到 76 个相关系数检验值.对这 76 个检验值作雷达图(图 1),可以看到,只有剔除第 32 号十四碳烯酸分子时,所得方程的相关系数值最大,说明该分子的存在对总体相关性影响较大.这可能与该分子链较长而保留时间相对较小有关,但总体而言,检验相关系数所得 r 的平均值为 0.883 6,与多元回归分析所得式(3)的相关系数值基本吻合,说明采用这 4 个变量来构建相关性模型稳健性可靠.同理,对式(4)用相同的方法进行检验,得到 68 个检验的相关系数.对这 68 个检验值作雷达图(图 2),检验相关系数所得 r 的平均值为 0.916,与多元回归分析所得式(4)的相关系数值完全一致,说明采用这 4 个变量来构建姬松茸挥发性成分保留时间的相关性模型同样稳健可靠.

图 1 松茸相关系数检验值 r 的雷达图图 2 姬松茸相关系数检验值 r 的雷达图

2 神经网络模型建构

2.1 松茸神经网络模型建构

由于多元回归方程的相关性不够理想, 为提高预测能力, 在多元回归方程基础上进一步用神经网络法进行研究. 按照许禄等^[25]及 Andrea 等^[26]的建议, 神经网络隐含层变量(Y)可按式(5)计算:

$$2.2 > n / [(S_i + 1)Y + (Y + 1)S_o] \geq 1.4 \quad (5)$$

式中: n , S_i , Y , S_o 分别为神经网络的分子样本数、输入层、隐含层和输出层变量.

将松茸挥发性成分的 A , E_2 , M_{15} 和 M_{21} 4 个分子结构参数作输入层变量, 色谱保留时间作输出层变量, 根据隐含层计算公式, 隐含层 Y 可取 6, 7 或 8, 经过计算分析比较, 发现当 Y 取 8 时, 所得结果最好, 故对松茸挥发性成分色谱保留时间预测的神经网络结构方式采用 4-8-1. 为防止建模过程中的过拟合, 将 76 个分子分为 3 组: 第 1 组为训练集组(以每 5 个分子为 1 组, 取其中第 1, 3, 5 个分子)、第 2 组为测试集组(每 5 个分子组中的第 2 个分子)、第 3 组为验证集组(每 5 个分子组中的第 4 个分子). 通过建模计算, 得到了松茸挥发性成分色谱保留时间预测模型各组集相关系数值为 $r_{\text{总}} = 0.9919$, $r_{\text{训练集}} = 0.9910$, $r_{\text{测试集}} = 0.9932$, $r_{\text{验证集}} = 0.9974$. 与多元回归方程(3)相比, 神经网络模型相关系数明显提高, 预测的松茸挥发性成分的色谱保留时间与文献实验值的吻合度较好, 两者的平均误差仅为 0.57. 将神经网络模型所得的预测值 2、实验值以及两者之间的误差列于表 1 中, 其关系见图 3.

2.2 姬松茸神经网络模型的建构

将姬松茸挥发性成分的 A , E_2 , M_{15} 和 M_{21} 4 个分子参数作为输入层变量, 相应色谱保留时间作为输出层变量, 根据隐含层计算公式, Y 可取 5, 6 或 7, 经计算分析当 Y 取 7 时, 所得的结果较好, 故对姬松茸挥发性成分色谱保留时间预测的神经网络结构方式采用 4-7-1. 模型的各个组集相关系数分别为 $r_{\text{总}} = 0.9929$, $r_{\text{训练集}} = 0.9905$, $r_{\text{测试集}} = 0.9971$, $r_{\text{验证集}} = 0.9963$. 利用该神经网络模型预测的姬松茸挥发性成分的色谱保留时间与文献实验值的平均误差为 0.58, 将相关分子的预测值 2 与实验值列入表 2, 其关系见图 4. 结果表明: 无论是松茸还是姬松茸, 根据模型所得预测值与实验值对应的点均在直线附近, 说明没有误差特别大的数据存在.

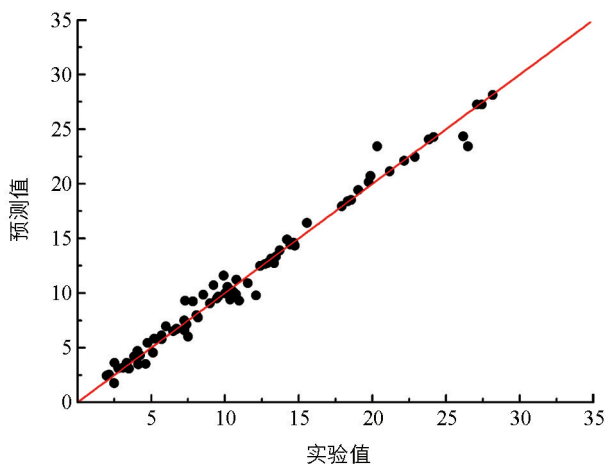


图 3 松茸保留时间实验值与预测值关系图

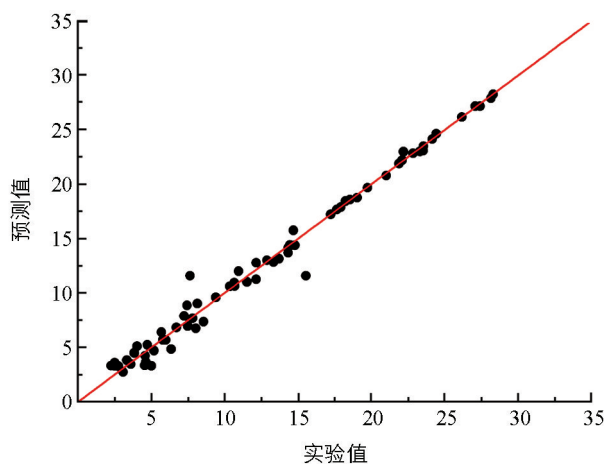


图 4 姬松茸保留时间实验值与预测值关系图

2.3 建构模型的检验

为检验所建构模型的可靠性, 以建构的姬松茸神经网络模型为例, 对姬松茸中其他挥发性成分的保留时间进行检验, 这里的姬松茸分子及相关数据随机选取于文献^[27]中的分子, 用模型对这些分子的保留时间进行预测, 可以看到预测值与实验值基本相符(表 3), 说明本法建构的模型基本可靠.

表 3 姬松茸挥发性成分保留时间的检验

序号	化合物	A	E_2	M_{15}	M_{21}	保留时间		
						实验值 ^[27]	预测值	误差
1	芳樟醇	16.675	1.670	2.169	3.955	16.013	14.950	-1.06
2	对异丙基苯甲醛	18.583	0	10.570	9.656	20.059	21.070	1.01
3	2,3-辛二酮	16.037	3.419	-0.063	3.347	12.081	13.903	1.82
4	3-辛烯-2-酮	13.897	3.419	-1.230	5.074	13.901	10.644	-3.26
5	肉豆蔻醚	25.009	1.066	14.620	0	26.447	31.200	4.75
6	1,3-二甲苯	12.928	0	5.952	0	8.035	10.340	2.30
7	1,4-二甲苯	12.928	0	5.792	0	8.093	10.267	2.17
8	1,2,3-三甲氧基苯	20.767	0	7.240	0	22.203	21.554	-0.65

3 讨论与结论

不同产地松茸和姬松茸中各种成分或含量均会存在一定的差异,而且这些挥发性成分种类较多,组成复杂,一般采用实验方法对其进行检测分析过程漫长而繁琐.本法通过建构一种新的原子指数 A,结合计算松茸和姬松茸挥发性成分的电性拓扑状态指数和电性距离矢量,并优化筛选了其中的几种结构参数,采用反向传播(BP)神经网络方法建构预测模型.建构的原子指数充分考虑了分子中原子本身的电性以及所处环境的空间结构特性,从而蕴含了对分子的贡献;选择的电性拓扑状态指数和电性距离矢量则同样能反映电性结构信息和空间拓扑结构连接性质.将他们结合可以充分反映松茸或姬松茸挥发性成分保留时间的变化规律,从表 1 和表 2 中松茸和姬松茸挥发性成分的结构参数和保留时间的大小可以看出,当检测条件一定时,除原子之间的范德华力、色散力、诱导力等会影响保留时间外,基团的大小、数量甚至连接情况均能影响保留时间的大小.在对所建模型的稳健性检验中可以看到,当剔除 32 号分子十四碳烯酸时,模型检验的相关系数相对明显高于其他值,说明长链的存在,由于体积越大,对保留时间影响也越大.本文建构一种新的原子指数 A,结合 E_2 , M_{15} 和 M_{21} 共 4 种结构参数,与松茸和姬松茸挥发性成分的色谱保留时间构建 BP 神经网络模型,模型总相关系数能达到 0.99 以上的优级相关,预测松茸保留时间与实验值的平均误差为 0.57,预测姬松茸的平均误差为 0.58,说明本法具有较好的可靠性,所建模型的预测能力较强,预测准确度较好.

新建构的原子指数 A、筛选的电性拓扑状态指数的 E_2 和电性距离矢量 M_{15} , M_{21} , 它们能充分反映松茸和姬松茸挥发性成分中非氢原子的数量和连接特性.通过分析挥发性成分的分子结构发现,基团—CH₃, —CH₂—和含有的杂原子(如氧原子、氮原子)是影响保留时间大小的一个主要片段,原子上连接的氢原子数量多少也会对性质有影响.从神经网络模型的相关系数明显高于多元回归模型可以看出,4 种分子结构参数与松茸、姬松茸挥发性成分色谱保留时间之间存在较好的非线性关系,与多元回归中反映出的线性关系不同.建构的松茸与姬松茸色谱保留神经网络模型,预测能力明显优于多元回归分析法,所得结果较为理想.

参考文献:

- [1] 王永辉,吴瑾,白家磊,等.不同产地松茸中氨基酸含量测定及比较[J].营养学报,2019,41(5):519-520.
- [2] 张沙沙,罗晓莉,何容,等.松茸风味物质研究进展及应用展望[J].农产品加工,2019(9):68-70.
- [3] TIDKE G, RAI M K. Biotechnological Potential of Mushrooms: Drugs and Dye Production [J]. International Journal of Medicinal Mushrooms, 2006, 8(4): 351-360.
- [4] 林英,张倩,班颖珏.松茸活性成分及研究进展[J].微生物学杂志,2018,38(1):118-122.

- [5] 解修超, 刘军生, 罗阳兰, 等. 4 株姬松茸胞外多糖含量和生物活性的对比分析 [J]. 食品研究与开发, 2019, 40(12): 68-75.
- [6] 房雷雷, 赵肖通, 辛慧洁, 等. 一种姬松茸多糖的纯度鉴定与单糖组成分析 [J]. 食品工业科技, 2018, 39(13): 94-97, 267.
- [7] 周宁. 姬松茸多糖对人体健康的改善作用及机制 [J]. 中国食用菌, 2020, 39(1): 38-41.
- [8] 魏晓梅, 吴丽芳, 柏旭, 等. 云南松茸多糖提取方案优化及其测定 [J]. 中国食用菌, 2018, 37(1): 46-49.
- [9] 申秀娟, 张磊, 杨美玉, 等. 醇解姬松茸多糖对小鼠免疫指标和肠道菌群的影响 [J]. 食品科技, 2019, 44(6): 215-220.
- [10] 石芳, 李瑶, 杨雅轩, 等. 不同干燥方式对松茸品质的影响 [J]. 食品科学, 2018, 39(5): 141-147.
- [11] ZHAI F H, WANG Q, HAN J R. Nutritional Components and Antioxidant Properties of Seven Kinds of Cereals Fermented by the Basidiomycete *Agaricus blazei* [J]. Journal of Cereal Science, 2015, 65: 202-208.
- [12] ONUMA H, HARA K, SUGITA K, et al. Purification and Characterization of a Glycoside Hydrolase Family 5 Endoglucanase from *Tricholoma matsutake* Grown on Barley Based Solid-State Medium [J]. Journal of Bioscience and Bioengineering, 2019, 128(6): 669-676.
- [13] ONUMA H, UCHIYAMA H, HARA K, et al. Enzymatic Characterization of an Extracellular Glucoamylase from *Tricholoma matsutake* and Its Cloning and Secretory Expression in *Pichia pastoris* [J]. Bioscience, Biotechnology, and Biochemistry, 2018, 82(12): 2180-2190.
- [14] GUO Y, CHEN D, DONG Y F, et al. Characteristic Volatiles Fingerprints and Changes of Volatile Compounds in Fresh and Dried *Tricholoma matsutake* Singer by HS-GC-IMS and HS-SPME-GC-MS [J]. Journal of Chromatography B, 2018, 1099: 46-55.
- [15] 朱湘临, 宋彦, 王博, 等. 基于改进布谷鸟算法-BP 神经网络的松茸发酵过程软测量建模 [J]. 计算机测量与控制, 2019, 27(5): 39-43.
- [16] 堵锡华, 李靖, 田林, 等. 含氮杂环类润滑油添加剂抗磨损性能的神经网络研究 [J]. 福州大学学报(自然科学版), 2019, 47(4): 560-565.
- [17] DU X H, ZHUANG W C, SHI X Q, et al. Research on Thermodynamic Properties of Polybrominated Diphenylamine by Neural Network [J]. Chinese Journal of Chemical Physics, 2015, 28(1): 59-64.
- [18] 堵锡华, 田林, 徐艳, 等. 枣花及枣花蜜香气成分保留指数的神经网络法研究 [J]. 西南大学学报(自然科学版), 2021, 43(9): 142-152.
- [19] 堵锡华, 吴琼, 陈艳, 等. 酚类化合物臭氧氧化速率的神经网络研究 [J]. 生态毒理学报, 2017, 12(3): 675-680.
- [20] 石佳超, 罗坤, 樊建人, 等. 基于 CMAQ 与前馈神经网络的区域大气污染物浓度快速响应模型 [J]. 环境科学学报, 2018, 38(11): 4480-4489.
- [21] 冯玉芳, 卢厚清, 殷宏, 等. 基于 BP 神经网络的故障诊断模型研究 [J]. 计算机工程与应用, 2019, 55(6): 24-30.
- [22] 瞿英, 王冕, 董文旭, 等. 基于 BP 神经网络的农田大气氨浓度预测 [J]. 中国生态农业学报, 2019, 27(4): 519-528.
- [23] 张宪臣, 张朋杰, 容裕棠, 等. 松茸和姬松茸挥发性成分比较分析 [J]. 食品科学, 2019, 40(10): 229-235.
- [24] 张婷, 梁逸曾, 赵晨曦, 等. 基于分子结构预测气相色谱程序升温保留指数 [J]. 分析化学, 2006, 34(11): 1607-1610.
- [25] 许禄, 邵学广. 化学计量学方法 [M]. 2 版. 北京: 科学出版社, 2004.
- [26] ANDREA T A, KALAYEH H. Applications of Neural Networks in Quantitative Structure-activity Relationships of Dihydrofolate Reductase Inhibitors [J]. Journal of Medicinal Chemistry, 1991, 34(9): 2824-2836.
- [27] 张艳荣, 吕呈蔚, 刘通, 等. 不同干燥方式对姬松茸挥发性风味成分分析 [J]. 食品科学, 2016, 37(10): 116-121.