

DOI:10.14077/j.issn.1007-7812.202209022

基于分子体积法的 CHONF 高能致爆材料 晶体密度的预测

文琳元^{1,2}, 何晓凯², 师进文¹, 刘英哲², 王伯周²

(1. 西安交通大学 能源与动力工程学院, 陕西 西安 710049; 2. 西安近代化学研究所, 陕西 西安 710065)

摘 要:针对现有的晶体密度预测研究大多围绕 CHON 类含能材料,在预测 CHONF 高能致爆材料晶体密度时会由于公式适用范围不同,出现预测误差较大的问题,为提高对该类材料的性能评估精度,从剑桥晶体数据中心中筛选出 56 个该类材料,在 B3PW91/6-31g(d,p) 计算水平下系统性评估了 Rice 以及 Politzer 公式在预测该类材料晶体密度的表现,然后在分子表面静电势相关的典型参数基础上利用符号回归方法和“加、减、乘、除”4 种算符构建出适用该体系的密度预测公式。结果表明,符号回归公式的均方根误差从 Rice 公式的 0.106 g/cm^3 降低到 0.045 g/cm^3 ,并将超过半数的含氟高能致爆材料的密度预测误差降低到 0.030 g/cm^3 以下,极大地提高了含氟高能致爆材料晶体密度预测精度。

关键词:量子化学; CHONF 类含能材料; 晶体密度预测; 符号回归; 高能量密度材料

中图分类号: Tj55; O641

文献标志码: A

文章编号: 1007-7812(2022)06-0807-07

Prediction of Crystal Density of CHONF High Energetic Explosophoric Materials Based on Molecular Volume Method

WEN Lin-yuan^{1,2}, HE Xiao-kai², SHI Jin-wen¹, LIU Ying-zhe², WANG Bo-zhou²

(1. School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an 710049, China;

2. Xi'an Modern Chemistry Research Institute, Xi'an 710065, China)

Abstract: Most of the existing studies on crystal density prediction focus on CHON energetic materials. Due to the different application ranges of the formulas, large prediction errors appear in predicting the crystal density of CHONF highly energetic explosophoric materials. In order to improve the performance evaluation accuracy, at the B3PW91/6-31g(d,p) level, the performances of the Rice and Politzer formulas in predicting the crystal density of 56 such materials screened from the Cambridge Crystal Data Center were first systematically studied. In addition, on the basis of the typical parameters related to the molecular surface electrostatic potential, the density prediction formula suitable for this system was constructed by using symbolic regression method and four operators of “addition, subtraction, multiplication, and division”. The results showed that the root-mean-square error of the symbolic regression formula was reduced from 0.106 g/cm^3 of the Rice formula to 0.045 g/cm^3 , and the predicted error of over 50% of materials were less than 0.030 g/cm^3 . The symbolic regression method greatly improved the crystal density prediction accuracy of fluorine-containing highly energetic explosophoric materials.

Keywords: quantum chemistry; CHONF-type energetic materials; crystal density prediction; symbolic regression; high-energy-density materials

收稿日期: 2022-09-29; 修回日期: 2022-10-31

基金项目: 陕西省杰出青年科学基金(No. 2022JC-10)

作者简介: 文琳元(1995—), 男, 博士研究生。E-mail: w466258394@stu.xjtu.edu.cn

通信作者: 师进文(1981—), 男, 副教授。E-mail: jinwen_shi@mail.xjtu.edu.cn

刘英哲(1986—), 男, 博士, 研究员。E-mail: liuyz_204@163.com

引用本文: 文琳元, 何晓凯, 师进文, 等. 基于分子体积法的 CHONF 高能致爆材料晶体密度的预测[J]. 火炸药学报, 2022, 45(6): 807-813.

WEN Lin-yuan, HE Xiao-kai, SHI Jin-wen, et al. Prediction of Crystal Density of CHONF High Energetic Explosophoric Materials Based on Molecular Volume Method[J]. Chinese Journal of Explosives & Propellants(Huozhayao Xuebao), 2022, 45(6): 807-813.

引 言

作为炸药、发射药、推进剂和火工品中高能量组分的含能材料,其能量密度水平一定程度上决定了武器发射、推进及毁伤能力,乃至国家整体军事装备的战略威慑和打击能力。追求更高能量密度,成为了含能材料领域科学家为之长期奋斗的目标。为此,研究者们提出了引入高能致爆基团构筑新型高能量密度含能材料的策略^[1]。同时,由于新材料的合成通常需要大量的时间与金钱,因此准确预测含能材料性能对于开发具有优良性能的新型含能材料至关重要。随着计算科学与理论化学的快速发展,新型含能材料的理论设计技术日渐成熟,具体为通过量化计算预估新化合物的生成焓、密度、爆轰速度、爆轰压力等参数,辅以性能指标要求筛选到最终的目标化合物。其中,因为 K-J 方程中爆轰速度与密度正相关且爆轰压力与密度的平方成正比,所以密度被认为是影响含能材料爆轰性能的关键参数,如何准确预测新型含能材料晶体密度也就成为新型含能材料研发中一大关键问题。

当前含能材料晶体密度预测方法包含物理模型与统计模型两种。其中物理模型可细分为基于分子体积^[2-4]以及基于晶体体积^[5-9]两类,而基于统计模型中较为主流的方法则是机器学习^[10]。目前基于分子体积方法用的较多的是分子表面静电势法,该方法是用分子摩尔质量(M)除以使用 Monte-Carlo 方法计算得到的分子 0.001 a. u. 电子密度等值面包裹内体积 $V_{0.001}$ 的值来表示晶体密度^[11-12],该方法未考虑分子间作用和晶体堆积,计算误差与速度适中。Politzer 在分子表面静电势法的基础上,针对 CHON 类材料提出了分子表面静电势参数修正的方法^[13-14],其预测密度与实验值误差小于 0.05 g/cm^3 ,被广泛运用于含能材料晶体密度预测。

基于晶体体积方法则是通过晶体结构预测,以获得含能材料准确的晶体结构,从而得到该材料的晶体密度。该方法需要准确描述分子间相互作用,存在计算速度较慢的不足,是当前含能材料晶体密度研究的一大挑战。而机器学习模型主要是通过各种回归算法构建材料结构与密度之间的关系,其中大多数是黑箱模型,缺乏解释性。近年来,不少研究者基于该方法在 CHON 类含能材料晶体密度预测上做了许多优秀工作,其中张朝阳研究员针对超过 2000 个硝基化合物构建了均方根误差(RMSE) = 0.049 g/cm^3 的图神经网络密度预测模型^[15],T. Yong-jin Han 针对 10251 个含有至少一根氮氧键的

CHON 类化合物开发了 RMSE = 0.062 g/cm^3 消息传递神经网络模型^[16],张庆华研究员对超过 1000 个 CHON 类含能材料利用核岭回归的方法构建了 MAE = 0.042 g/cm^3 的密度预测模型^[17]。

从上述研究可以发现,其样本量均大于 10^3 ,这是因为机器学习模型的计算精度十分依赖数据样本,样本量过少会使得模型过拟合。这其中,符号回归作为一种回归分析方法,在挖掘变量间内在关系以及帮助理解物理机制并发现隐藏的数学表达式方面取得一定的成果^[18-19]。

由于当前晶体密度预测方法大多针对 CHON 类含能材料,没有考虑氟代偕二硝基($-\text{CF}(\text{NO}_2)_2$)、二氟氨基($-\text{C}(\text{NF}_2)$)以及氟代偕硝基($-\text{CF}(\text{NO}_2)$)等含氟高能致爆基团对密度预测的影响,难以保证上述方法在预测含氟高能致爆材料晶体密度的精度。因此,有必要建立一个适用于含氟高能致爆材料的晶体密度预估方法。其中,基于晶体体积方法存在计算速度较慢、缺乏准确描述分子间相互作用的不足,因此本研究仅采用基于分子体积与基于机器学习的方法。先在 Politzer 基于 CHON 类含能材料密度预测公式基础上,利用多元线性拟合方法得到适合含氟高能致爆材料体系的晶体密度预测公式。考虑到含有上述含氟高能致爆基团材料数量较少的情况,本研究将机器学习模型中的符号回归算法与基于分子体积方法相结合,从多种静电势参数中确定出更适用于该研究体系的变量表现形式,构造出了适用于含氟高能致爆材料体系的 $R^2 = 0.77$ 、RMSE = 0.045 g/cm^3 的晶体密度预测公式。所得公式能提高含氟高能致爆材料晶体密度预测精度,以期对含氟高能致爆材料的高效准确设计奠定基础。

1 流程及原理

从剑桥晶体结构数据库(CCDC)中,按照实验密度大于 1.6 g/cm^3 、晶体结构为单晶、分子中存在含氟高能致爆基团等标准,筛选出 56 个化合物。再利用 Gaussian16^[20] 软件的 B3PW91/6-31g(d,p) 方法,对上述化合物的分子进行几何结构优化并得到能量最优的构型,经频率分析确定无虚频。

根据分子表面静电势理论,利用 Multiwfn^[21] 计算得到各参数,表 1 列出几个典型静电势参数。基于多元线性拟合方法,对 Politzer 所提的密度拟合公式进行修正,并在静电势参数基础上采用符号回归方法得到最优的公式形式。

表 1 56 个含氟高能致爆材料的部分静电势参数及实验晶体密度

Table 1 Electrostatic potential parameters and experimental crystal densities for 56 fluorine-containing high energetic explosives

refcode	$(M/V_m)/$ ($g \cdot cm^{-3}$)	$M/$ ($g \cdot mol^{-1}$)	$\sigma_{tot}^2/$ ($kJ^2 \cdot mol^{-2}$)	$\sigma_+^2/$ ($kJ^2 \cdot mol^{-2}$)	$\nu\sigma_{tot}^2/$ ($kJ^2 \cdot mol^{-2}$)	$\Pi/$ ($kJ \cdot mol^{-1}$)	$A_s/$ (Å^2)	$\rho/$ ($g \cdot cm^{-3}$)
KUPYUT	1.816	220.131	1886.2	898.9	470.5	262.4	192.249	1.667
CIFRER	1.748	235.134	5037.5	3264.5	1149.0	332.3	222.299	1.680
YORYIN	1.727	195.106	3235.5	2062.5	747.7	262.9	189.445	1.691
YIJIW	1.783	285.144	2198.8	1477.4	484.7	243.8	265.218	1.722
JIVXUK	1.836	249.114	4556.0	3259.2	927.6	284.8	226.262	1.726
GIHVUO	1.886	390.166	1609.3	1096.7	349.3	208.8	310.831	1.736
HEGSIV	1.833	464.246	1456.0	1023.5	304.1	201.0	402.925	1.742
YAZXIJ	1.774	235.127	2255.9	1476.2	510.2	281.7	217.303	1.744
FADNAN	1.862	219.103	2124.7	1404.0	476.2	252.3	197.352	1.757
JIVYAR	1.838	249.114	3957.5	3098.6	672.5	256.2	225.755	1.783
YEQNEQ	1.764	220.119	7156.9	4568.3	1652.3	316.3	206.912	1.785
LUQNIU	1.970	332.130	2528.9	2169.5	308.3	231.9	258.249	1.786
HIHSID	1.901	352.206	1908.5	1512.7	313.8	219.8	259.758	1.787
XOKXED	2.016	647.275	3378.2	3138.4	222.8	237.2	450.896	1.798
NABNAF	1.992	334.103	2433.2	2192.2	217.1	206.8	280.284	1.805
CIFRIV	1.861	280.131	3238.8	2433.2	605.2	284.6	239.788	1.809
TAXIEU01	1.963	426.162	2971.3	2592.1	330.8	231.4	352.352	1.832
GEXDAO	1.897	384.171	2514.3	2184.0	286.9	226.1	326.045	1.839
ZIZFEV	2.019	624.195	2431.7	2269.4	151.4	208.3	475.070	1.842
JIVYIZ	1.841	249.114	6157.2	3982.2	1406.7	309.4	225.140	1.852
JIVYEV	1.847	206.092	5195.2	3620.9	1097.2	281.0	188.852	1.861
WOZILU	1.835	280.131	2900.1	2354.3	443.1	265.8	253.122	1.870
ACUPAS	1.925	393.153	2513.2	2181.3	288.0	236.2	325.157	1.871
FLNOAM	1.993	425.147	5115.1	4606.1	458.4	268.0	307.600	1.871
TICKAU	1.923	303.125	4968.8	3465.6	1048.4	282.9	257.547	1.872
CELFII	2.014	362.113	2577.7	2237.7	295.2	214.6	274.203	1.873
WISKEO	1.970	348.129	1707.9	1413.6	243.6	196.0	275.983	1.876
KIVLOS	1.953	420.152	3449.9	2416.4	723.9	245.4	355.007	1.884
CUVXUM	2.034	405.122	3910.2	3685.6	211.7	207.1	295.296	1.884
EDEHAZ	1.912	372.157	3623.5	3224.6	355.0	248.3	312.645	1.894
XAJYOX	2.100	243.088	2505.3	1651.4	562.9	195.4	196.420	1.897
LILGAQ	1.888	154.054	4034.8	3619.3	372.7	256.8	147.287	1.899
FANIBZ	1.958	264.100	2599.0	2193.6	342.2	251.1	221.310	1.902
PALDOW	2.101	243.088	2104.2	1371.5	477.6	183.8	196.256	1.902
CUVFAA	2.037	351.108	4080.5	3853.8	214.1	226.9	274.148	1.906
IVUGEN	1.959	386.141	2839.9	2540.3	268.0	223.2	324.606	1.914
VAYVIA	2.029	632.221	3607.2	3262.5	311.8	237.8	452.032	1.914
AWEMAS	2.050	410.122	1902.1	1721.8	163.2	175.1	316.466	1.920
LELXOP	2.048	780.262	4443.9	4111.6	307.4	238.9	522.429	1.921
FNETAM10	2.012	334.106	4049.5	3737.0	288.3	242.1	263.383	1.922
CELFOO	2.108	332.087	2975.4	2519.2	386.2	190.4	247.969	1.925
KENGES	2.094	470.135	3069.8	2700.7	324.7	214.9	351.382	1.926
ACUNUK	1.981	362.113	2647.7	2370.5	248.2	219.0	304.983	1.930
UEZOC	1.998	408.145	3058.0	2761.9	267.4	235.6	312.765	1.938
YUFRIC	1.952	242.076	3725.2	3071.4	539.0	258.5	215.448	1.939

Continued

refcode	$(M/V_m)/$ (g · cm ⁻³)	$M/$ (g · mol ⁻¹)	$\sigma_{tot}^2/$ (kJ ² · mol ⁻²)	$\sigma_+^2/$ (kJ ² · mol ⁻²)	$\nu\sigma_{tot}^2/$ (kJ ² · mol ⁻²)	$\Pi/$ (kJ · mol ⁻¹)	$A_s/$ (Å ²)	$\rho/$ (g · cm ⁻³)
DFINBD	2.047	306.092	3593.4	3349.4	227.5	214.2	238.781	1.946
XOKXIH	2.030	529.227	3566.9	3296.6	249.8	255.8	377.182	1.947
BIZKIG	2.054	382.109	2214.8	1983.7	207.0	179.6	307.687	1.960
CUVDUS	2.040	378.115	3066.5	2848.8	202.2	199.2	284.254	1.965
IVUGIR	2.001	418.139	3694.4	3134.2	475.3	253.8	343.420	2.007
SABMOA	2.021	280.085	2811.5	2439.7	322.6	224.6	228.472	2.011
HEGRUG	2.095	314.075	2274.2	2064.8	190.1	179.5	252.070	2.024
YOQWUY	2.099	169.026	1782.4	1650.0	122.6	155.5	146.986	2.033
BIZNIJ	2.054	382.109	2215.9	1984.9	206.9	179.6	307.687	2.040
BIWREE	2.122	396.088	1763.1	1646.4	108.9	152.1	285.722	2.048
KUPYUP	2.101	154.029	2996.1	2676.6	285.4	233.3	138.236	2.053

注: A_s 为 0.001 a. u 电子密度等值面总面积(Å²)。

其中,选取了 Rice 在 2007 年所提的公式(1)^[12]、Politzer 在 2009 年所提的两个公式(2)和(3)^[13]作为对比,并计算了各公式在预测含氟高能致爆材料时的均方根误差(RMSE)。

$$\rho = \frac{M}{V_m} \quad (1)$$

$$\rho = 0.9183 \left(\frac{M}{V_m} \right) + 0.0490 (\nu\sigma_{tot}^2) + 0.0443 \quad (2)$$

$$\rho = 0.9186 \left(\frac{M}{V_m} \right) + 0.0506 (\Pi) - 0.0632 \quad (3)$$

式中: ρ 为计算密度, g/cm³; M 为摩尔质量, g/mol; V_m 为 0.001 a. u. 电子密度包裹的分子体积, cm³/mol; $\nu\sigma_{tot}^2$ 可描述静电相互作用趋势, kJ²/mol²; Π 为静电势平均偏差, kJ/mol。为后续表述简便,记 M/V_m 前系数为 α , $\nu\sigma_{tot}^2$ 前系数为 β 。

2 结果与讨论

根据公式(1),其实验晶体密度与预测密度关系如图 1 所示。

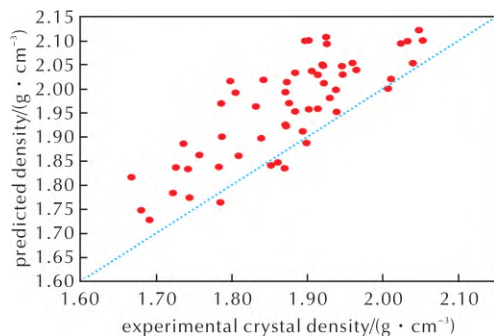


图 1 Rice 公式下预测密度与实验晶体密度关系
Fig. 1 The relationship between predicted density and experimental crystal density under Rice formula

从图 1 可以清楚地发现,其预测密度几乎都大

于对应的实验晶体密度,这首先说明氟元素加入改变了其分子间相互作用强弱,使得 0.001 a. u. 电子密度等值面包裹的体积 V_m 小于等效的晶体体积;同时,按照 Goh 的标准^[22],公式(1)计算中仅有 17.9% 的化合物预测密度可视为准确(误差小于 0.03 g/cm³),25.0% 的化合物预测密度可视为有用(误差小于 0.05 g/cm³),而公式(1)预测密度的均方根偏差高达 0.106 g/cm³,不足以达到晶体密度预测的要求。

而考虑 Politzer 提出的静电势修正公式(2),即用静电势的方差 σ_{tot}^2 与量化静电势平衡程度的 ν 之积,其预测的晶体密度与实际晶体密度关系如图 2 所示。

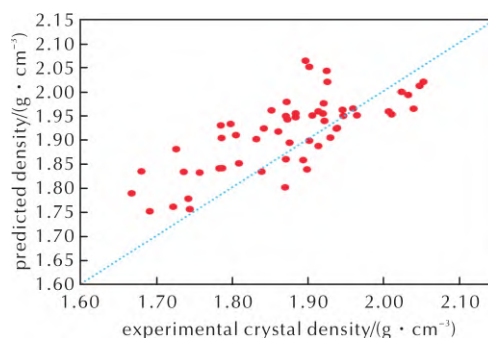


图 2 Politer- $\nu\sigma_{tot}^2$ 公式下预测密度与实验晶体密度关系
Fig. 2 The relationship between predicted density and experimental crystal density under Politer- $\nu\sigma_{tot}^2$ formula

从图 1 和图 2 可以看出,公式(2)得到预测密度比公式(1)预测密度小,这主要在于公式(2)中主要影响因素(M/V_m)前面的系数 α 小于公式(1)中的系数 $\alpha=1$ 。该公式使得预测结果更偏向于实际晶体密度,但从图 2 看,其仍呈现出高估了晶体密度的现象。相较没考虑分子间相互作用的公式(1),公

式(2)计算得到的晶体密度与实际晶体密度的偏差减小,均方根偏差从 0.106 g/cm^3 降到 0.075 g/cm^3 ,在 56 种含氟高能致爆材料中,预测晶体密度可视为准确的比例提高到 26.8%,而视为有用的比例提高到 44.6%,几乎是成倍地增长。这说明,加入静电势修正项,是符合物理本质且对于含氟晶体密度预测有用的。但是,其中仍有 21.4%的结果偏差是大于 0.100 g/cm^3 ,说明这一套仅考虑 C、H、O、N 元素的静电势修正拟合参数,不足以满足现有的需求。

接下来,同样基于 Politer 分子静电势修正的理论,不过把表示静电势作用的 $\nu\sigma_{\text{tot}}^2$ 改成描述静电势平均偏差的 Π ,按照公式(3)得到相应的预测晶体密度,并将其与实际晶体密度绘制在图 3 中。

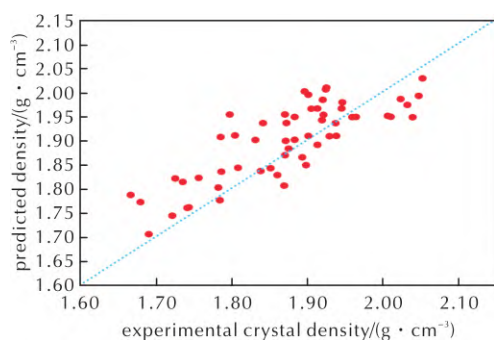


图 3 Politer-II 公式下预测密度与实验晶体密度关系

Fig. 3 The relationship between predicted density and experimental crystal density under Politer-II formula

从图 3 中可以发现,用静电势的平均偏差来量化其分子内部相互作用在含氟高能致爆材料晶体密度预测中也是有作用的,而且其均方根偏差从用 $\nu\sigma_{\text{tot}}^2$ 表示静电势作用的 0.075 g/cm^3 降至 0.060 g/cm^3 ,预测晶体密度可视为准确的比例提高到 39.3%,而视为有用的比例提高至 51.8%。但是,用该方法仍然有 8.9%的预测偏差大于 0.100 g/cm^3 。

由于上述使用 Politer 所提的分子表面静电势修正公式在预测含氟高能致爆材料晶体密度时误差仍较大,因此为了更为准确地描述氟元素加入后分子间相互作用对预测公式的影响。综合考虑 (M/V_m) 在公式(1)、(2)、(3)中关键作用,在公式(2)、(3)的基础采用多元线性拟合方法修正密度预测公式,得到如下式(4)、(5):

$$\rho = 0.80924 \left(\frac{M}{V_m} \right) + 0.0115 (\nu\sigma_{\text{tot}}^2) + 0.27207 \quad (4)$$

$$\rho = 0.84699 \left(\frac{M}{V_m} \right) + 0.0296 (\Pi) + 0.12002 \quad (5)$$

发现修正公式(4)和(5)的决定系数与均方根误差都为 0.65 和 0.055 g/cm^3 ;而公式(4)在密度预

测误差小于 0.05 g/cm^3 和小于 0.03 g/cm^3 上的表现优于公式(5),因此在后续仅讨论公式(4)的效果。其预测密度与实际晶体密度如图 4 所示,相较于前面提及的 3 种方法,改进后预测结果的相对误差都在 5%以内,而且均方根偏差也降至 0.055 g/cm^3 。预测晶体密度可视为准确的比例为 33.9%,而视为有用的比例提高到 60.7%。不仅如此,预测晶体密度的偏差大于 0.100 g/cm^3 的比例也降到 1.8%。

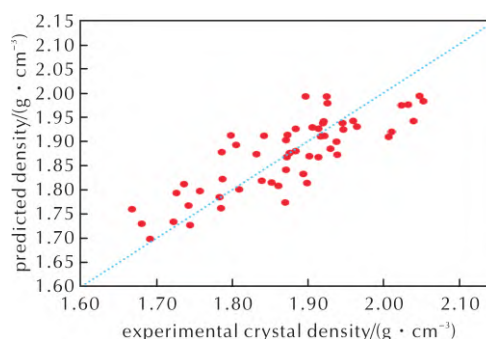


图 4 多元线性拟合改进后密度预测公式中预测密度与实验晶体密度关系

Fig. 4 The relationship between predicted density and experimental crystal density in the improved density prediction formula by multiple linear regression

对比公式(2)、(4),可以发现 α 由 0.9183 减小到 0.80924,而 β 也从 0.0490 减小到 0.0115,这说明在使用上述参数拟合时:用 0.001 a. u 电子密度包裹体积来等效表示含氟高能致爆材料晶体体积,还是不足以体现其复杂的内部相互作用对密度的影响且会在数值上低估其等效晶体体积,这可能是与氟原子更多的电荷以及更强的电负性相关;同时, $\nu\sigma_{\text{tot}}^2$ 前面较小的系数说明了该项修正量的贡献不如在 CHON 类含能材料体系,也就是说该项修正量不足以描述含氟高能致爆材料分子间相互作用。进而在公式(4)中,密度预测结果主要取决于 M/V_m ,缺乏静电势修正理论对含氟高能致爆材料分子间相互作用的体现。也就是将含氟体系与 CHON 类含能材料的差异都表现在等效体积一项上,而使用 0.001 a. u 电子密度等值面包裹的体积作为晶体等效体积是在数值上低估的,反映在 M/V_m 值上就是该值偏大,也就解释了为何公式(4)中 α 比公式(2)、(3)中都小。

虽然公式(4)中静电势方差参数在含氟高能致爆材料晶体密度预估公式中作用不明显,但将公式(4)中静电势方差替换为其他静电势参数的组合,可能会更好地体现含氟高能致爆材料体系内相互作用。由此课题组基于符号回归概念,开展含氟高能致爆材料的晶体密度预测研究。在本研究中,考

考虑到公式(2)以及(3)中静电势参数主要的运算方式为“加、减、乘、除”(分别对应于“add、sub、mul、div”等 4 个算符),

因此,在符号回归中基于上述 4 种运算符构建模型。如图 5 所示,利用符号回归模型后得到的含氟高能致爆材料晶体密度预测公式的二叉树表示,其中每个叶节点都是变量或者常数,而内部节点是选定的运算符。

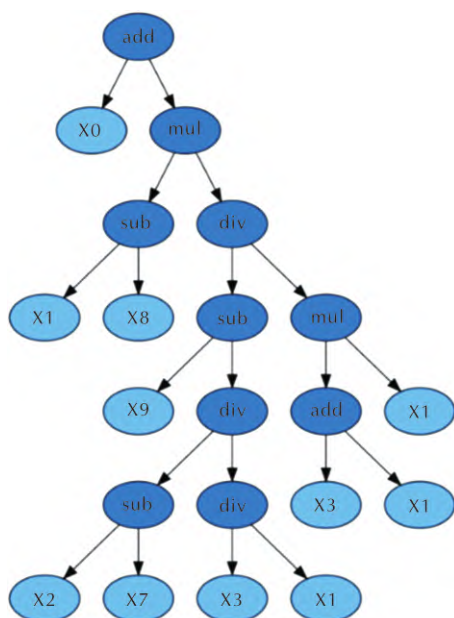


图 5 含氟高能致爆材料晶体密度预测公式-二叉树表示
Fig. 5 Density prediction formula of fluorine-containing high energetic explosives-binary tree representation

根据图 5 的二叉树,可以得到其表达式,见公式(6):

$$\rho = X0 + \frac{(X1 - X8) \left(X9 - \frac{(X2 - X7) X1}{X3} \right)}{(X3 + X1) X1} \quad (6)$$

变量 $X0 \sim X11$ 分别对应于 $M/V_m, M, \sigma_{\text{tot}}^2, \sigma_+^2, \sigma_-^2, \nu, \nu\sigma_{\text{tot}}^2, \Pi, A_s, A_s^+, A_s^-, V_m$, 将对应变量带入公式(6)可得:

$$\rho = \frac{M}{V_m} + \frac{\left(1 - \frac{\Pi}{M}\right) \left(\frac{A_s}{M} - \frac{(\sigma_{\text{tot}}^2 - \nu\sigma_{\text{tot}}^2)}{\sigma_+^2}\right)}{(\sigma_+^2 + M)} \quad (7)$$

符号回归得到的密度预测公式,其 $R^2 = 0.77$, $\text{RMSE} = 0.045 \text{ g/cm}^3$, 优于利用多元线性拟合方法修正的公式(4)。预测密度与实验晶体密度的分布如图 6 所示,可以发现该公式弥补之前数个公式在预测实验密度低于 1.75 g/cm^3 的材料时高估密度以及预测实验密度高于 2.00 g/cm^3 的材料时低估密度的问题,表现在图中就是头尾的数据更加趋近于蓝色点线。如表 2 所示,相比于多元线性拟合方法,

符号回归在降低预测均方根误差的同时,还提高了预测公式的综合表现。具体而言,一半以上材料的密度预测误差控制在 0.03 g/cm^3 以下,超过 70% 材料的预测误差低于 0.05 g/cm^3 。结合之前对于公式(2)、(3)和(4)的分析,公式(7)等号右侧第二项是由静电势方差、静电势偏差、 0.001 a.u 电子密度等值面面积以及分子摩尔质量组合的新变量,该参数大致体现含氟高能致爆材料复杂的分子间相互作用并以数学表达式呈现,使得该公式能够弥补之前谈到的 M/V_m 对密度的高估问题,从而也就解释了为何该公式在密度预测时较优的表现。

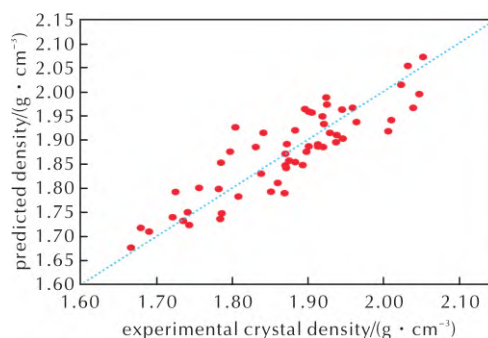


图 6 符号回归方法改进后密度预测公式中预测密度与实验晶体密度关系

Fig. 6 The relationship between the predicted density and the experimental crystal density in the density prediction formula after the improvement of the symbolic regression method

表 2 各公式预测精度对比

Table 2 Comparison of the prediction accuracy of different formulas

formula	error distribution/%			RMSE/ ($\text{g} \cdot \text{cm}^{-3}$)
	<0.03 (g/cm^3)	<0.05 (g/cm^3)	<0.100 (g/cm^3)	
(1)	17.9	25.0	58.9	0.106
(2)	26.8	44.6	78.6	0.075
(3)	39.3	51.8	91.1	0.060
(4)	33.9	60.7	98.2	0.055
(5)	32.1	57.1	96.4	0.055
(7)	51.8	71.4	98.2	0.045

3 结 论

(1) 基于分子体积方法,在 Politzer 的 CHON 晶体密度预测公式基础上,运用多元线性拟合得到了均方根误差为 0.055 g/cm^3 的修正公式。

(2) 结合符号回归方法构建了适合含氟高能致爆材料晶体密度预测的公式,其均方根误差为 0.045 g/cm^3 , 极大地提高了预测精度,其中有超过 50% 的化合物预测误差低于 0.03 g/cm^3 。

(3) 符号回归构建的预测公式提高了新型含氟

高能致爆材料晶体密度预测精度,有助于加速含氟高能致爆材料的设计。与此同时,符号回归方法还可以运用于其他研究体系,以破译材料性能间高维复杂的关系,构建出可理解的数学表达式,为材料设计提供思路。

参考文献:

- [1] 王伯周. 高能致爆基团构筑新型高能量密度材料策略[J]. 含能材料, 2020, 28(7): 588-590.
WANG Bo-zhou. Synthetic strategies for new high energy density materials based on high energetic explosophoric groups[J]. *Chinese Journal of Energetic Materials*, 2020, 28(7): 588-590.
- [2] 舒远杰, 李华荣, 熊鹰, 等. 含能材料理论设计中的几个问题[J]. 含能材料, 2013, 21(2): 166-172.
SHU Yuan-jie, LI Hua-rong, XIONG Ying, et al. Some problems in theoretical design of energetic materials[J]. *Chinese Journal of Energetic Materials*, 2013, 21(2): 166-172.
- [3] SINGH H, MUKHERJEE U, SAINI R S. Computational studies on nitro derivatives of 1-hydroxy-1, 2, 4-triazole[J]. *Journal of Energetic Materials*, 2012, 30(3): 265-281.
- [4] MA P, JIANG J C, ZHU S. Synthesis, XRD and DFT studies of a novel cocrystal energetic perchlorate amine salt; methyamine triethylenediamine triperchlorate[J]. *Combustion, Explosion, and Shock Waves*, 2017, 53(3): 319-328.
- [5] CHAN H C S, KENDRICK J, NEUMANN M A, et al. Towards ab initio screening of co-crystal formation through lattice energy calculations and crystal structure prediction of nicotinamide, isonicotinamide, picolinamide and paracetamol multi-component crystals[J]. *CrystEngComm*, 2013, 15(19): 3799-3807.
- [6] WU H, HABGOOD M, PARKER J E, et al. Crystal structure determination by combined synchrotron powder X-ray diffraction and crystal structure prediction; 1: 1 L-ephedrine D-tartrate [J]. *CrystEngComm*, 2013, 15 (10): 1853-1859.
- [7] ZHANG L, JIANG S L, YU Y, et al. Phase transition in octahydro-1, 3, 5, 7-tetranitro-1, 3, 5, 7-tetrazocine (HMX) under static compression; an application of the first-principles method specialized for CHNO solid explosives [J]. *The Journal of Physical Chemistry B*, 2016, 120(44): 11510-11522.
- [8] GHUE V D, NIRWAN A. Role of forcefield in density prediction for CHNO explosives [J]. *Structural Chemistry*, 2018, 29(5): 1375-1382.
- [9] WANG H, WANG Y, LV J, et al. CALYPSO structure prediction method and its wide application[J]. *Computational Materials Science*, 2016, 112: 406-415.
- [10] 王丽莉, 熊鹰, 谢炜宇, 等. 含能晶体密度预测的研究进展[J]. 含能材料, 2020, 28(1): 1-12.
- [11] WANG Li-li, XIONG Ying, XIE Wei-yu, et al. Review of crystal density prediction methods for energetic materials [J]. *Chinese Journal of Energetic Materials*, 2020, 28(1): 1-12.
- [12] QIU L, XIAO H, GONG X, et al. Crystal density predictions for nitramines based on quantum chemistry[J]. *Journal of hazardous materials*, 2007, 141(1): 280-288.
- [13] RICE B M, HARE J J, BYRD E F C. Accurate predictions of crystal densities using quantum mechanical molecular volumes[J]. *The Journal of Physical Chemistry A*, 2007, 111(42): 10874-10879.
- [14] POLITZER P, MARTINEZ J, MURRAY J S, et al. An electrostatic interaction correction for improved crystal density prediction[J]. *Molecular Physics*, 2009, 107(19): 2095-2101.
- [15] POLITZER P, MARTINEZ J, MURRAY J S, et al. An electrostatic correction for improved crystal density predictions of energetic ionic compounds [J]. *Molecular Physics*, 2010, 108(10): 1391-1396.
- [16] YANG C, CHEN J, WANG R, et al. Density prediction models for energetic compounds merely using molecular topology[J]. *Journal of Chemical Information and Modeling*, 2021, 61(6): 2582-2593.
- [17] NGUYEN P, LOVELAND D, KIM J T, et al. Predicting energetics materials' crystalline density from chemical structure by machine learning[J]. *Journal of Chemical Information and Modeling*, 2021, 61(5): 2147-2158.
- [18] SONG S, WANG Y, CHEN F, et al. Machine learning-assisted high-throughput virtual screening for on-demand customization of advanced energetic materials[J]. *Engineering*, 2022, 10: 99-109.
- [19] WENG B, SONG Z, ZHU R, et al. Simple descriptor derived from symbolic regression accelerating the discovery of new perovskite catalysts [J]. *Nature Communications*, 2020, 11(1): 1-8.
- [20] LIU Y, GU Z, HUGHES D J, et al. Understanding mixed mode ratio of adhesively bonded joints using genetic programming (GP)[J]. *Composite Structures*, 2021, 258: 113389.
- [21] LEE C, YANG W, PARR R G. Development of the collessavetti correlation-energy formula into a functional of the electron density[J]. *Physical Review B*, 1988, 37(2): 785-789.
- [22] LIU T, CHEN F. Multiwfn: a multifunctional wavefunction analyzer[J]. *Journal of Computational Chemistry*, 2012, 33(5): 580-592.
- [23] KIM C K, CHO S G, KIM C K, et al. Prediction of densities for solid energetic molecules with molecular surface electrostatic potentials[J]. *Journal of Computational Chemistry*, 2008, 29(11): 1818-1824.