

研究论文

DOI: 10.11949/j.issn.0438-1157.20171453

石油馏分基础数据模型建模方法

梅华^{1,2}, 黄彪², 钱锋¹

(¹华东理工大学化工过程先进控制和优化技术教育部重点实验室, 上海 200237; ²Department of Chemical & Materials Engineering, University of Alberta, Edmonton, Alberta T6G 1R1, Canada)

摘要: 石油馏分属性数据是石油化工生产过程中的重要基础数据, 但是海量的现场数据包含了大量的冗余信息和测量误差, 给化工过程实际生产带来很大的困扰。基于石油馏分状态空间表征法提出一种石油馏分基础数据模型建模方法。该方法通过非负矩阵分解算法得到一组初始基础馏分数据模型并在此基础上采用迭代更新策略, 在保证模型预测精度的前提下尽可能地减少模型库的规模。仿真结果验证了本方法的有效性和实用性, 在石油化工生产过程中具有广阔的应用前景。

关键词: 石油馏分; 状态空间表征; 基础馏分数据模型; 非负矩阵分解; 迭代优化

中图分类号: TE 622

文献标志码: A

文章编号: 0438—1157 (2018) 03—0931—05

Modeling basic fraction data of petroleum distillation

MEI Hua^{1,2}, HUANG Biao², QIAN Feng¹

(¹Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China; ²Department of Chemical & Materials Engineering, University of Alberta, Edmonton, Alberta T6G 1R1, Canada)

Abstract: Properties of petroleum fractions are important data of petrochemical processes. However, tremendous on-site data containing redundant information and measurement errors pose a great challenge to routine operation of chemical processes. A basic fraction data modelling method was proposed from characterization techniques of state space of petroleum fractions, in which an initial basic fraction data model was obtained via non-negative matrix factorization and updated by an iterative strategy so that the scale of the model base set was minimized as much as possible under circumstance of assured modelling accuracy. The results of simulation study verify that the proposed method is effective and suitable for a wide application to petrochemical processes.

Key words: petroleum fractions; state space representation; basic fraction data model; non-negative matrix factorization; iterative optimization

引言

石油及其衍生产品是现代化学工业和能源行业的基础, 石油化工行业是整个国民经济领域的支

柱产业之一。石油及其产品的主要成分是各类碳氢化合物, 由于原油产地和加工工艺不同等因素使得即使是同类石油馏分产品的内部分子组成也存在非常大的差异, 而这对后续生产单元的运行状况与经

2017-11-01 收到初稿, 2017-11-07 收到修改稿。

联系人及第一作者: 梅华 (1977—), 男, 博士, 讲师。

基金项目: 国家自然科学基金重点项目 (61333010); 国家自然科学基金优秀青年基金项目 (61422303); 国家留学基金委资助项目 (201606745013)。

Received date: 2017-11-01.

Corresponding author: MEI Hua, hmei2006@ecust.edu.cn

Foundation item: supported by the National Natural Science Foundation of China (61333010, 61422303) and the China Scholarship Council (201606745013).

济效益都会产生巨大的影响。例如石脑油是重要的石油精炼中间产物,也是乙烯工业、汽油调合等石化工业重要的生产原料。通常直馏石脑油包含 40%~70% (质量分数,下同) 的饱和烷烃、20%~50% 的环烷烃、5%~20% 的芳烃以及不到 5% 的烯烃;而 FCC 裂解石脑油中的烯烃能够达到 20% 左右^[1]。对于乙烯工业来说,通常要求石脑油中的烷烃含量尽可能多,而芳烃含量尽可能少^[2]。但对于汽油调合过程来说,则希望芳烃含量较多以提高汽油的辛烷值^[3]。因此,为保证石化工业生产装置的生产安全,优化生产工艺流程和运行参数,提高整体运行效益,需要对石油烃类生产原料进行细致的分析分类,并建立较为完备的原料特性模型库。

石油馏分的特性包括密度、黏度、馏程、PONA 值等宏观属性和分子属性两大类^[4]。宏观属性的优点是比较好测量,但只能粗略地表征石油烃类的宏观特性。分子属性从分子的层面刻画石油烃类的各项性质,但由于石油馏分内的分子组分非常复杂,只能依靠各类色谱或质谱等分析仪表进行分析,而且分析过程需要耗费大量的人力和时间,分析设备也比较昂贵^[5]。因此,通过数据分析的方法建立石油烃类的分子模型是国内外研究的重点。石油馏分分子建模的复杂性在于石油馏分内的分子种类和分子结构千变万化,为方便刻画这些分子组成并简化建模过程的计算量需要采用合理的分子模型形式。Quann 等^[6]提出一种结构导向集总模型(SOL),该模型用向量来表示分子,向量的元素代表了分子中不同分子结构的个数。Ghosh 等^[7-8]将 SOL 模型用于估算柴油的十六烷值和汽油的辛烷值。邱彤等^[9]在 SOL 模型基础上利用 Monte Carlo 方法随机生成 1000 个虚拟分子,并根据信息熵最大化原则调整虚拟分子的含量,从而实现石油馏分的分子重构。同分异构分子矩阵模型(MTHS)由 Peng^[10]于 1999 年提出,该模型用一个矩阵的形式表征石油烃,其中矩阵的行表示分子中的碳原子数,而矩阵的列表示分子结构特征(同分异构体),矩阵的每一个元素被视为一个虚拟组分。目前 MTHS 模型在分子重构^[11-13]、石化过程分子管理与优化^[14-15]方面获得较好的应用。MTHS 模型的最大特点在于从几何的角度刻画了石油馏分分子组成,即在一个三维空间中用一系列离散点来刻画一种不同分子在空间中分布情况。然而,当需要比较不同组成的石油烃,上述三维空间并不能简单明了地表征这

些石油馏分分子分布的差异性。Mei 等^[16-17]通过将石油馏分内各种分子定义为状态变量,从而构建了一个多维的状态空间,任何一种石油馏分都是这个空间内的一个点。在这一状态空间表征方法的基础上,又进一步地提出基础油品的概念,即对于某一大类石油馏分可以视为少数几种基础油品按特定比例混合得到的产物。根据这一概念,石油馏分基础数据库内的样品数量可以大大减少,并且减少了样品中可能存在的冗余信息与测量误差。

本文在文献[17]的基础上,提出一种基础馏分数据模型建模方法,该方法主要是为了解决原有基础数据库无法准确刻画新的样品特性时,如何对原有基础数据库进行扩展的问题。

1 石油馏分的状态空间表征

假设某石油馏分中含有 n 种成分,分别用 n 个单位向量来表示这些成分,即有

$$\begin{aligned} \text{成分} 1^{\#}: \mathbf{e}_1 &= [1, 0, \dots, 0]^T \\ \text{成分} 2^{\#}: \mathbf{e}_2 &= [0, 1, \dots, 0]^T \\ &\vdots \\ \text{成分} n^{\#}: \mathbf{e}_n &= [0, 0, \dots, 1]^T \end{aligned} \quad (1)$$

显然,这些单位向量张成了一个 n 维欧氏空间 $\mathbf{S} \triangleq \text{span}(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$, 它的一个坐标轴代表了一种物质成分,其长度即为该物质的量(物质的量、质量或体积)。令石油馏分中各物质成分的量分别为 $w_1, w_2, \dots, w_n$, 因此该石油馏分可以表示为如下向量和的形式

$$\begin{aligned} \mathbf{v} &= \sum_{i=1}^n w_i \mathbf{e}_i, (w_i \geq 0) = \\ &(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)(w_1, w_2, \dots, w_n)^T = \\ &(w_1, w_2, \dots, w_n)^T \end{aligned} \quad (2)$$

令 $\|\mathbf{v}\| = \sum_{i=1}^n |w_i|$, 则有

$$\bar{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|} = \sum_{i=1}^n \frac{w_i}{\sum_{i=1}^n |w_i|} \mathbf{e}_i = \sum_{i=1}^n \bar{w}_i \mathbf{e}_i \quad (3)$$

式中, $\bar{w}_i$ 为摩尔分数、质量分数或体积分。

令 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ 为不同的石油馏分,其状态空间表达式分别为 $\mathbf{v}_k = \sum_{i=1}^{n_k} w_{k,i} \mathbf{e}_{k,i}, \forall k=1$ 。若所有石油馏分的内部组成都相同,即有 $\mathbf{e}_{1,i} = \mathbf{e}_{2,i} = \dots = \mathbf{e}_{k,i} = \mathbf{e}_i$ 和 $n_1 = n_2 = \dots = n_k = n$, 则这些石油馏分混合物的表达式为

$$\begin{aligned} \mathbf{v} &= \sum_{k=1}^m \mathbf{v}_k = \sum_{k=1}^m \sum_{i=1}^n w_{i,k} \mathbf{e}_i = \\ &\sum_{i=1}^n \left(\sum_{k=1}^m w_{i,k} \right) \mathbf{e}_i = \sum_{i=1}^n w_i \mathbf{e}_i \end{aligned} \quad (4)$$

对式(4)进行归一化,有

$$\begin{aligned} \bar{\mathbf{v}} &= \sum_{k=1}^m \frac{\mathbf{v}_k}{\|\mathbf{v}\|} = \sum_{k=1}^m \frac{\|\mathbf{v}_k\|}{\|\mathbf{v}\|} \bar{\mathbf{v}}_k = \\ &\sum_{k=1}^m \sum_{i=1}^n h_k \bar{w}_{i,k} \mathbf{e}_i = \end{aligned} \quad (5)$$

$$\begin{bmatrix} \bar{w}_{1,1} & \bar{w}_{1,2} & \cdots & \bar{w}_{1,m} \\ \bar{w}_{2,1} & \bar{w}_{2,2} & \cdots & \bar{w}_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{w}_{n,1} & \bar{w}_{n,2} & \cdots & \bar{w}_{n,m} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{bmatrix}$$

其中, $h_k \triangleq \|\mathbf{v}_k\|/\|\mathbf{v}\|$, 且 $\sum_{k=1}^m h_k = 1$ 。由式(5)可知,

任意石油馏分可以看作是石油馏分 $\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2, \dots, \bar{\mathbf{v}}_m$ 按照一定比例混合得到的产物。如果 $\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2, \dots, \bar{\mathbf{v}}_m$ 是线性无关的, 也即是 $\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2, \dots, \bar{\mathbf{v}}_m$ 中任意一个石油馏分不能由其他 $m-1$ 个石油馏分混合得到, 则 $\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2, \dots, \bar{\mathbf{v}}_m$ 被称为状态空间 $\mathcal{S}$ 内的一组基础馏分。

若 $\mathbf{e}_{k,i}$ 是不同的组成, 则 $\{\mathbf{e}_{k,1}, \mathbf{e}_{k,2}, \dots, \mathbf{e}_{k,n_k}\}$ 构成了状态空间中的 m 个子空间, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ 分别是这些子空间中的一个向量, 且 $\mathbf{v}_1 \perp \mathbf{v}_2 \perp \dots \perp \mathbf{v}_m$ 。 $\mathbf{v}_k$ 可以被认为是由 $\{\mathbf{e}_{k,1}, \mathbf{e}_{k,2}, \dots, \mathbf{e}_{k,n_k}\}$ 构成的虚拟组分。显然, 其组分 $w_{k,i}$ 不同, 虚拟组分 $\mathbf{v}_k$ 所表现出来的宏观属性也不相同。因此, 在利用虚拟组分进行石油馏分分子重构时, 可根据宏观属性的要求按照不同的集总规则, 例如根据实沸点曲线^[18]、质量分数^[19]或摩尔分数^[20]、碳原子数^[21]和反应动力学特性^[22-23]等, 选择合适的分子集及其组分。

2 基础馏分数据模型建模

2.1 初始模型的建立

石油馏分是由大量碳氢化合物分子构成, 即使是同一类石油馏分, 由于内部分子组分的差异也会表现出不同的物理化学性质。在石化工业应用领域, 石油馏分分子管理是进行生产流程优化的基础^[24], 但是海量的石油馏分分子组分信息的存储和使用是一项极大的挑战。同类石油馏分的不同样本之间存在着某种相似性但又不完全相同, 这意味着通过工业采集方法得到的石油馏分样本中包含了大量的冗余信息以及测量误差。因此, 从实际应用的角度出发, 需要对已有的样本数据进行适当地处理, 降低样本数据的维度, 同时去除数据中的冗余信息。

由式(5)可知, 假设存在一组基础馏分 $\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2, \dots, \bar{\mathbf{v}}_m$, 现有 N 个石油馏分样本数据 $\bar{\mathbf{v}}^{(1)}, \bar{\mathbf{v}}^{(2)}, \dots, \bar{\mathbf{v}}^{(N)}$, 则它们的状态空间表示为

$$\begin{aligned} \mathbf{V} &\triangleq [\bar{\mathbf{v}}^{(1)}, \bar{\mathbf{v}}^{(2)}, \dots, \bar{\mathbf{v}}^{(N)}] = \\ &\begin{bmatrix} \bar{w}_1^{(1)} & \bar{w}_1^{(2)} & \cdots & \bar{w}_1^{(N)} \\ \bar{w}_2^{(1)} & \bar{w}_2^{(2)} & \cdots & \bar{w}_2^{(N)} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{w}_n^{(1)} & \bar{w}_n^{(2)} & \cdots & \bar{w}_n^{(N)} \end{bmatrix}_{n \times N} = \\ &\begin{bmatrix} \bar{w}_{1,1} & \bar{w}_{1,2} & \cdots & \bar{w}_{1,m} \\ \bar{w}_{2,1} & \bar{w}_{2,2} & \cdots & \bar{w}_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{w}_{n,1} & \bar{w}_{n,2} & \cdots & \bar{w}_{n,m} \end{bmatrix}_{n \times m} \begin{bmatrix} h_1^{(1)} & h_1^{(2)} & \cdots & h_1^{(N)} \\ h_2^{(1)} & h_2^{(2)} & \cdots & h_2^{(N)} \\ \vdots & \vdots & \ddots & \vdots \\ h_m^{(1)} & h_m^{(2)} & \cdots & h_m^{(N)} \end{bmatrix}_{m \times N} = \mathbf{WH} \end{aligned} \quad (6)$$

式中, $\mathbf{W}$ 为基础馏分数据矩阵, $\mathbf{H}$ 为样本数据对应该组基础馏分的混合系数矩阵。

式(6)表明, 通过对已知的石油馏分样本数据进行最大秩矩阵分解即可得到基础馏分品数据模型。这里由于石油馏分分子组分和混合系数都为非负的实数, 因此必须采用非负矩阵分解(non-negative matrix factorization, NMF)方法求解。NMF 算法的一般表达式为

$$(\mathbf{W}, \mathbf{H}) = \arg \min_{\mathbf{W} \geq 0, \mathbf{H} \geq 0} \|\mathbf{V} - \mathbf{WH}\|_F^2 \quad (7)$$

由于 $\mathbf{W}$ 和 $\mathbf{H}$ 都是未知的, 因此式(7)是一个 NP-hard 问题^[25]。Lee 等^[26-27]将其视为一类非凸优化问题, 并提出一种乘性迭代算法, 其迭代规则如式(8)所示。

$$\begin{aligned} \mathbf{H} &\leftarrow \mathbf{H} \cdot * (\mathbf{W}^T (\mathbf{V} / (\mathbf{WH}))) \\ \mathbf{W} &\leftarrow \mathbf{W} \cdot * ((\mathbf{V} / (\mathbf{WH})) \mathbf{H}^T) \\ \mathbf{W}_{m \times n} &\leftarrow \mathbf{W}_{m \times n} / \sum_m \mathbf{W}_{m \times n} \end{aligned} \quad (8)$$

然而, 这种乘性迭代算法的结果受迭代初始值的影响很大, 并且不能保证所得到的结果是最优且唯一的^[28], 为提高 NMF 算法的精确性, 可采用多次平方差最小化测试^[29-30]或分步算法^[31]等。目前关于 NMF 算法的鲁棒性和收敛性仍然是有待进一步深入研究的问题。

2.2 模型的迭代更新策略

NMF 算法计算结果的不唯一性使得每次迭代运算得到的基础馏分都不相同, 这会严重影响基础馏分数据模型的稳定性, 会对石油馏分样本对应的混合系数及其物化性质的预测结果造成不利影响。因此, 在获得一组基础馏分数据模型后应保持已有模型不变。当需要更新基础馏分数据模型以适应新的石油馏分样本分析数据时, 应在保

持原基础馏分数据的基础上添加新的基础馏分数据模型。

基于这一要求,假设已获得一组基础馏分数据矩阵 W_0 , 现有一批新的样本分析数据 $\bar{v}^{(1)}, \bar{v}^{(2)}, \dots, \bar{v}^{(M)}$ 。首先, 需要判断新的石油馏分样本是否能够由已有的基础馏分模型混合得到, 即判断如下约束最小二乘问题是否有解

$$\begin{aligned} h^{(i)} &= \arg \min \| \bar{v}^{(i)} - W_0 h^{(i)} \|_2^2, i=1, \dots, m \\ \text{s.t. } h^{(i)} &\geq 0 \\ \| h^{(i)} \|_1 &= 1 \end{aligned} \quad (9)$$

若式 (9) 有解, 则表明针对样本 $\bar{v}^{(i)}$ 基础馏分数据模型 W_0 不需要扩展; 否则, 就需要增加新的基础馏分模型。一种简单的方法是对 $\bar{v}^{(1)}, \bar{v}^{(2)}, \dots, \bar{v}^{(M)}$ 进行非负矩阵分解, 得到

$$V_1 = W_1 H_1 = [W_0 | W_1] \begin{bmatrix} 0 \\ H_1 \end{bmatrix} \quad (10)$$

则新的基础馏分数据模型为 $[W_0 | W_1]$ 。但这种方法的不足之处在于 W_0 和 W_1 可能蕴含相同的分子组成信息。为克服这一不足, 可采用如下迭代更新基础馏分数据模型的策略。

求解如下最小二乘问题

$$\begin{aligned} h^{(i)} &= \arg \min \| \bar{v}^{(i)} - W_{i-1} h^{(i)} \|_2^2, i=1, \dots, m \\ \text{s.t. } h^{(i)} &\geq 0 \\ \| h^{(i)} \|_1 &\leq 1 \\ W_0 h^{(i)} &\leq \bar{v}^{(i)} \end{aligned} \quad (11)$$

若 $h^{(i)}$ 有非零解且 $\| h^{(i)} \|_1$ 远小于 1, 则新的基础馏分模型为

$$\bar{w}_i = (\bar{v}^{(i)} - W_0 h^{(i)}) / (1 - \| h^{(i)} \|_1) \quad (12)$$

若 $\| h^{(i)} \|_1$ 等于或接近于 0 或者无可行解, 则新的基础馏分即为 $\bar{w}_i = \bar{v}^{(i)}$, 更新后的基础馏分模型 $W_i = [W_{i-1} | \bar{w}_i]$ 。若 $\| h^{(i)} \|_1$ 接近于 1, 则基础馏分模型不需要更新。

2.3 仿真实例

已知一组共 35 个石脑油的实沸点 (TBP) 属性, 采用非负矩阵分解方法得到一组基础馏分数据模型 (包含 10 个基础馏分), 其基础馏分模型矩阵如图 1 所示。此时, 又采集得到 24 个石脑油样品 TBP 属性。采用 2.2 节方法求得新的基础馏分数据模型 (在原模型基础上又增加 3 个基础馏分), 如图 2 所示。图 3 为根据新的基础馏分模型估计得到的样品 TBP 属性与实际 TBP 属性之间的平方误差, 并且最大平方误差控制在 3% 以内。

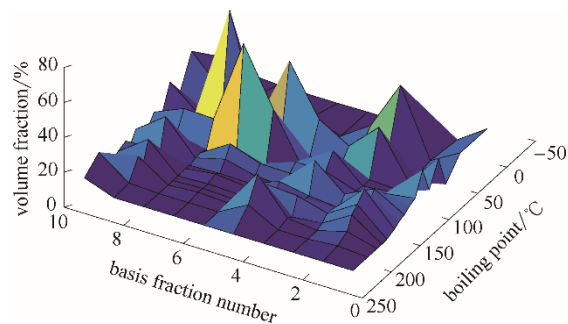


图 1 初始基础馏分分布

Fig. 1 Distribution of initial basis fractions

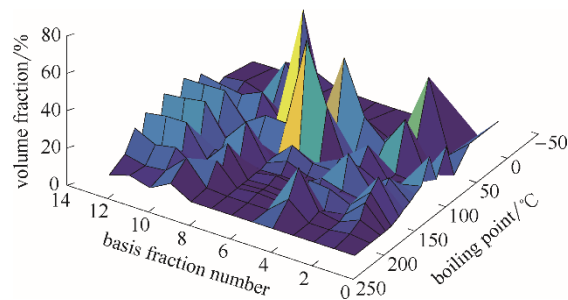


图 2 更新后的基础馏分分布

Fig. 2 Distribution of updated basis fractions

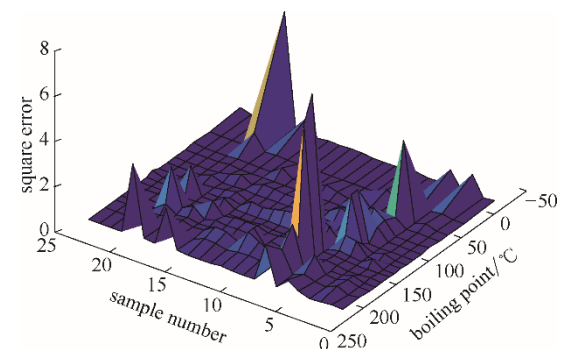


图 3 实沸点估计值与实际值平方误差分布

Fig. 3 Distribution of square errors between predicted and measured true boiling points

3 结 论

石油馏分属性数据是石油化工生产过程中重要的基础数据, 然而海量的现场数据中包含了大量的冗余信息与测量误差, 这给生产过程中石油馏分属性数据的管理与应用带来了很大的困扰。本文在文献[17]工作基础上, 提出了一种基于石油馏分状态空间表征方法的石油馏分基础数据模型建模方法。该方法首先利用非负矩阵分解算法对现有石油馏分属性数据进行筛选, 建立初始的基础馏分数据模型库。当获得新的石油馏分属性数据时, 采用迭代更新的策略对初始基础馏分模型数据库进行更新。这一策略的优点体现为在保证模型精度的前提下能够

有效地减少基础馏分模型数据库的规模,因此在石化生产过程的实际应用中具有广阔的应用前景。最后,通过一组石脑油 TBP 属性数据验证了本文提出的基础馏分数据模型建模方法。仿真结果显示,样品的 TBP 属性估计值与实际值之间的平方误差,并且最大平方误差控制在 3%以内。

References

- [1] ANTOS G J, AITANI A M. Catalytic Naphtha Reforming, Revised and Expanded[M]. Florida: CRC Press, 2004: 2-3.
- [2] 王松汉. 乙烯装置技术与运行[M]. 北京: 中国石化出版社, 2009: 41-43.
WANG S H. Technology and Operation of Ethylene Plants[M]. Beijing: Sinopec Press, 2009: 41-43.
- [3] 蔡智, 黄维秋, 李伟民, 等. 油品调合技术[M]. 北京: 中国石化出版社, 2006: 57-58.
CAI Z, HUANG W Q, LI W M, *et al.* Petroleum Blending Technology[M]. Beijing: Sinopec Press, 2006: 57-58.
- [4] RIAZI M. Characterization and Properties of Petroleum Fractions[M]. Philadelphia: ASTM International, 2005: 10-12.
- [5] VAN GEEM K M, HUDERBINE D, REYNIEERS F, *et al.* Molecular reconstruction of naphtha steam cracking feedstocks based on commercial indices[J]. Computer & Chemical Engineering, 2007, **31**: 1020-1034.
- [6] QUANN R J, JAFFE S B. Structural-oriented lumping: describing the chemistry of complex mixtures[J]. Industrial & Engineering Chemistry Research, 1992, **31**: 2483-2497.
- [7] GHOSH P, HICKEY K J, JAFFE S B. Development of a detailed gasoline composition-based octane model[J]. Industrial & Engineering Chemistry Research, 2006, **45**(1): 337-345.
- [8] GHOSH P. Predicting the effect of cetane improvers on diesel fuels[J]. Energy & Fuel, 2008, **22**(2): 1073-1079.
- [9] 邱彤, 陈金财, 方舟. 基于结构导向集总的石油馏分分子重构模型[J]. 清华大学学报(自然科学版), 2016, **56**(4): 424-429.
QIU T, CHEN J C, FANG Z. Molecular reconstruction model for petroleum fractions based on structure oriented lumping[J]. Journal of Tsinghua University (Science and Technology), 2016, **56**(4): 424-429.
- [10] PENG B. Molecular modelling of refinery processes[D]. Manchester: University of Manchester Institute of Science and Technology, 1999.
- [11] ZHANG Y. A molecular approach for characterization and property predictions of petroleum mixtures with applications to refinery modelling[D]. Manchester: University of Manchester, 1999.
- [12] AYE M, ZHANG N. A novel methodology in transforming bulk properties of refining streams into molecular information[J]. Chemical Engineering Science, 2005, **60**: 6702-17.
- [13] GOMEZ-PRADO J, ZHANG N, THEODOROPOULOS C. Characterization of heavy petroleum fractions using modified molecular-type homologous series (MTHS) representation[J]. Energy, 2008, **33**: 974-987.
- [14] WU Y. Molecular management for refining operations[D]. Manchester: University of Manchester, 2010.
- [15] LIU L. Molecular characterization and modelling for refining processes[D]. Manchester: University of Manchester, 2015.
- [16] 梅华, 杜玉鹏, 王振雷, 等. 基于分子同系物向量表示的石脑油特征提取方法[J]. 清华大学学报(自然科学版), 2016, **56**(7): 723-727.
MEI H, DU Y P, WANG Z L, *et al.* Naphtha characterization based on a molecular-type homologous vector representation[J]. Journal of Tsinghua University (Science and Technology), 2016, **56**(7): 723-727.
- [17] MEI H, CHENG H, WANG Z, *et al.* Molecular characterization of petroleum fractions using state space representation and its application for predicting naphtha pyrolysis product distributions[J]. Chemical Engineering Science, 2017, **164**: 81-89.
- [18] ECKERT E, VANEK T. New approach to the characterisation of petroleum mixtures used in the modelling of separation processes[J]. Computers & Chemical Engineering, 2005, **30**: 343-356.
- [19] HONG K S. Lumped-component characterization of crude oils for compositional simulation[C]//SPE Enhanced Oil Recovery Symposium. USA: Society of Petroleum Engineers, 1982: 4-7.
- [20] LEE S, JACOBY R H, CHEN W H, *et al.* Experimental and theoretical studies on the fluid properties required for simulation of thermal processes[J]. Society of Petroleum Engineers Journal, 1981, **21**(5): 535-550.
- [21] WHITSON C H. Characterizing hydrocarbon-plus fractions[J]. Society of Petroleum Engineers Journal, 1983, **23**(4): 683-694.
- [22] RANZI E, DENTE M, GOLDANIGA A, *et al.* Lumping procedures in detailed kinetic modeling of gasification, pyrolysis, partial oxidation and combustion of hydrocarbon mixtures[J]. Progress in Energy and Combustion Science, 2001, **27**(1): 99-139.
- [23] RANZI E, PIERUCCI S, DENTE M, *et al.* Correct molecular reconstruction of cracking feeds: a need for the accurate prediction of ethylene yields[J]. Chemical Engineering Transaction, 2015, **43**: 871-876.
- [24] WU Y, ZHANG N. Molecular management of gasoline streams[J]. Chemical Engineering Transaction, 2009, **18**: 749-754.
- [25] VAVASIS S A. On the complexity of nonnegative matrix factorization[J]. SIAM Journal on Optimization, 2009, **20**(3): 1364-1377.
- [26] LEE D D, SEUNG S H. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, **401**(6755): 788-791.
- [27] LEE D D, SEUNG S H. Algorithms for nonnegative matrix factorization[C]//Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference. London: MIT Press, 2000: 556-562.
- [28] KUMAR A, SINDHWANI V, KAMBADUR P. Fast conical hull algorithms for near separable non-negative matrix factorization[C]//Proceedings of the 30th International Conference on Machine Learning. USA: ACM, 2013, **28**: 231-239.
- [29] BITTORF V, RECHT B, RE C, *et al.* Factoring nonnegative matrices with linear programs[J]. Advances in Neural Information Processing Systems, 2012, **2**: 1214-1222.
- [30] TEPPER M, SAPIRO G. Compressed nonnegative matrix factorization is fast and accurate[J]. IEEE Trans. Signal Process, 2016, **64**(9): 2269-2283.
- [31] WANG D, NIE F, HUANG H. Fast robust non-negative matrix factorization for large-scale human action data clustering[C]//Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI-16). USA: ACM, 2016: 2104-2110.