

基于分子同系物向量表示的石脑油特征提取方法

梅 华, 杜玉鹏, 王振雷, 钱 锋

(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)

摘 要: 根据石脑油馏分的特点, 提出一种同系物向量表征方法。该方法将石脑油内部的每一个同系物分子定义为一个状态变量, 这些状态变量构成了一个多维的向量空间。因此, 任意一种石脑油都对应着该向量空间内的一个点, 并且能够由该向量空间内的一组相互独立的石脑油馏分(基础油品)线性表示。在此基础上, 提出一种基于非负矩阵分解(non-negative matrix factorization, NMF)方法的基础油品选择方法。该方法将具有较高维数的石脑油样品数据矩阵分解为一个较低维数的特征矩阵及其系数矩阵。在研究实例中, 从 59 组石脑油样本数据中可以提取出 21 组基础油品, 并且由它们还原得到的石脑油模型与样本数据相比, 其相对误差不超过原数据的 2.5%。

关键词: 石脑油; 详细族组成; 分子同系物向量表示; 非负矩阵分解(NMF)

中图分类号: TE 622

文献标志码: A

文章编号: 1000-0054(2016)07-0723-05

DOI: 10.16511/j.cnki.qhdxxb.2016.24.022

Naphtha characterization based on a molecular-type homologous series vector representation

MEI Hua, DU Yupeng, WANG Zhenlei, QIAN Feng

(Key Laboratory of Advanced Control and Optimization for Chemical Processes of the Ministry of Education, East China University of Science and Technology, Shanghai 200237, China)

Abstract: A novel homologous series vector representation method was developed for naphtha in which each homologous molecule of naphtha is defined as a state variable and all these variables are then used to construct a high dimension vector space. Thus, any variation of naphtha as one point in this vector space can be blended linearly by a group of independent naphthas named Basis Oils. These basis oils are obtained using the non-negative matrix factorization (NMF) method with the components data matrix of a huge number of naphtha samples factorized into a characteristic matrix with a lower dimension and its coefficient matrix. In a case study, a naphtha model containing 21 groups of naphtha bases was extracted from 59 groups of naphtha samples with a maximum representation error of less than 2.5 percent of the original data.

Key words: naphtha; detailed group components; molecular-type homologous series vector representation; non-negative matrix factorization (NMF)

石脑油是一种重要的原油蒸馏产品, 被广泛用作芳烃重整、蒸汽裂解制烯烃的主要原料之一。然而, 由于原油来源及生产条件的不同, 使得不同来源或批次的石脑油馏分的内部分子组成及特性差异非常大, 因此在工业乙烯生产过程中需要尽可能地获取石脑油内分子组成的信息。目前, 最直接的获取裂解原料组成信息的方法是通过气相色谱(GC)、二维气相色谱(2D-GC)^[1]或者气相色谱-质谱联用(GC-MS)^[2]等现代分析技术进行组分分析测量。然而这些分析方法的普遍缺点是耗时太长, 并且设备价格昂贵, 不便于工业现场生产应用。

为克服分析仪器耗时长缺点, 一种解决办法是采用分子重构等软测量技术来对石油烃进行特性刻画, 即通过石油烃的一些诸如碳氢比、族组成(paraffins, iso-paraffins, olefins, naphthenes and aromatics, PIONA)值、沸点曲线等易测参数得到石油烃的分子组成信息^[3-5]。传统的石油烃特性描述方法包括虚拟组分方法^[6]、平均结构参数方法和混合物聚类分析方法^[7-8], 它们具有分析速度快的优点, 但由于采用集总组成使得它们无法表征某些具有复杂结构的石油烃混合物, 因而可扩展性不高。1999 年, Peng^[9]提出一种分子型同系物(molecular-type homologous series, MTHS)表征法, 该方法用矩阵的形式来表征石油烃组分, 其中矩阵的行代表分子结构中的碳原子数, 矩阵的列代

收稿日期: 2015-08-20

基金项目: 国家自然科学基金重点基金资助项目(61333010);
国家自然科学基金青年基金资助项目(21206037)

作者简介: 梅华(1977—), 男, 讲师。

E-mail: hmei2006@ecust.edu.cn

表同系物(如链烷烃、环烷烃、芳烃等),而矩阵中的元素表示每一种组分的摩尔分率或质量分率。基于 MTHS 表征法,假设任意一种未知的石油烃可以被认为是几种已知的石油烃按照某个比例混合而成。Zhang 等^[10]和 Ahmad 等^[11]通过对比已知的石油烃内各组分的沸点及密度与整体性质之间的关系,从而得到未知石油烃的 MTHS 矩阵,进而预测出该石油烃的整体性质。

MTHS 矩阵表征法的特点在于将石油烃内的分子组成分布情况在一个三维空间内直观地呈现出来。然而石脑油包含有几十种分子同系物,同系物分布的不同将使得石脑油表现出不同的特性,这种差异性在三维空间内很难进行定量刻画。因此,本文提出一种石脑油向量表征方法,在此基础上运用非负矩阵分解方法提取石脑油分子组成分布的共性特征。

1 石脑油馏分的向量表征

石脑油通常由从碳三到碳十二的各类碳氢化合物分子混合而成。根据族组成分布并且对异构体做集总化处理,将这些组成分别记为:正构烷烃 \mathbf{nP}_i , 异构烷烃 \mathbf{iP}_i , 烯烃 \mathbf{O}_i , 环烷烃 \mathbf{Na}_i 和芳烃 \mathbf{A}_i , ($i=3\sim 12$)。由于这些分子组成具有不同的物理化学特性,因此可以将它们视为互为正交的状态变量,显然它们构成了一组正交基,并张成了一个 n 维的状态空间(记为 $\mathbf{S}=(\mathbf{nP}_i, \mathbf{iP}_i, \mathbf{O}_i, \mathbf{Na}_i, \mathbf{A}_i), i=3\sim 12$)。于是,任意一种石脑油 \mathbf{NAP}_k 都可以表示为基底 $\{\mathbf{nP}_i, \mathbf{iP}_i, \mathbf{O}_i, \mathbf{Na}_i, \mathbf{A}_i | i=3\sim 12\}$ 的线性组合,即有

$$\mathbf{NAP}_k = \sum_{i=3}^{12} (a_i^k \cdot \mathbf{nP}_i + b_i^k \cdot \mathbf{iP}_i + c_i^k \cdot \mathbf{O}_i + d_i^k \cdot \mathbf{Na}_i + e_i^k \cdot \mathbf{A}_i). \quad (1)$$

其中系数 $a_i^k, b_i^k, c_i^k, d_i^k, e_i^k$ 分别为各分子组分的摩尔分率或质量分率。

式(1)也可以写成如下向量形式:

$$\begin{aligned} \mathbf{NAP}_k &= (\mathbf{nP}_i \quad \mathbf{iP}_i \quad \mathbf{O}_i \quad \mathbf{Na}_i \quad \mathbf{A}_i) \cdot \\ &\quad (a_i^k, b_i^k, c_i^k, d_i^k, e_i^k)^T \\ &\triangleq \mathbf{S} \cdot \mathbf{w}_k. \end{aligned} \quad (2)$$

其中 $\mathbf{w}_k = (a_i^k, b_i^k, c_i^k, d_i^k, e_i^k)^T$ 为各同系物组分在状态空间 \mathbf{S} 内的坐标向量。为简单起见,可以直接用坐标向量 \mathbf{w}_k 来表征该种石脑油。假设有 m 种石脑油,它们的坐标向量 $\mathbf{w}_k (k=1, 2, \dots, m)$ 互不相关,则这 m 种石脑油则构成了状态空间 \mathbf{S} 内的一个 m 维子空间,并可以表示成式(3)的矩阵形式,即

$$\mathbf{M} = \begin{bmatrix} a_i^1 & a_i^2 & \cdots & a_i^m \\ b_i^1 & b_i^2 & & b_i^m \\ c_i^1 & c_i^2 & & c_i^m \\ d_i^1 & d_i^2 & & d_i^m \\ e_i^1 & e_i^2 & \cdots & e_i^m \end{bmatrix}_{n \times m}. \quad (3)$$

若 $m=n$, 则子空间 \mathbf{M} 即是原料状态空间 \mathbf{S} , 这就意味着任意一种石脑油都是这 m 种石脑油的一个线性组合。换言之,可由这 m 种石脑油按照某种特定比例混合得到任意一种石脑油,因此,这 m 种石脑油也被称为基础油品。由于石脑油中组分种类繁多,要找到整个原料状态空间内的基础油品几乎不可能,但在实际应用中所涉及的石脑油样品的馏程范围相对比较固定。因此,只需要用一个维数较小的原料子空间内来刻画实际应用中所涵盖的石脑油样品特性。

设任意一种石脑油 \mathbf{NAP} 可由这 m 种基础油品按照一定比例混合而成,即有

$$\mathbf{NAP} = \sum_{k=1}^m h_k \cdot \mathbf{NAP}_k = \sum_{k=1}^m \mathbf{S} \cdot (h_k \mathbf{w}_k) = \mathbf{S} \cdot \mathbf{v}, \quad (4)$$

$$\mathbf{v} = \sum_{k=1}^m h_k \mathbf{w}_k = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m] \cdot [h_1, h_2, \dots, h_m]^T = \mathbf{W} \cdot \mathbf{h}^T. \quad (5)$$

其中:矩阵 $\mathbf{W} \triangleq [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m]$ 为基础油品构成的原料状态子空间的矩阵描述; $\mathbf{h} \triangleq [h_1, h_2, \dots, h_m]$ 为石脑油 \mathbf{NAP} 关于基底 \mathbf{W} 的坐标; $h_k (k=1, 2, \dots, m)$ 为各基础油品所对应的混合系数。

2 基于非负矩阵分解 NMF 的石脑油特征提取

假设从工业现场采集到 N 个石脑油样品且 $N \gg m$, 通过实验室化验分析得到的这 N 个石脑油样品的详细组成数据构成一个样本集,记为

$$\mathbf{V} = \begin{bmatrix} a_i^1 & a_i^2 & \cdots & a_i^N \\ b_i^1 & b_i^2 & & b_i^N \\ c_i^1 & c_i^2 & & c_i^N \\ d_i^1 & d_i^2 & & d_i^N \\ e_i^1 & e_i^2 & \cdots & e_i^N \end{bmatrix}_{n \times N} \triangleq [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_N]. \quad (6)$$

将式(5)代入式(6)得到

$$\mathbf{V} = [\mathbf{W} \cdot \mathbf{h}_1^T \quad \mathbf{W} \cdot \mathbf{h}_2^T \quad \cdots \quad \mathbf{W} \cdot \mathbf{h}_N^T] = \mathbf{W}_{n \times m} \cdot \mathbf{H}_{m \times N}. \quad (7)$$

因此,通过对石脑油现场采集样品组分矩阵 \mathbf{V} 进行满秩分解即可得到基础油品的表征向量以及样

品关于这些基础油品的混合系数。通常可以采用奇异值分解(SVD)或者正交分解求解得到矩阵 \mathbf{W} 和 \mathbf{H} 。然而, SVD 分解和正交分解得到的结果往往包含负数, 即基准油品的分子组成摩尔分率以及混合系数可能为负数, 这显然不符合工程实际。因此, 必须采取非负矩阵分解(NMF)方法来求解。

NMF 分解最早由 Lee 和 Seung 提出^[12], 在图像分析、文本聚类、数据挖掘、语音处理等方面得到了广泛应用。NMF 可以用式(8)所示优化问题描述, 即

$$\begin{aligned} & \text{已知 } \mathbf{V} \in \mathbb{R}_+^{n \times N}, \text{ 求解} \\ & \min_{\mathbf{W}, \mathbf{H}} \|\mathbf{V} - \mathbf{W} \cdot \mathbf{H}\|_F, \\ & \text{s. t. } \mathbf{W} \in \mathbb{R}_+^{n \times m}, \quad \mathbf{H} \in \mathbb{R}_+^{m \times N}. \end{aligned} \quad (8)$$

其中: $\mathbb{R}_+^{n \times m}$ 表示 $n \times m$ 维非负实数空间; $\|\cdot\|_F$ 为矩阵的 Euclidean 范数。

式(8)所给出的优化问题迭代求解步骤有如下 4 个步骤:

Step 1 基础油品维数的确定。

对样品组成矩阵 \mathbf{V} 进行 SVD 分解, 所有大于 1 的奇异值个数即为基础油品维数 m 。

Step 2 \mathbf{W} 和 \mathbf{H} 的初始化。

随机给定 \mathbf{W} 和 \mathbf{H} 元素的非负的初始值。

Step 3 根据乘性迭代准则计算 \mathbf{W} 和 \mathbf{H} 。

计算公式为

$$\begin{aligned} \mathbf{H} & \leftarrow \mathbf{H} \cdot (\mathbf{W}^T (\mathbf{V} / (\mathbf{W} \cdot \mathbf{H}))), \\ \mathbf{W} & \leftarrow \mathbf{W} \cdot ((\mathbf{V} / (\mathbf{W} \cdot \mathbf{H})) \cdot \mathbf{H}^T), \\ \mathbf{W}_{m \times n} & \leftarrow \frac{\mathbf{W}_{m \times n}}{\sum_m \mathbf{W}_{m \times n}}. \end{aligned}$$

Step 4 误差判断。

判断 $\|\mathbf{V} - \mathbf{W} \cdot \mathbf{H}\|_F$ 是否小于某个预设值或者迭代次数大于 300 次。如果否, 返回 Step 3 重新计算 \mathbf{W} 和 \mathbf{H} 。

3 实例研究

收集国内某石化企业石脑油连续 2 a 的总共 59 个样品的石脑油详细 PIONA 值分析数据。剔除某些含量极小的同系物, 还剩下 35 个具有显著分布的同系物, 如图 1 所示。从图 1 中不难发现, 有多个样品具有类似的组成分布特征, 这使得样品数据中包含了大量的冗余信息。

根据上一节介绍的 NMF 算法对这 59 种石脑油样本数据进行特征提取, 得到 21 种基础油品及其对应的混合系数矩阵。图 2 为基础油品同系物分布情况, 与图 1 相比, 基础油品之间具有显著的特征差异, 冗余信息大大减少。

图 3 为上述基础油品对应的混合系数分布情况。图 3 中峰值越高, 表明其所对应的基础油品在混合后的石脑油样品中的比例越高。

图 4 为根据基础油品重构得到石脑油组成分布与样品组成分布之间的误差分布。由图 4 可见, 与原始矩阵的误差不超过 2.5%。

需要指出的是, NMF 迭代求解过程是一个非凸优化过程, 其优化结果并不能保证是全局最优的。因此 NMF 优化结果的精度受 \mathbf{W} 和 \mathbf{H} 的初始值影响很大, 可经过多次尝试找到一组满意的初始值。

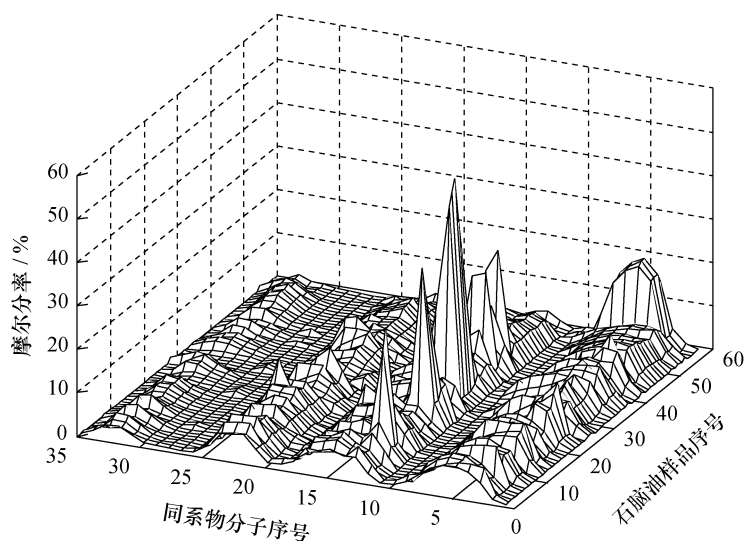


图 1 石脑油样品组成分布情况

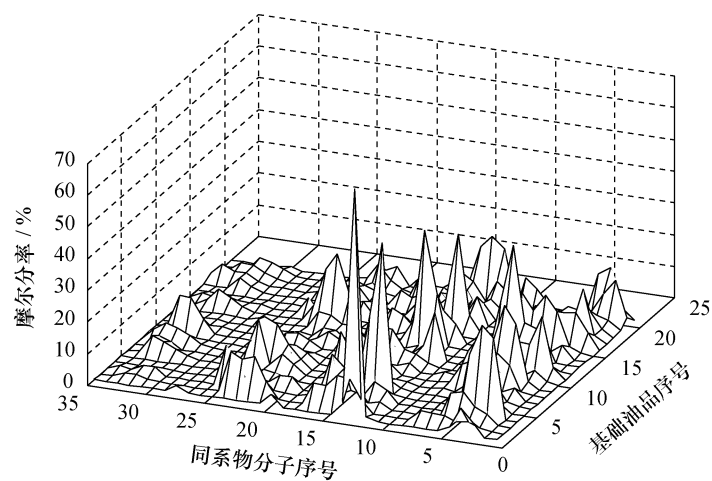


图2 基础油品组成分布情况

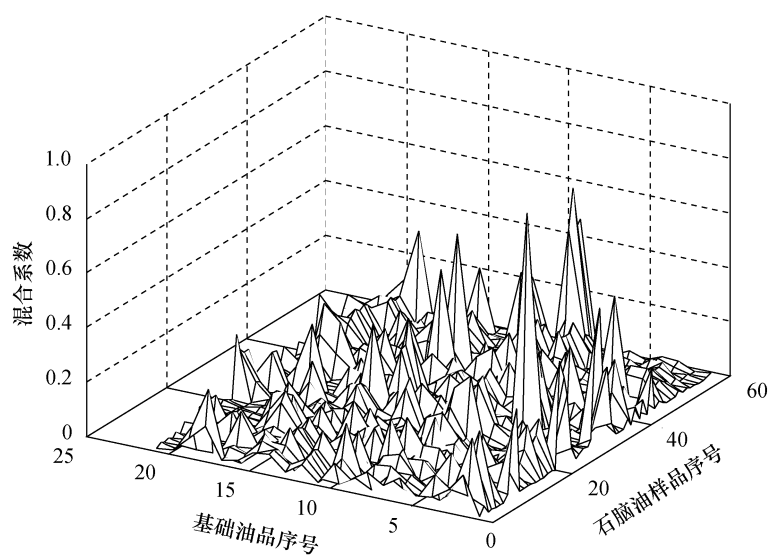


图3 石脑油样品关于基础油品的混合系数分布情况

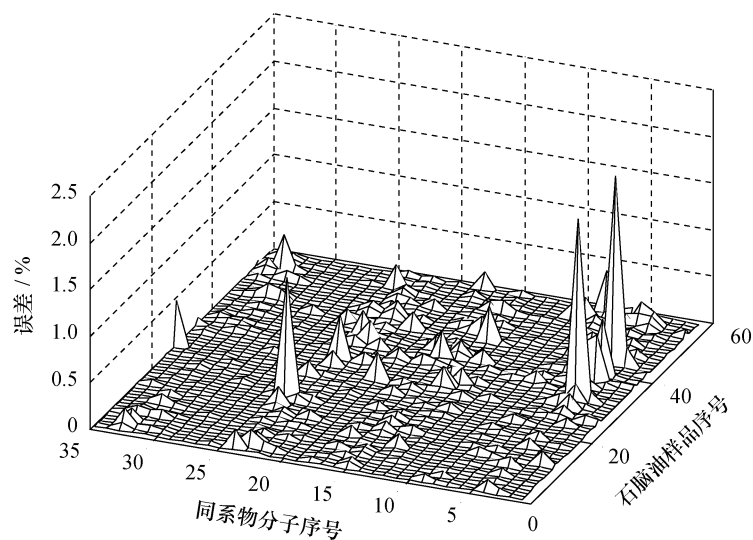


图4 重构油品与样品组成误差分布

4 结 论

石脑油是一种重要的化工原料,广泛应用于芳烃重整、管式蒸汽裂解制烯烃等领域。由于内部组成复杂,使得不同来源及生产批次的石脑油具有显著的物性差异。本文根据石脑油馏分的特点,提出一种同系物向量表征方法,即所有的同系物分子构成一个多维的向量空间,任意一种石脑油都是该向量空间内的一个点,并且能够由该向量空间内的任意一组线性无关向量组(基础油品)线性表示。上述表征方法具有清晰的几何涵义,揭示了石脑油馏分内在的物性特征。在此基础上,本文进一步提出一种基于非负矩阵分解方法的基础油品选择方法。在实例研究中,应用非负矩阵分解算法,一个由59组样本石脑油组成构成的维样本数据矩阵被分解为一个维矩阵和一个维的矩阵,且二者乘积与原始矩阵的相对误差不超过2.5%。这一结果表明,用这21组基础油品按照一定比例进行混合,可以得到任意一种石脑油样品,从而证明了本文所提出的石脑油特征提取方法的有效性和可行性。

参考文献 (References)

- [1] Vendeuvre C, Bertoncini F, Duval J L, et al. Comparison of conventional gas chromatography and comprehensive two-dimensional gas chromatography for the detailed analysis of petrochemical samples [J]. *Journal of Chromatography A*, 2004, **1056**(1-2): 331-347.
- [2] Qian K N, Dechert G J. Recent advances in petroleum characterization by GC field ionization time-of-flight high-resolution mass spectrometry [J]. *Analytical Chemistry*, 2002, **74**(16): 3977-3983.
- [3] Pyl S P, Van Geem K M, Reyniers M, et al. Molecular reconstruction of complex hydrocarbon mixtures: An application of principal component analysis [J]. *AIChE Journal*, 2010, **56**(12): 3174-3188.
- [4] Saine Aye M, Zhang N. A novel methodology in transforming bulk properties of refining streams into molecular information [J]. *Chemical Engineering Science*, 2005, **60**(23): 6702-6717.
- [5] Van Geem K M, Hudebine D, Reyniers M F, et al. Molecular reconstruction of naphtha steam cracking feedstocks based on commercial indices [J]. *Computers and Chemical Engineering*, 2007, **31**(9): 1020-1034.
- [6] Dente M, Ranzi E. Detailed prediction of olefin yields from hydrocarbon pyrolysis through a fundamental simulation program (SPYRO) [J]. *Computers and Chemical Engineering*, 1979, **3**(1-4): 61-75.
- [7] Speight J G. The Chemistry and Technology of Petroleum, [M]. 3rd Ed. New York, USA: Marcel Dekker, 1998.
- [8] Riazi M R. Characterisation and Properties of Petroleum Fractions [M]. Philadelphia, USA: ASTM International, 2005.
- [9] PENG Bin. Molecular Modeling of Refinery Process [D]. Manchester, UK: The University of Manchester, 1999.
- [10] ZHANG Yan. A Molecular Approach for Characterization and Property Predictions of Petroleum Mixtures with Applications to Refinery Modeling [D]. Manchester, UK: The University of Manchester, 1999.
- [11] Ahmad M I, Zhang N, Jobson M. Molecular components-based representation of petroleum fractions [J]. *Chemical Engineering Research and Design*, 2011, **89**: 410-420.
- [12] Lee D, Seung H. Learning the parts of objects by non-negative matrix factorization [J]. *Nature*, 1999, **401**: 788-791.