

# 烃类液体黏度的定量结构性质预测模型

蔡广庆, 张霖宙\*

(中国石油大学(北京)重质油国家重点实验室, 北京 102249)

**摘要:**开发了结构性质关联模型(QSPR),实现了基于烃类化合物的结构特征预测黏温特性的功能。搜集了 254 种烃类化合物不同温度下的黏度数据,选择改进的 Andrade 方程来描述烃类化合物的黏温特性曲线,并通过对实验数据进行回归,获得了化合物的 Andrade 方程参数  $B$  和  $T_0$ 。在此基础上,采用分子质量和 15 个基团作为分子的结构特征参数,建立神经网络模型预测 Andrade 模型参数  $B$  和  $T_0$ ,平均相对误差分别为 3.59% 和 1.27%。通过所预测的 Andrade 模型参数计算化合物的黏温性能,预测值与实验数据相比绝对平均误差为 0.42 mPa·s。

**关键词:** QSPR; 黏温特性; 基团贡献法; 人工神经网络

**中图分类号:** TE622.5

**文献标志码:** A

**文章编号:** 0253-4320(2018)07-0204-04

**DOI:** 10.16606/j.cnki.issn.0253-4320.2018.07.047

## Prediction model of QSPR for hydrocarbon liquid viscosities

CAI Guang-qing, ZHANG Lin-zhou\*

(State Key Laboratory of Heavy Oil Processing, China University of Petroleum, Beijing 102249, China)

**Abstract:** In this paper, a quantitative structure property relationship (QSPR) model is developed to predict the viscosity vs. temperature profile based on structural features of hydrocarbon compounds. The viscosity data sets of 254 hydrocarbon compounds at different temperatures are collected. An improved Andrade equation is chosen to describe viscosity vs. temperature curves of hydrocarbon compounds and two Andrade equation parameters (named  $B$  and  $T_0$ ) of compounds are obtained through regressing the experimental data. On this basis, an associated model is constructed between structural parameters of compounds and Andrade equation parameters. Molecular weights and 15 groups are served as the structure characteristic parameters of molecular to establish an artificial neural network (ANN) model to estimate Andrade model parameters  $B$  and  $T_0$  in terms of structural information of compounds. The average relative errors are 3.59% and 1.27%, respectively. The viscosity vs. temperature performances of compounds are calculated based on the predicted Andrade model parameters and the absolute mean deviation of predicted values is 0.42 mPa·s compared with experimental data.

**Key words:** QSPR; viscosity vs temperature relationship; group contribution method; artificial neural network

液体的黏度是科学研究和工程应用的重要流动特征,通常使用黏度计测量获得,测量工作耗时耗力,许多情况下无法获得目标化合物在目标温度条件下的黏度实验数据。因此,开发可靠的理论模型来预测化合物在不同条件下的黏度,显得尤为重要。

常规的化工应用及产品质量测量中,液体往往处于中压或者低压环境,在此情况下压力对于液体黏度的影响可以忽略,这样液体黏度与温度的变化关系就可以用较为简单的方程进行描述,常见的模型有 Andrade<sup>[1]</sup>方程和 Korsten<sup>[2]</sup>方程。在已有的研究工作中最为流行的为基团贡献法和基于化学描述符的 QSPR (quantitative structure property relationship) 方法。

Joback 等<sup>[3]</sup>提出一个简单的基团贡献法。Simon<sup>[4]</sup>、Klincewicz 等<sup>[5]</sup>、Lyman 等<sup>[6]</sup>提出的基团贡献法也被广泛应用。Velzen 等<sup>[7]</sup>提出了基于 De

Guzman-Andrade 方程的液体黏度、温度和化学结构之间的相关性。Allan 等<sup>[8]</sup>以有效碳原子为参数来预测液体的黏度。基团贡献法虽然简单易用,可以从分子结构直接计算化合物的物理性质,但是采用的模型数学可变性不够强,结构对黏温特性的描述具有局限性。

近年来 QSPR<sup>[9-10]</sup>方法在预测化合物的物理和热力学性质方面受到广泛关注。Kauffman 等<sup>[11]</sup>基于 QSPR 模型,使用多元线性回归和计算神经网络方法预测液体的黏度。QSPR 方法实现了大量化学描述符和先进的数据挖掘方法一起使用,具有较强的可变性和描述性,但是由于许多化学描述符(比如临界参数)本身难以计算,容易产生偏差。

基于上述问题,本文中寻找出决定黏温特性的主要化学特征,并将其作为分子的描述符,构建基于神经网络的 QSPR 模型,实现对烃类化合物在不同温度下黏度的预测。

收稿日期:2017-12-06;修回日期:2018-05-07

基金项目:国家自然科学基金青年基金项目(21506254);中国石油大学(北京)科研基金(2462014YJRC020)

作者简介:蔡广庆(1994-),男,硕士生;张霖宙(1987-),男,副教授,研究方向为重油化学及转化规律,通讯联系人, Lzz@cup.edu.cn。

## 1 方法

图1是预测化合物黏温曲线的流程图。预测过程包括数据收集和准备、化学特性的生成与筛选和QSPR模型构建3个阶段。第一阶段,从API databook<sup>[12]</sup>搜集单体烃的黏温数据,使用多元线性回归法从黏温数据拟合得到改进的Andrade方程中的参数,建立黏度模型参数的数据集;第二阶段,把化学特性放到以所有数据为训练集的ANN模型中,对其进行评估,得到分子的化学特性数据集;最后,采用得到的化学特性数据集和Andrade参数数据集来建立QSPR模型,实现预测烃类的黏温曲线的功能。

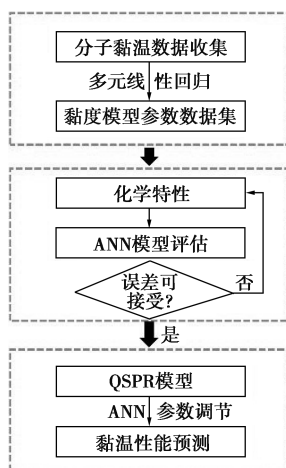


图1 预测单体烃黏温性能流程图

## 2 数据收集

本研究主要的目标应用体系是石油馏分<sup>[13]</sup>,搜集254组单体烃不同温度下的黏度数据,覆盖了从C<sub>3</sub>~C<sub>64</sub>的碳数范围,由5种不同类型的化学族组成,具体分布如表1所示,其中大部分化合物都符合石油中分子的结构特征<sup>[14]</sup>。

表1 黏温性质数据库组成

烃族组成	正构烷烃	异构烷烃	环烷烃	烯烃/炔烃	芳香烃
化合物数量	31	52	33	75	63

用于估计黏度关于温度的函数使用改进的Andrade方程<sup>[15]</sup>:

$$\mu = \exp[B(1/T - 1/T_0)] \quad (1)$$

其中 $B$ 和 $T_0$ 是方程中的2个参数。通过多元线性回归技术对所有化合物的黏温数据进行回归,得到Andrade方程中的 $B$ 和 $T_0$ ,拟合的平均相对误差为0.92%。随后,寻找分子结构和改进Andrade方程

参数 $B$ 和 $T_0$ 之间的相关性。

## 3 化学特性的生成与筛选

为生成和筛选出决定黏温特性的化学特征作为分子的描述符,需寻找影响烃类黏度的一系列基团,并尽可能减少其数量以防止过拟合,故采用基本基团来描述分子的化学特征。基本基团应能代表所有的烃类化合物(包括脂族烃,环烷烃和芳烃)、结构简单、在一个化合物分子中不能与其他基本基团重叠、具有环状和非环状结构之间的区别能力。参考Joback等<sup>[3]</sup>的基团贡献法,定义15个基本基团的集合,这些基本基团已经能够反映分子的结构特征。

采用人工神经网络(ANN)对已生成的基团进行评估<sup>[16]</sup>。所选ANN的结构如图2所示。

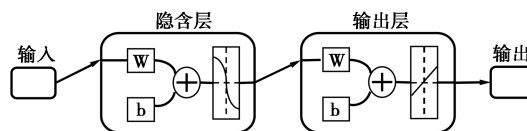


图2 三层神经网络结构图

图2显示了应用的三层神经网络的结构示意图。把基本基团作为ANN的输入,参数 $B$ 和 $T_0$ 作为ANN的输出。将数据库中的所有数据放入ANN训练集中,即训练集是100%,以此来评价基本基团是否有效。最终获得了一系列基本基团(见表2)。

表2 基本基团表

基团信息	评价
—CH <sub>3</sub>	链烷烃中的伯碳
—CH <sub>2</sub> —	链烷烃中的仲碳
$\begin{array}{c} \diagdown \\ \text{CH} \\ \diagup \end{array}$	链烷烃中的叔碳
$\begin{array}{c} \diagdown \\ \diagup \\ \text{C} \\ \diagdown \\ \diagup \end{array}$	链烷烃中的季碳
=CH <sub>2</sub>	烯烃中的端位碳
=CH—	烯烃中连着1个双键和1个单键的碳
=C—	烯烃中连着1个双键和2个单键的碳
=C=	烯烃中连着2个双键的碳
≡CH	炔烃中连着三键的碳
≡C—	炔烃中连着三键和1个单键的碳
—CH <sub>2</sub> —(cyclic)	环状结构的仲碳
$\begin{array}{c} \diagdown \\ \text{CH} \\ \diagup \end{array}$ (cyclic)	环状结构的叔碳
$\begin{array}{c} \diagdown \\ \diagup \\ \text{C} \\ \diagdown \\ \diagup \end{array}$ (cyclic)	环状结构的季碳
=CH—(cyclic)	环状结构中含1个双键和1个单键的碳
=C—(cyclic)	环状结构中含1个双键和2个单键的碳

添加了分子质量作为额外的化学特征,一方面添加分子质量后模型精度有明显提升,另一方面分子质量的添加有助于提高模型的外推性能。

图 3 和图 4 分别展示了采用不同信息输入时,实验和预测的参数  $B$  及  $T_0$  值的准确度。从图 3 可以看出,使用分子质量(a)、基本基团(b)和分子质量+基本基团(c)作为 100% 训练集输入,预测的参数  $B$  的相关系数分别是 0.838 4、0.966 1、0.990 3。由图 3(a)可以看出,仅输入分子质量时,存在一定的趋势变化,但训练结果较差。由图 3(b)所示,以基本基团作为输入时,相关系数得到了较大提升。

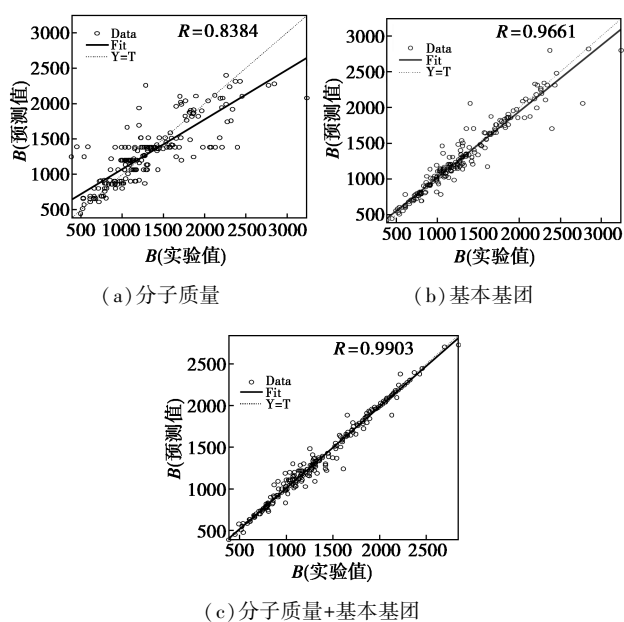


图 3  $B$  预测值和实验值的对比图

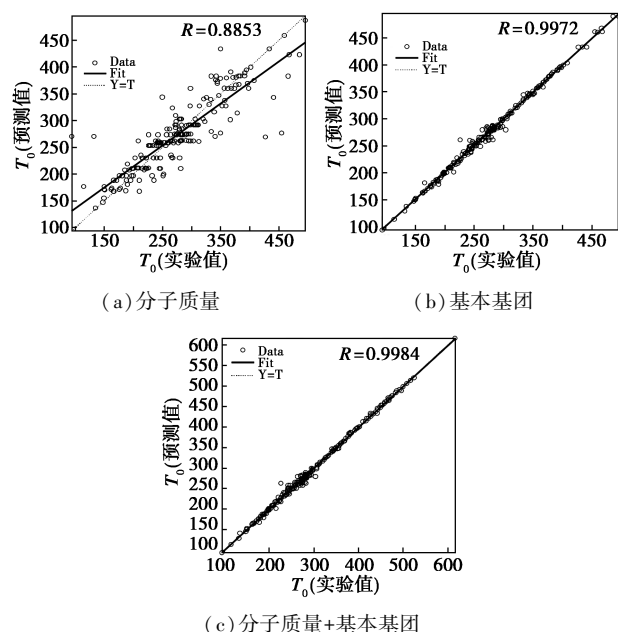


图 4  $T_0$  预测值和实验值的对比图

分子质量和基本基团一起作为输入时,评价模型得到的预测精度进一步提高,可见影响  $B$  参数的特征已经被总结出来。

图 4 展示了不同信息输入情况下  $T_0$  实验和预测值的比较,相关系数分别是 0.885 3、0.997 2、0.998 4,与参数  $B$  的训练过程一样。从图 3 和图 4 中可以看出,随着化学特性的更加详细,相应的相关系数有所增加。所以采用分子质量和基本基团已经能够描述分子的 Andrade 方程参数  $B$  和  $T_0$  的变化情况。

#### 4 QSPR 模型的开发

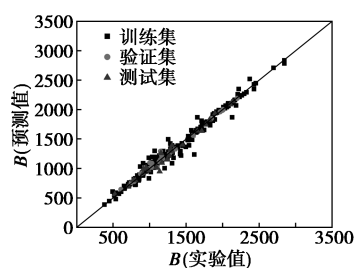
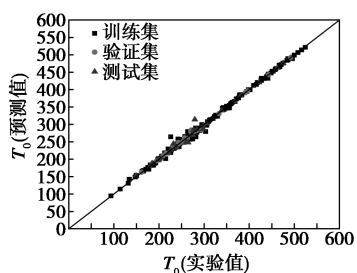
评价模型由于将所有数据放到 ANN 模型训练集中,因此模型仅能反映基团的描述能力,结果存在过拟合的问题,并不能直接应用。在取得化学特征的基础上,开始构建具有泛化能力的 QSPR 模型。

254 个单体烃数据被分为训练集、验证集、测试集,各占总样本数的 80%、10%、10%。为确保验证集和测试集涵盖不同的烃类结构,抽取一定数量的不同结构烃类放入验证集和测试集中。QSPR 模型同样采用 ANN 模型作为基本关联方法,不断调整 ANN 模型中隐含层的层数和数量来寻找最优的参数组合,得到了结构为 16-7-1 的三层神经网络来分别预测参数  $B$  和  $T_0$ 。表 3 列出了神经网络的最优参数组合。

表 3 神经网络的训练参数

训练参数	方法
训练函数	贝叶斯正则化算法
表现函数	相对误差和相关系数
隐含层激活函数	tansig 函数
输出层激活函数	purelin 函数
输入层节点数	16
隐含层节点数	7
输出层节点数	1

通过上述工作,得到了结合基团贡献法和人工神经网络来预测纯烃类化合物黏度的 QSPR 模型。图 5 比较了实验和预测的 Andrade 模型参数  $B$ (a) 和  $T_0$ (b) 的值。对于参数  $B$  的训练集、验证集及测试集,相关系数分别为 0.994 3、0.996 6、0.984 0,平均相对误差为 3.89%、0.82%、1.19%。对于  $T_0$ ,相关系数为 0.998 9、0.998 3、0.991 3,平均相对误差为 1.17%、0.44%、0.54%。从统计结果来看,建立的 ANN 模型的预测性能是可信的。

(a) 参数  $B$ (b) 参数  $T_0$ 图5 Andrade模型参数  $B$  和  $T_0$  实验值与预测值对比图

QSPR模型最终是为了预测不同温度下烃类的液体黏度,图6展示了通过预测Andrade模型参数

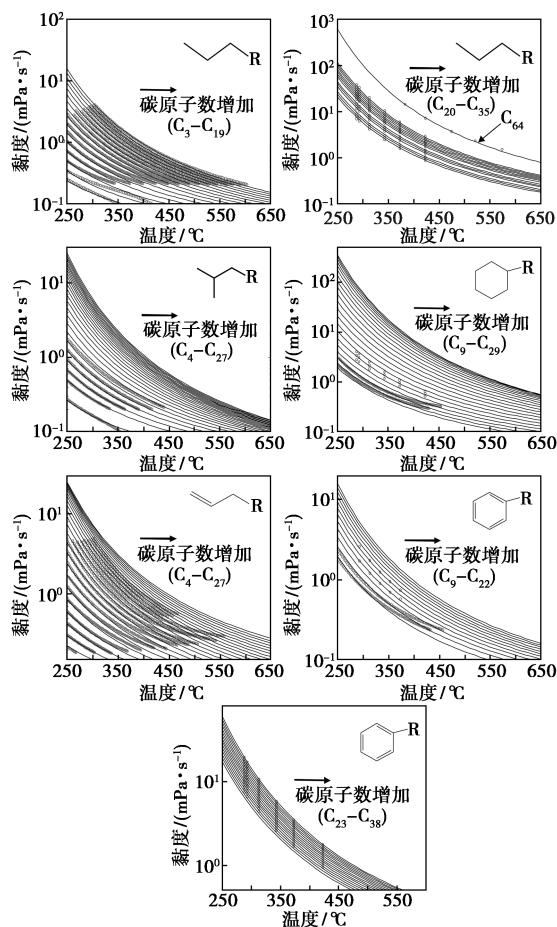


图6 5个系列同系物不同温度下黏度的预测值与实验值对比图

计算得到的5个系列同系物不同温度下黏度预测值和实验值的对比。随着碳原子数的增加,预测值与实验值有很好的—致性。在温度较高区间黏度预测性能变差,但是该区间液体的黏度都小于  $1 \text{ mPa}\cdot\text{s}$ ,这些误差是可以接受的。结果表明,该QSPR模型能够在不同的温度下准确估算纯化合物的黏度。

## 5 结论

选择了16个影响烃类纯化合物液体黏度的化学特征,包括分子质量和基本基团,采用ANN构建了基于结构特征的烃类液体黏度预测的QSPR模型。模型首先通过烃类分子的化学特征,预测Andrade模型的参数  $B$  和  $T_0$ ,平均相对误差分别为3.59%和1.27%。之后再计算任意温度下的液体黏度。单体烃化合物的预测值与实验值有很好的—致性,绝对平均误差为  $0.42 \text{ mPa}\cdot\text{s}$ 。研究结果表明,本文中提出的QSPR模型可以准确预测纯烃类化合物的液体黏度。

## 参考文献

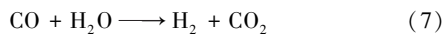
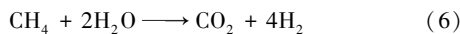
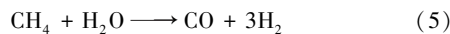
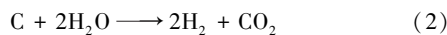
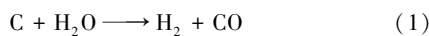
- [1] Reid R C, Prausnitz J M, Poling B E. The properties of gases and liquids [M]. New York: McGraw-Hill, 1987.
- [2] Korsten H. Viscosity of liquid hydrocarbons and their mixtures [J]. Aiche Journal, 2010, 47(2): 453-462.
- [3] Joback K G, Reid R C. Estimation of pure-component properties from group-contributions [J]. Chemical Engineering Communications, 1987, 57(1/2/3/4/5/6): 233-243.
- [4] Simon R H M. Estimation of critical properties of organic compounds by the method of group contributions [J]. Aiche Journal, 1956, 2(3): 12.
- [5] Klinciewicz K M, Reid R C. Estimation of critical properties with group contribution methods [J]. Aiche Journal, 1984, 30(1): 137-142.
- [6] Lyman, Warren J. Handbook of chemical property estimation methods: Environmental behavior of organic compounds [M]. Washington: American Chemical Society (ACS), 1990.
- [7] Velzen D V, Cardozo R L, Langenkamp H. A liquid viscosity-temperature-chemical constitution relation for organic compounds [J]. Ind Eng Chem Fundamen, 1972, 11(1): 20-25.
- [8] Allan J M, Teja A S. Correlation and prediction of the viscosity of defined and undefined hydrocarbon liquids [J]. Canadian Journal of Chemical Engineering, 2010, 69(4): 986-991.
- [9] Suzuki T, Ohtaguchi K, Koide K. Computer-assisted approach to develop a new prediction method of liquid viscosity of organic compounds [J]. Computers & Chemical Engineering, 1996, 20(2): 161-173.

(下转第209页)

## 1 建立模型

### 1.1 石油焦气化制氢机理

石油焦由底部加入分别经历干燥区、裂解区、还原区<sup>[9]</sup>。在干燥区中,石油焦中含有的水分与热源进行热交换,使石油焦中的水分蒸发出去;在裂解区,由于温度较高,石油焦将发生热裂解而析出挥发分,挥发分主要由碳、氢气、甲烷、一氧化碳、二氧化碳、焦油组成;在还原区,水蒸汽与剩余的碳、一氧化碳、二氧化碳发生还原反应,生成一氧化碳、氢气和甲烷等气体。石油焦气化制氢主要考虑的化学反应为:



### 1.2 气化模型

利用 Aspen Plus 软件进行石油焦模拟时,不再考虑气化炉的尺寸大小,热力学平衡为零维,需做如下假设<sup>[10]</sup>:①模拟在稳定状态下进行,所有参数不会随时间发生改变;②石油焦进入气化炉之前进行粉碎处理,粒径极小,不需要考虑内外扩散的影响,且颗粒内温度分布运行,无温度梯度;③在模型内,石油焦除 C 外,其他气化产物全部以气态形式存在,C 元素以固态形式存在;④气化炉内无压力梯度;⑤石油焦中灰分为惰性物体不参加反应;⑥假设气相物质在瞬间和固相物质混合均匀,也就是说所有的气相反应速度都很快,气相和固相的混合是均匀的,瞬间完成并且达到平衡;⑦气化介质与石油焦在气化炉内瞬间混合完毕。

由于石油焦热裂解过程非常复杂,无法获得真

正的化学方程式,所以对于石油焦裂解过程使用产率反应器(RYield),根据石油焦的元素分析与工业分析,依靠 Fortran 语言建立热裂解反应方程式。由于目前对于石油焦气化过程化学反应和动力学都没有一个公认的理论,因此运用吉布斯自由能最小化原则来模拟实际的石油焦气化反应是常用的方法<sup>[11]</sup>。由于吉布斯反应器(RGibbs)根据吉布斯自由能趋于最小值原则,计算同时达到化学平衡和相平衡的系统组成和分布,并可以限制平衡温度来修正模型,所以使用 RGibbs 来模拟石油焦气化反应。建立如图 1 气化模型。

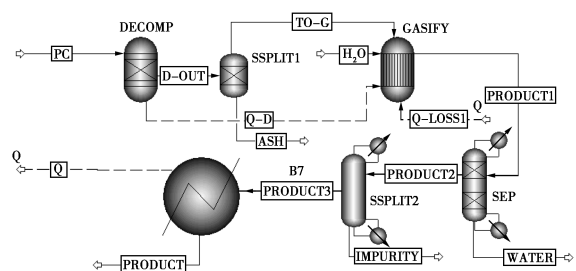


图 1 石油焦气化模型

该模型中具有 6 个模块,11 个物流,3 个热流。具体流程为:石油焦(PC)导入到产率反应器(DECOMP)进行裂解,裂解为 C、S、H<sub>2</sub>O、O<sub>2</sub>、N<sub>2</sub>、H<sub>2</sub>、灰分(ASH);然后将裂解产物(D-OUT)送入气固分离器(SSPLIT1),去除灰分(ASH),将固体碳与挥发分气体导入到 GASIFY 模块并通入气化剂(H<sub>2</sub>O)进行气化;气化产物(PRODUCT1)通入 SEP 模块将水分分离出来;之后产物(PRODUCT2)通入 SSPLIT2 模块将杂质(IMPURITY)分离出来;最后导入到换热器模块(HEATER)进行冷却得到最终气体(PRODUCT)。热流方面:裂解热(Q-D)传导气化炉中,并伴随有热量损失,损失热量 Q-loss1 按石油焦低位热值的 2%来计算<sup>[9]</sup>。

(上接第 207 页)

- [10] Rajappan R, Shingade P D, Natarajan R, et al. Quantitative structure-property relationship (QSPR) prediction of liquid viscosities of pure organic compounds employing random forest regression [J]. Industrial & Engineering Chemistry Research, 2009, 48(21): 9708-9712.
- [11] Kauffman G W, Jurs P C. ChemInform Abstract: Prediction of surface tension, viscosity, and thermal conductivity for common organic solvents using quantitative structure-property relationships [J]. Cheminform, 2001, 32(22): 348-354.
- [12] American Petroleum Institute (API). Technical data book,

petroleum refining[M]. Washington: API, 1976.

- [13] 安谧,曹青,马晨菲,等.全二维气相色谱在重馏分油分析中的应用[J].现代化工,2015,35(2):172-175.
- [14] 刘美,赵锁奇,杨影影.重质油分子组成和结构的质谱分析方法进展[J].现代化工,2013,33(6):140-142.
- [15] Akimitsu Murata, Katsumi Tochigi, Hiroshi Yamamoto. Prediction of the liquid viscosities of pure components and mixtures using neural network and ASOG group contribution methods[J]. Molecular Simulation, 2004, 30(7): 451-457.
- [16] 王志奎. 人工神经网络在化学工程中的应用[J]. 现代化工, 1996, (9): 17-21. ■