

文章编号 1004-924X(2016)11-2814-07

动态光散射颗粒分布软测量

田慧欣*, 彭 晓, 朱新军, 孟 博

- (1. 天津工业大学 电气工程与自动化学院, 天津 300387;
2. 天津工业大学 电工电能新技术天津重点实验室, 天津 300387)

摘要: 考虑传统动态光散射颗粒粒度分布测量用的反演算法复杂、精度不够、抗噪能力差, 本文基于大数据思想, 提出了一种动态光散射颗粒分布软测量方法。该方法通过调节颗粒粒度分布形状参数获得大量自相关函数及其对应颗粒分布的数据; 使用这些数据对子学习机进行训练。最后, 针对训练数据维数较高的特点对传统 Bagging 算法进行改进, 并利用改进的 Bagging 集成算法集成子学习机以提高软测量模型的精度及泛化能力。通过模拟单峰数据和对 300 nm 标准粒径进行软测量开展了验证实验。结果表明, 该方法能够较好地测量出不同动态光散射颗粒分布的峰值及分布宽度, 模拟单峰数据测量峰值精度可达 1 nm, 300 nm 和 503 nm, 标准粒径测量精度分别可达 3 nm 和 4 nm, 优于一般的反演算法。该软测量方法为动态光散射颗粒分布测量开辟了新的途径。

关键词: 动态光散射; 颗粒分布测量; 软测量; Bagging 算法

中图分类号: TP301; TB92 **文献标识码:** A **doi:** 10.3788/OPE.20162411.2814

Soft sensing of particle size distribution in dynamic light scattering measurement

TIAN Hui-xin*, PENG Xiao, ZHU Xin-jun, MENG Bo

- (1. School of Electrical Engineering & Automatic, Tianjin Polytechnic University, Tianjin 300387, China;
 2. Key Laboratory of Advanced Electrical Engineering and Energy Technology in Tianjin, Tianjin Polytechnic University, Tianjin 300387, China)
- * Corresponding author, E-mail: tianhui.xin@tjpu.edu.cn

Abstract: As the traditional inversion algorithms for particle size distribution measurement by dynamic light scattering show complex computation, lower accuracy and poorer anti-noise capacity, this paper proposes a soft sensing method for particle size distribution based on improved Bagging algorithm by using idea big data. The data of autocorrelation function and particle sizing distribution were obtained by changing the parameters of particle distribution shape. Then the learning machines were trained by the data. Finally, the traditional Bagging algorithm was improved on the basis of the character of high dimensional data. The improved Bagging strategy was used to aggregate the machines for bettering the model accuracy and its generalization performance. A validation experiment was performed by simulating the single peak data and soft sensing for the standard particles with a diameter of 300 nm. Experiment results demonstrate that the proposed method predicts the peak position and the width of

收稿日期: 2016-07-02; 修订日期: 2016-09-01.

基金项目: 国家自然科学基金资助项目 (No. 61403277; No. 71602143)

particle sizing distribution accurately, and the best accuracy of peak position measurement is 1 nm. Meanwhile, the accuracies for standard particles with diameters of 300 nm and 503 nm are 3 nm and 4 nm, respectively. The proposed method provides a new way for the particle size distribution measurement in dynamic light scattering.

Key words: dynamic light scattering; particle size distribution measurement; soft sensing; Bagging algorithm

1 引言

亚微米颗粒以及纳米颗粒在电、磁、光、热、声等各个方面都有其特殊性。随着新材料工业生产的迅猛发展,纳米颗粒由于具有表面效应和宏观量子隧道等特点,其应用也日益增多^[1-2]。此外,在动态的乳胶聚合过程中以及对 DNA、细胞等特性研究中,纳米颗粒的自身粒径也会给实验结果带来较大影响^[3]。通常用来表征纳米颗粒性能的参数是粒度的大小及其粒度分布。对同质纳米而言,不同的颗粒大小或是不同的颗粒分布都将影响到其特性能^[4]。因此,颗粒粒度及其分布的精准测量对于实际应用具有重要意义。

动态光散射技术是测量亚微米颗粒及纳米颗粒粒度分布的主要测量技术。该技术采用密封测量的方式,不会破坏和干扰样品的原有状态,具有快速、便捷、可靠以及非侵入性等特点,在生物、化学和高分子材料等领域应用广泛^[5]。动态光散射技术一般通过散射光相关函数反演求解的方法来获取颗粒粒度分布。在反演求解第一类 Fredholm 积分方程过程中,相关函数的测量需在无噪声的环境下进行,在数据的运算中也应尽量避免舍入误差的存在。但求解此类积分方程属于一个不适定病态数学问题,数据中一个微小的扰动都极有可能产生巨大误差,从而无法得到准确的颗粒粒度分布,使测量失去意义。因此,找寻最接近于真实粒度分布的反演算法是动态光散射技术中的重难点问题。针对上述问题,近些年来学者们提出了多种反演算法,其中被广泛使用的有累积法、指数法和非负约束最小二乘法,但这些方法对宽峰数据的分布测量效果较差,而 CONTIN 以及贝叶斯算法^[6-9]等对不同颗粒的测量效果不同,算法稳定性有待提高。为了提高反演算法的精度,王静雅等结合 Tikhonov 正则化与瀑布型多重网格技术将原反演问题分解到多尺度网格空间,以降低求解难度^[10],曹丽霞等改进 Chahine

算法以提高反演结果的平滑性及稳定性^[11]。另一方面,一些研究者通过分析光学相关谱、散射光谱估计等理论,并与光学系统、光电探测器等硬件测量技术相结合,完成对动态光散射颗粒分布的测量^[12-13]。大多数反演算法都需要大量先验知识来确定相关的参数,因此人工干预较多,存在抗噪能力差,测量精度低等不足。使用硬件设备虽然测量精度高,但仪器价格昂贵。针对已有测量方法的不足,本文将软测量概念引入动态光散射颗粒分布测量中,提出一种基于改进 Bagging 的人工智能集成方法,从大量颗粒分布与相关函数的数据中获取信息,建立颗粒分布软测量模型,巧妙地回避了解 Fredholm 积分方程这一难题,有效地克服了传统方法需要大量先验知识、抗噪能力差的不足,为动态光散射技术测量颗粒分布提供了新的可靠途径。

2 改进 Bagging 的颗粒分布软测量

2.1 动态光散射颗粒分布软测量模型的建立

对随机的散射光信号进行相关运算并作归一化处理,可以得到归一化的散射光强自相关函数 $g^{(2)}(\tau)$,它与归一化的电场自相关函数 $g^{(1)}(\tau)$ 的关系为:

$$g^{(2)}(\tau) = A(1 + \beta[g^{(1)}(\tau)]), \quad (1)$$

式中: τ 是延迟时间, A 是实验基线, β 是相干因子。

对于多分散颗粒系,归一化的电场自相关函数的分布积分为:

$$g^{(1)}(\tau) = \int_0^{\infty} G(\Gamma) \exp(-\Gamma\tau) d\Gamma, \quad (2)$$

式中: $G(\Gamma)$ 是归一化散射光电场强度自相关函数的线宽分布函数,其中 $\int_0^{\infty} G(\Gamma) d\Gamma = 1$, Γ 为衰减常数或线宽。

在求解方程(2)得到 $G(\Gamma)$ 后,通过衰减线宽—平移扩散系数—颗粒粒径换算后,可求得颗粒粒径的分布 $G(d)$ 。式(2)为第一类 Fredholm 积

分方程,而且在 $g^{(1)}(\tau)$ 中不可避免地存在着噪声,在此情况下求解颗粒粒径分布属于病态问题,解通常会严重偏离,甚至无法求解^[14-16]。而如果已知 $G(\Gamma)$,则可以很容易地求得 $g^{(1)}(\tau)$,因此可以考虑利用这一特性获得分布函数及自相关函数的大量数据,再使用人工智能方法对这些数据进行学习,并建立颗粒粒径分布函数的软测量模型,从而实现对颗粒分布的测量,如图 1 所示。



图 1 基于 Bagging 的颗粒分布软测量结构图

Fig. 1 Structure diagram of soft measurement of particle distribution based on Bagging

2.2 基于改进 Bagging 的颗粒分布软测量集成算法

作为一种新的机器学习方式,集成学习的测量精度较传统单一学习算法高。其基本思想是将多个精度较低的弱学习机集成,得到一个较强的学习机^[17]。Bagging 集成算法首先通过 bootstrap,即有放回式的抽样方法,得到相应数量的子训练集,然后使用子训练集对弱学习机进行训练,最后将训练弱学习机得到的结果进行集成得到强学习机,并输出最终结果。在训练过程中,各子数据集相互独立,测量子函数并行生成,因此,集成后的最终模型精度以及泛化性能均得到了很大的提高^[18]。

改进后用于颗粒分布软测量的 Bagging 集成方法结构图如图 2 所示。

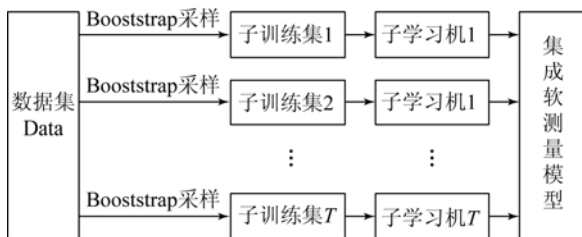


图 2 基于 Bagging 的集成软测量模型结构图

Fig. 2 Structure diagram of integrated soft measurement model based on Bagging

该算法首先确定子数据集个数 T ,并从训练数据集 Data 中有放回地抽取 L 个训练数据放入子数据集 S_i 中,每个数据样本对应的输出 y 为 R 维;然后,用训练子集来训练子学习机得到子测量模型 $f_i(x) \rightarrow y$,并计算出各个子模型的权重因子 β_i ;最后根据权重因子将训练好的子模型集成为最终的软测量模型。改进方法在传统的 Bagging 集成方法的基础上加入了权重因子,用于最终模型的集成。求解权重因子的过程如下:首先计算子测量模型的训练误差,并将误差与预先设定的阈值进行比较,然后统计误差大于阈值数据的个数用于计算误差率,这样就可以得到该子模型的权重因子了。权重因子的引入加大了训练效果好的子模型的权重,使得整个集成算法的精度得到了有效地提升。此外,由于考虑了动态光散射颗粒分布的特点,改进后的算法所建立的软测量模型更适用于高维数据。与低维数据不同,改进后的算法在计算子模型的训练误差与误差率时,需要分别计算每一维输出数据的绝对误差以及对应的训练误差、标准差、误差率等参数,最终权重因子取其计算结果的平均值。改进后的算法有效地考虑了多维数据对算法精度的影响。软测量建模方法的具体步骤如下。

输入:

(1)数据集 Data 样本数量为 m :

$$Data = [(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)],$$

其中 $y = y_1, y_2, \dots, y_R$ 。

(2)确定子学习机算法。

(3)确定子学习机个数 T 以及子训练集的大小 L 。

(4)设定阈值参数 $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_R) (\varphi > 0)$ 。

循环: $t = 1, 2, 3, \dots, T$

(1)采用 bootstrap 抽样方法抽取 L 个数据组成的训练子集 S_i 。

(2)根据训练数据训练弱学习机,得到测量模型: $f_i(x) \rightarrow y$ 。

(3)计算 $f_i(x)$ 的训练误差:

$$Er_i^r(i) = \frac{|f_i^r(x_i) - y_i^r|}{\sigma_i^r}, \quad (3)$$

其中 σ_i^r 为 $(f_i^r(x_i) - y_i^r)$ 的标准差。

(4)计算 $f_i(x)$ 的误差率:

$$\epsilon_i^r = \frac{l_i^r}{L}, \quad (4)$$

l_i^r 为 $Er_i^r > \varphi_r$ 的个数。

(5)计算权重因子:

$$\beta_t = \frac{1}{R} \sum_{r=1}^R \epsilon_r^t. \quad (5)$$

输出:

$$F(x) = \frac{\sum_t \left[\log \frac{1}{\beta_t} \right] f_t(x)}{\sum_t \left[\log \frac{1}{\beta_t} \right]}. \quad (6)$$

较传统的反演推导算法而言,使用软测量方法获取动态光散射颗粒分布的精度较高,操作简单,只需在实验训练过程中调整阈值参数以获取准确的模型,无需繁复的先验背景知识以及复杂的数学推导过程,为研究动态光散射颗粒分布提供了另一种可能。

3 实验

3.1 测试数据测量实验

在用已知 $G(\Gamma)$ 反推求解 $g^{(1)}(\tau)$ 的过程中,通过改变 100~800 nm 单峰颗粒 Johnson's S_B 参数 u 与 σ ,获取实验所需 175 种颗粒粒度分布数据用于训练模型,其中 $u \in (0.8, 3)$, $\sigma \in (1.8, 6.8)$,数据噪声水平为 10^5 。训练数据的输入为 200 维,输出为 70 维。测试数据分为 5 组,其参数如表 1 所示。

表 1 测试数据对应参数表

Tab. 1 Parameters corresponding to test data

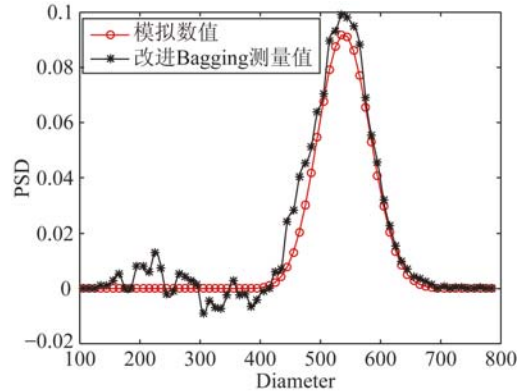
数据	u	σ	噪声水平 (n)
1	1.5	6.8	10^5
2	3	2.6	10^5
3	1.5	2.6	10^5
4	1.5	2.6	10^{10}
5	1.5	2.6	10^{15}

均方根误差 RMSE 的计算公式如式 (7) 所示:

$$RMSE = \sqrt{\frac{1}{R} \sum_{r=1}^R (y^r - f^r(x))^2}. \quad (7)$$

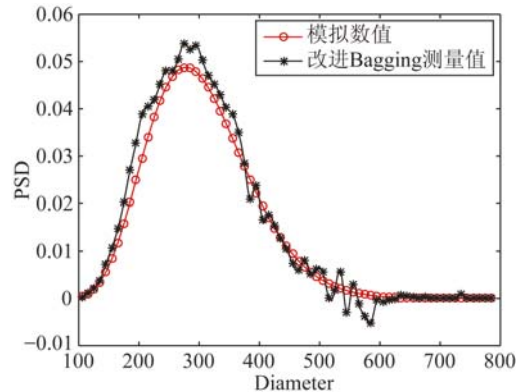
实验采用三层的 BP(Back Propagation)神经网络作为子学习机,子学习机个数为 40,子训练集大小为 100。建立好软测量模型后,使用 5 组

测试数据进行实验仿真。1、2、3 组数据加入的噪声水平相同,峰值不同;3、4、5 组数据峰值相同加入的噪声水平不同。噪声水平的值越大说明该数据噪声越小。5 组数据的峰值、宽度分布及均方根误差如表 2 所示,测量分布图如图 3、图 4 所示。



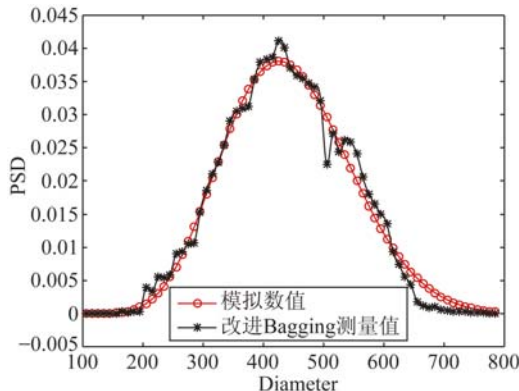
(a)第 1 组参数测量结果 ($u=1.5, \sigma=6.8$)

(a) Result of the 1st group parameters ($u=1.5, \sigma=6.8$)



(b)第 2 组参数测量结果 ($u=3, \sigma=2.6$)

(b) Result of the 2nd group parameters ($u=3, \sigma=2.6$)

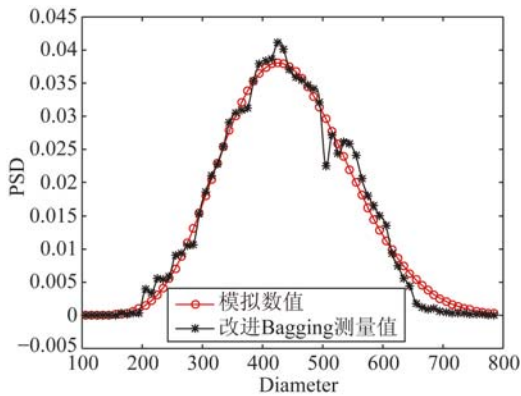


(c)第 3 组参数测量结果 ($u=1.5, \sigma=2.6$)

(c) Result of the 3rd group parameters ($u=1.5, \sigma=2.6$)

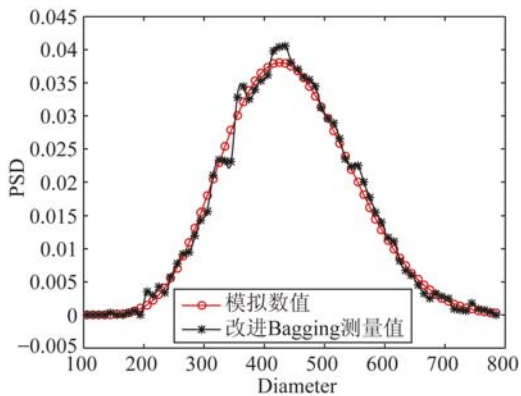
图 3 相同噪声水平下的数据测量结果

Fig. 3 Results of test data with the same noise factor $n=1 \times 10^5$



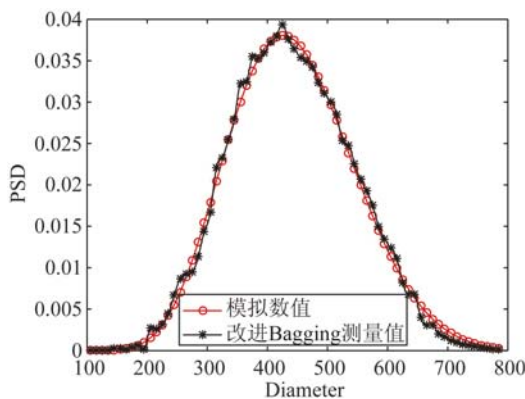
(a)第 3 组参数测量结果(noise factor= 1×10^5)

(a) Result of the 3rd group parameters(noise factor= 1×10^5)



(b)第 4 组参数测量结果(noise factor= 1×10^{10})

(b)Result of the 4th group parameters(noise factor= 1×10^{10})



(c)第 5 组参数测量结果(noise factor= 1×10^{15})

(c) Result of the 5th group parameters(noise factor= 1×10^{15})

图 4 相同峰值的数据测量结果

Fig. 4 Result of test data with same peak position

($u=1.5, \sigma=2.6$)

表 2 测试数据测量结果

Tab. 2 Results of test data in simulation experiment

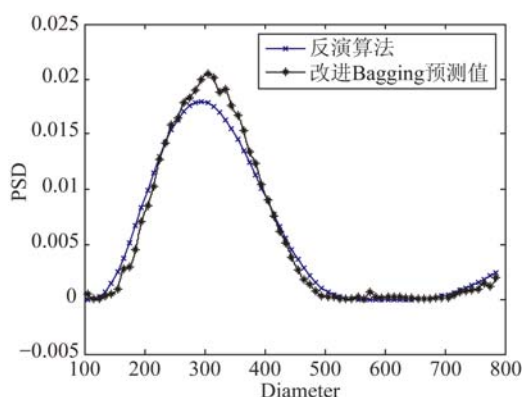
模拟测试	预测分布宽度/nm	真实分布宽度/nm	预测峰值/nm	真实峰值/nm	均方根误差/nm
数据 1	415~695	405~675	539	532	0.006 1
数据 2	105~595	105~575	277	271	0.003 0
数据 3	195~675	195~735	430	422	0.001 8
数据 4	215~715	195~735	426	422	0.001 3
数据 5	215~715	195~735	423	422	0.000 8

实验结果表明,软测量方法获得的颗粒分布精度较高且抗噪性能良好。针对不同的动态光散射颗粒分布,测量总体效果较好,分布宽度的预测值与真实值大致相同,峰值误差较小,精度可达到 8 nm。在加入噪声后,软测量建模算法的误差增大,但随着噪声的增加,误差增长幅度较小,有一定的抗噪性能。

3.2 真实数据测量实验

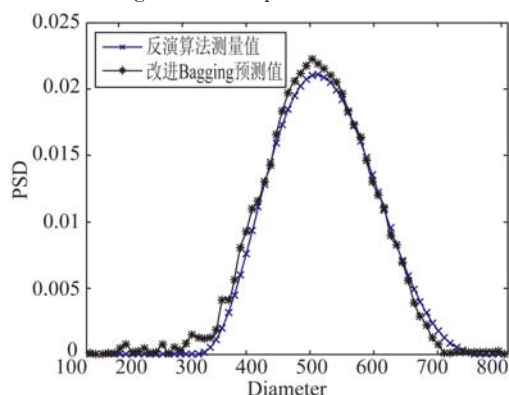
300 nm 与 503 nm 标准颗粒采用 MilliPore Milli-Q/ Milli-Rho 系统去离子蒸馏水稀释,稀释后的 300 nm (3 300 A) 颗粒浓度为 3.0×10^{-4} wt%, 503 nm (3 500 A) 颗粒浓度为 2.0×10^{-4} wt%。样品池温度控制在 25 °C, Ar+激光器的激光功率为 50 mW,将通道 BI-2030AT 72 相关器测量的粒径作为标准粒径用于数据的对比。对于 300 nm 标准粒径,相关器测量值为 $324 \text{ nm} \pm 2 \text{ nm}$,对于 503 nm 标准粒径,相关器测量值 $520 \text{ nm} \pm 4 \text{ nm}$ 。

实验分别使用改进 Bagging 的软测量方法与基于二阶光强相关函数的改进反演算法^[2]对 300 nm 与 503 nm 标准颗粒数据进行分布测量,其中反演算法采用约束 Tikhonov 正则化,参数选取为固定值 0.005。图 5 为两种算法得到的 300 nm 颗粒与 503 nm 颗粒分布图。



(a) 300 nm 颗粒反演结果与软测量结果对比图

(a) Comparison of inversion method and soft sensing in 300 nm particle



(b) 503 nm 颗粒反演结果与软测量结果对比图

(b) Comparison of inversion method and soft sensing in 503 nm particle

图 5 标准颗粒反演结果与软测量结果对比图

Fig. 5 Results of inversion method and soft sensing in standard particles

参考文献:

- [1] 王雅静,郑刚,申晋,等. 用于动态光散射颗粒测量的改进双指数算法[J]. 中国激光, 2009, 36(s2): 173-177.
WANG Y J, ZHENG G, SHEN J, et al.. Improving double exponential algorithm for particle sizing in dynamic light scattering [J]. *Chinese Journal of Lasers*, 2009, 36(s2): 173-177. (in Chinese)
- [2] ZHU X, SHEN J, WANG Y, et al.. The reconstruction of particle size distributions from dynamic light scattering data using particle swarm optimization techniques with different objective functions [J]. *Optics & Laser Technology*, 2011, 43(43): 1128-1137.
- [3] SCHURTENBERGER P, SCHEFFOLD F, SHALKEVICH A, et al.. PCS particle sizing in

图 5 表明,软测量方法得到的 300 nm 标准颗粒峰值为 303 nm,精度为 3 nm;503 nm 标准颗粒峰值为 507 nm,精度为 4 nm。而反演算法得到的 300 nm 标准颗粒峰值为 295 nm,精度为 5 nm;503 nm 标准颗粒峰值为 513 nm,精度为 10 nm。较真实的标准颗粒而言,软测量方法精度较高,而且软测量得到的分布宽度比反演算法得到的分布宽度窄,较为稳定。

4 结 论

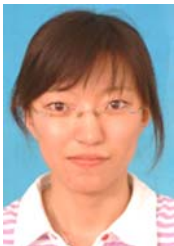
本文根据动态散射光颗粒的特点,与软测量技术相结合,提出了改进 Bagging 的高维度软测量方法。首先通过反求解的方法求解第一类 Fredholm 积分方程获得大量不同的动态光散射颗粒分布数据。然后用所获取的数据训练子学习机。最后将子学习机集成为最终的软测量模型并使用测试数据进行检验。实验结果表明,对于模拟数据,该软测量方法的峰值精度最高可达 1 nm,对于真实数据精度最高可达 3 nm,精度优于反演算法。本文所研究的软测量技术为动态光散射颗粒分布的测量提供了新的可靠途径。

turbid suspensions: scope and Limitations [J]. *Modeling Analysis & Simulation of Computers & Telecommunication Systems, mascots, ie*, 2004, 21(1):28-46.

- [4] SHI W, SONG J, ZHU X, et al.. Design of weighing transmitter in corn yield measurement system [C] *Proceedings of the First Electric Information and Control Engineering, Wuhan, P. R. China: ICEICE*, 2011:324-327.
- [5] MORRISON I D, GRABOWSKI E F, HERB C A. Improved techniques for particle size determination by quasi-elastic light scattering [J]. *Langmuir*, 2002, 1(4):496-501.
- [6] PROVENCHER S W. A constrained regularization method for inverting data represented by linear algebraic or integral equations [J]. *Computer Physics Communications*, 1982, 27(3):213-227.

- [7] CLEMENTI L A, VEGA J R, GUGLIOTTA L M, *et al.*. A Bayesian inversion method for estimating the particle size distribution of latexes from multiangle dynamic light scattering measurements [J]. *Chemometrics & Intelligent Laboratory Systems*, 2011, 107(1):165-173.
- [8] ZHU X, SHEN J, THOMAS J C. Analysis of noisy dynamic light scattering data using constrained regularization techniques [J]. *Applied Optics*, 2012, 51(31):7537-7548.
- [9] LIU X Y, SHEN J, THOMAS J C, *et al.*. Multiangle dynamic light scattering analysis using angular intensity weighting determined by iterative recursion [J]. *Applied Optics*, 2012, 51(7):846-854.
- [10] 王雅静, 申晋, 郑刚, 等. Tikhonov 正则化与多重网格技术相结合的动态光散射反演[J]. *光学精密工程*, 2012, 22(5):963-971.
WANG Y J, SHENG G, ZHEN G, *et al.*. Inversion of dynamic light scattering combining Tikhonov regularization with multi-grid technique [J]. *Opt. Precision Eng.*, 2012, 22(5):963-971. (in Chinese)
- [11] 曹丽霞, 赵军, 孔明, 等. 基于改进的 Chahine 迭代算法的粒径分布反演[J]. *红外与激光工程*, 2015, 44(9):2837-2843.
CAO L X, ZHAO J, KONG M, *et al.*. Inversion of particle size distribution based on improved Chahine algorithm [J]. *Infrared and Laser Engineering*, 2015, 44(9):2837-2843. (in Chinese)
- [12] 杨晖, 郑刚, 张仁杰. 用动态光散射时间相干度法测量纳米颗粒粒径[J]. *光学精密工程*, 2011, 19(7):1546-1551.
YANG H, ZHENG G, ZHANG R J. Measurement of nanoparticle sizes by variance of temporal coherence of dynamic light scattering [J]. *Opt. Precision Eng.*, 2011, 19(7):1546-1551. (in Chinese)
- [13] 杨晖, 郑刚, 王雅静. 用动态光散射现代谱估计法测量纳米颗粒[J]. *光学精密工程*, 2010, 18(9):1996-2001.
YANG H, ZHENG G, WANG Y J. Measurement of nano-particles by dynamic light scattering based on spectral estimation [J]. *Opt. Precision Eng.*, 2010, 18(9):1996-2001. (in Chinese)
- [14] GAO S, SHEN J, THOMAS J C, *et al.*. Effect of scattering angle error on particle size determination by multiangle dynamic light scattering [J]. *Applied Optics*, 2015, 54(10):2824-2831.
- [15] KIM J, AHN S, LEE H, *et al.*. Estimation of particle size distribution using photon autocorrelation function from dynamic light scattering considering unknown baseline[J]. *Optics Letters*, 2013, 38(11):1757-1759.
- [16] CLEMENTI L A, VEGA J R, GUGLIOTTA L M. Particle size distribution of multimodal polymer dispersions by multiangle dynamic light scattering. solution of the inverse problem on the basis of a genetic algorithm [J]. *Particle & Particle Systems Characterization*, 2012, 27(5-6):146-157.
- [17] TIAN H X, MAO Z Z, WANG A. A new incremental learning modeling method based on multiple models for temperature prediction of molten steel in LF[J]. *Isij International*, 2009, 49(1):58-63.
- [18] TIAN H X, MAO Z Z, WANG Y. Hybrid modeling of molten steel temperature prediction in LF [J]. *Isij International*, 2008, 48(1):58-62.

作者简介:



田慧欣(1978—),女,辽宁抚顺人,副教授,2005年、2009年于东北大学分别获得硕士、博士学位,主要从事人工智能软测量及应用研究。E-mail: tianhuixin@tjpu.edu.cn



彭晓(1992—),女,湖北孝感人,硕士研究生,主要研究方向为人工智能软测量方法及光干涉测量的研究。E-mail: 853050658@qq.com