

# 基于 SVM 理论的涤纶织物分散染料上染率模型研究

巫 静<sup>a</sup>, 田彦杰<sup>b</sup>, 汪 澜<sup>a</sup>, 林俊雄<sup>b</sup>

(浙江理工大学, a. 先进纺织材料与制备技术教育部重点实验室; b. 机械与自动控制学院, 杭州 310018)

**摘 要:** 以分散红 FB 对涤纶织物染色为例, 利用在解决小样本、非线性及高维模式识别中表现出许多特有优点的支持向量机(SVM)理论, 建立了涤纶织物分散染料染色工艺因素—上染率多因素模型。并通过实验进行验证, 结果表明: 实验值与模型预测值的相关系数达到了 0.9778, 误差均在 3% 以内, 说明该数学模型具有较高的精确度, 能较准确地反映分散染料对涤纶织物的实际上染情况, 并可预测染色结果。

**关键词:** 涤纶织物; 分散染料; 上染率; 支持向量机; 多因素模型

**中图分类号:** TS190.9      **文献标志码:** A

## 0 引 言

近年来, 印染企业往往因染色一次成功率低导致染色过程中出现重染、修色、剥色改染等情况, 从而增加了额外的能耗和排放。在纺织品染色过程中, 染料的上染率主要受系统因素和工艺因素两方面影响<sup>[1]</sup>。在固定系统因素的实际生产中, 一般靠人工经验调节工艺参数来改变染料上染率, 故很难保证染色品的质量。因此建立染色工艺因素与染料上染率的数学模型, 精确预测染料上染率, 适当调整相应工艺参数, 使染料充分上染, 是提高染色一次成功率的有效途径<sup>[2]</sup>。

目前国内在染色工艺因素与染料上染率的数学模型方面研究较少, 国外在这方面虽有一些进展, 但主要针对温度因素对上染率的影响而建立染色温度—上染率单因素模型, 忽略了其他工艺因素对上染率的影响, 故在实际生产中无法准确地预测上染率。因此, 本文以分散红 FB 对涤纶织物染色为例, 首先建立涤纶织物分散染料染色工艺因素—上染率多因素模型, 然后通过实验验证该模型的准确性和可行性。

本文建立的多因素模型基于支持向量机

(SVM)理论, 该理论是建立在统计学习理论的 VC 维理论和结构风险最小原理的基础上, 即根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折衷。SVM 理论能够尽量提高学习机的推广能力, 即使由有限数据集得到的判别函数对独立的测试集仍能够得到较小的误差。与其他学习机相比, SVM 具有较好的泛化能力、非线性处理能力及高维处理能力<sup>[3]</sup>, 故特别适合用于影响因素多、且样本不够均匀的染色过程中。

## 1 实验部分

### 1.1 材料与药品

涤纶织物(涤平纺); 200%分散红 FB(浙江龙盛集团有限公司); 醋酸(AR), 扩散剂 NNO(工业品), 皂片(工业品)等。

### 1.2 仪器与设备

IR-12 红外染色机(台湾新瑞开发科技有限公司), UV-2 550 紫外/可见分光光度计(Shimadzu Corporation), 雷磁 PHS-3C 型 pH 计(东莞市创瑞检测试验设备有限公司), DHG-9140A 型电热恒温鼓风干燥箱(英博联科技公司), FA2004 电子天平(天津市天有利有限公司)等。

收稿日期: 2012-06-11

基金项目: 国家自然科学基金(61074154)

作者简介: 巫 静(1987-), 女, 江苏扬州人, 硕士研究生, 主要从事新型染整技术研究。

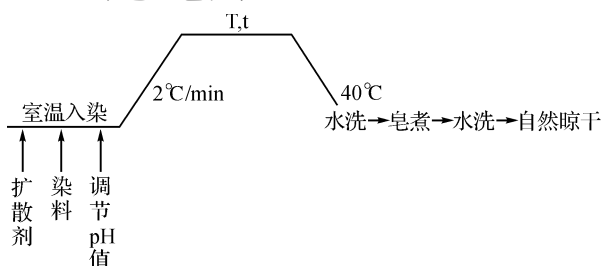
通信作者: 汪 澜, E-mail: wlan\_cn@yahoo.com.cn

### 1.3 实验及测试方法

#### 1.3.1 染色工艺处方

|               |                                   |
|---------------|-----------------------------------|
| 分散染料/%(owf)   | 2;                                |
| 扩散剂 NNO/(g/L) | 2;                                |
| 温度/℃          | 100~140, 每隔 2℃ 作为一个实验点;           |
| 时间/min        | 0~70, 每隔 5 min 作为一个实验点;           |
| pH 值(用醋酸调节)   | 4~6, 每隔 0.2 作为一个实验点, 增加 4.5, 5.5; |
| 浴比            | 1:50。                             |

#### 1.3.2 染色工艺流程



#### 1.3.3 上染百分率( $E\%$ )的测定

采用剥色法<sup>[4]</sup>:称取 0.1 g 染色试样于 50 mL 烧瓶内,加入 10 mL DMF 置于甘油浴中于 130℃ 下回流 5 min,同时振荡烧瓶,冷却滤出剥色残液,将剥色残液移至 50 mL 容量瓶中,重复上述操作 1~5 次,直至样品变为无色。将剥色残液全部移入 50 mL 容量瓶中,并用 DMF 剥色液定容,摇匀。采用紫外/可见分光光度计测得萃取染液中的吸光度  $A$ 。

按下式计算上染百分率:

$$E\% = A/A_0 \times 100\% \quad (1)$$

式中: $A$ —剥色液的吸光度;

$A_0$ —空白染液的吸光度。

## 2 结果与讨论

在涤纶织物分散染料高温高压染色中,对染料上染率影响较大的工艺因素有染色温度、保温时间、染液 pH 值和浴比<sup>[5]</sup>,但由于在实际染色过程中,浴比因素无法实时调节,故以分散红 FB 为例,对染色温度、保温时间和染液 pH 值 3 工艺因素与上染率进行多因素建模。

### 2.1 染色过程建模

SVM 理论基本上不涉及概率测度的定义及大数定律等,因此它不同于现有的统计方法。它的最终决策函数只由少数的支持向量所确定,而计算的复杂性取决于支持向量的数目,而不是样本空间的维数,这

就在某种意义上避免了“维数灾”<sup>[6]</sup>。少数支持向量决定了最终结果,这不但可以帮助我们抓住关键样本,“剔除”大量冗余样本,而且注定了该方法不但算法简单,而且具有较好的“鲁棒”性。由于 SVM 理论有较为严格的统计学习理论作保证,从而利用 SVM 理论建立的模型具有较好的推广能力。SVM 方法可以给出所建模型的推广能力确定的界,这是目前其它任何学习方法所不具备的;建立任何一个数据模型,人为的干预越少就越客观。因此与其他方法相比,利用 SVM 理论建立的模型所需要的先验干预较少<sup>[7]</sup>。

支持向量机回归有线性回归和非线性回归。由于涤纶分散染料染色过程中各工艺因素对上染率的影响属于非线性,本文就利用非线性支持向量机回归方法来建立涤纶织物分散染料染色工艺因素—上染率模型。

#### 2.1.1 非线性支持向量机回归原理

非线性支持向量机就是通过一个非线性映射  $\Phi$  将数据  $x$  映射到高维特征空间,其映射方式为  $x \rightarrow \Phi(x) = (a_1 \Phi_1(x), a_2 \Phi_2(x), \dots, a_n \Phi_n(x))$ ,  $a_n \in R$ ,  $\Phi_n \in R^n$ ,并在这个空间进行线性回归。

在高维特征空间的线性回归考虑用线性回归函数  $f(x) = w \cdot x + b$  来估计数据  $(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_l, y_l)$ ,  $x_i, y_i \in R$ ,为了处理函数在  $\epsilon$  精度不能估计的数据,引入松弛变量  $\xi_i, \xi_i^*$ ,因此优化问题为:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*)$$

$$\text{约束条件为:} \begin{cases} y_i - w \cdot x_i - b \leq \epsilon + \xi_i \\ w \cdot x_i + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

引入拉格朗日函数和对偶变量  $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ :

$$L = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l \alpha_i (\xi_i + \epsilon - y_i + w \cdot x_i + b) - \sum_{i=1}^l \alpha_i^* (\xi_i^* + \epsilon + y_i - w \cdot x_i - b) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*)$$

其中对偶变量  $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$  大于或等于 0,  $C > 0$ ,再根据 KKT 条件

$$\begin{cases} \frac{\partial L}{\partial b} = \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, l \\ \frac{\partial L}{\partial w} = w - \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i = 0 \Rightarrow w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i \\ \frac{\partial L}{\partial \xi_i^{(*)}} = C - \alpha_i^{(*)} - \eta_i^{(*)} \Rightarrow C = \alpha_i^{(*)} + \eta_i^{(*)} \end{cases}$$

在约束条件下最大化

$$W(\alpha, \alpha_i) = -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(x_i \cdot x_j) + \sum_{i=1}^l (\alpha_i - \alpha_i^*)y_i - \sum_{i=1}^l (\alpha_i + \alpha_i^*)\epsilon$$

求得参数  $\alpha_i, \alpha_i^*$  代入  $w = \sum_{i=1}^l (\alpha_i - \alpha_i^*)x_i$ , 得到线性回归函数为:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*)(x_i \cdot x) + b \quad (2)$$

在高维特征空间的线性回归就对应于低维输入空间的非线性回归,其具体是通过核函数来实现。通过这样的转换,就可以采用解决线性回归的方法来解决非线性回归,而同时免去了在高维空间计算复杂的点积运算<sup>[8]</sup>。

直接利用核函数  $K(x_i, x)$  代替  $x_i \cdot x$ , 求得  $\alpha_i, \alpha_i^*$ , 得到非线性回归函数为:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*)K(x_i \cdot x) + b \quad (3)$$

## 2.1.2 支持向量机算法步骤

1) 给定训练集  $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (R^n \times y)^l$ , 其中  $x_i \in R^n, y_i \in y = R, i = 1, \dots, l$ ;

首先用分散红 FB 对涤纶织物进行高温高压染色,按染色工艺处方设计的实验方案,得出对应的 49 组上染率数据,见表 1;然后将染色温度、保温时间、染液 pH 值变量作为输入变量,上染率作为输出变量输入程序。

2) 选取核函数  $K(x_i, x)$  及适当的精度  $\epsilon$ , 求取参数  $C, g$ ;

核函数选取高斯径向基函数  $K(x, x_i) = \exp\left\{-\frac{\|x - x_i\|^2}{2\sigma^2}\right\}$  (经过试验对比,发现应用径向基函数作为本研究的核函数得到的数据准确率最高<sup>[9]</sup>)。

精度  $\epsilon$  取 0.1。

本文采用网格划分(grid search)与遗传算法(GA)分别对惩罚系数  $C$ , 核函数参数  $g\left(\frac{1}{\sigma^2}\right)$  进行寻优选取,得到最佳参数  $C, g$ , 对比发现遗传算法获得的最优解能使所建立的模型准确度更高。其在提高回归结果的准确率的同时,解决了以前凭经验进行参数设置的不确定性<sup>[10]</sup>。

通过对工艺参数 - 上染率输入样本数据的训练学习,得出参数最优解为:

$$C = 92.309 \quad g = 0.0066757.$$

3) 构造并求解凸二次规划问题;

依据上面叙述的算法得解  $\bar{\alpha}^{(*)} = (\bar{\alpha}_1, \bar{\alpha}_1^{(*)}, \dots, \bar{\alpha}_l, \bar{\alpha}_l^{(*)})^T$ ;

凸二次规划问题即寻求最优解问题,将第二步已计算得出的惩罚系数  $C$  和精度  $\epsilon$  代入公式寻求  $\bar{\alpha}^{(*)}$  最优解。

$$\text{进一步求得 } \beta = \bar{\alpha}_i - \bar{\alpha}_i^* = \begin{bmatrix} 40.33 \\ -81.64 \\ \vdots \\ 54.06 \end{bmatrix}$$

4) 计算  $\bar{b}$ : 选取位于开区间  $(0, C)$  中  $\bar{\alpha}^{(*)}$  的分量  $\bar{\alpha}_j$  或  $\bar{\alpha}_k^*$ ;

若选到的是  $\bar{\alpha}_j$  则  $\bar{b} = y_j - \sum_{i=1}^l (\bar{\alpha}_i - \bar{\alpha}_i^*)K(x_i, x_j) + \epsilon$ ;

若选到的是  $\bar{\alpha}_k^*$ , 则  $\bar{b} = y_k - \sum_{i=1}^l (\bar{\alpha}_i - \bar{\alpha}_i^*)K(x_i, x_k) - \epsilon$ ;

利用  $\bar{\alpha}^{(*)}$  最优解计算常数  $\bar{b} = 60.49$ ,

以上参数寻优与凸二次规划问题的求解均借助林智仁的 Libsvm 工具箱<sup>[11]</sup> 在 Matlab 平台中计算实现。

5) 构造决策函数  $y = g(x) = \sum_{i=1}^l (\bar{\alpha}_i - \bar{\alpha}_i^*)K(x_i, x) + \bar{b}$

决策函数中  $\beta = \bar{\alpha}_i - \bar{\alpha}_i^*, \bar{b}$  已在前 2 步计算得出,从而最终得到的决策函数为:

$$f(x) = \sum_{i=1}^{49} \begin{bmatrix} 40.33 \\ -81.64 \\ \vdots \\ 54.06 \end{bmatrix} \cdot \exp\left\{-\frac{\|x - x_i\|^2}{299.59}\right\} + 60.49 \quad (4)$$

将模型的训练样本实验值与通过计算机编程获得的回归值进行对比,结果见表 1。

表 1 训练样本实验值与回归值对比

| 样本 | 实验方案   |          |         | 上染率/% |       |
|----|--------|----------|---------|-------|-------|
|    | 染色温度/℃ | 保温时间/min | 染液 pH 值 | 实验值   | 回归值   |
| 1  | 100    | 60       | 5.0     | 19.24 | 19.24 |
| 2  | 102    | 60       | 5.0     | 19.87 | 19.87 |
| 3  | 104    | 60       | 5.0     | 20.33 | 23.60 |
| 4  | 106    | 60       | 5.0     | 30.98 | 29.99 |
| 5  | 108    | 60       | 5.0     | 40.54 | 38.14 |
| 6  | 110    | 60       | 5.0     | 45.00 | 46.86 |
| 7  | 112    | 60       | 5.0     | 53.94 | 55.03 |

表 1 续

| 样本 | 实验方案   |          |         | 上染率/% |       |
|----|--------|----------|---------|-------|-------|
|    | 染色温度/℃ | 保温时间/min | 染液 pH 值 | 实验值   | 回归值   |
| 8  | 114    | 60       | 5.0     | 61.85 | 61.85 |
| 9  | 116    | 60       | 5.0     | 67.50 | 67.06 |
| 10 | 118    | 60       | 5.0     | 70.87 | 70.87 |
| 11 | 120    | 60       | 5.0     | 73.85 | 73.79 |
| 12 | 122    | 60       | 5.0     | 75.37 | 76.30 |
| 13 | 124    | 60       | 5.0     | 78.37 | 78.64 |
| 14 | 126    | 60       | 5.0     | 80.59 | 80.71 |
| 15 | 128    | 60       | 5.0     | 83.47 | 82.26 |
| 16 | 130    | 60       | 5.0     | 83.27 | 83.04 |
| 17 | 132    | 60       | 5.0     | 82.61 | 83.06 |
| 18 | 134    | 60       | 5.0     | 82.32 | 82.57 |
| 19 | 136    | 60       | 5.0     | 81.96 | 81.96 |
| 20 | 138    | 60       | 5.0     | 81.76 | 81.50 |
| 21 | 140    | 60       | 5.0     | 81.24 | 81.24 |
| 22 | 128    | 0        | 5.0     | 37.61 | 37.61 |
| 23 | 128    | 5        | 5.0     | 45.32 | 45.32 |
| 24 | 128    | 10       | 5.0     | 56.30 | 54.56 |
| 25 | 128    | 15       | 5.0     | 58.26 | 58.26 |
| 26 | 128    | 20       | 5.0     | 59.90 | 59.90 |
| 27 | 128    | 25       | 5.0     | 62.95 | 62.95 |
| 28 | 128    | 30       | 5.0     | 66.74 | 66.74 |
| 29 | 128    | 35       | 5.0     | 71.02 | 71.02 |
| 30 | 128    | 40       | 5.0     | 74.31 | 74.31 |
| 31 | 128    | 45       | 5.0     | 75.68 | 76.27 |
| 32 | 128    | 50       | 5.0     | 78.70 | 78.70 |
| 33 | 128    | 55       | 5.0     | 80.89 | 80.89 |
| 34 | 128    | 60       | 5.0     | 83.47 | 82.26 |
| 35 | 128    | 65       | 5.0     | 83.38 | 83.38 |
| 36 | 128    | 70       | 5.0     | 82.02 | 82.02 |
| 37 | 128    | 60       | 4.0     | 81.34 | 81.84 |
| 38 | 128    | 60       | 4.2     | 81.95 | 81.95 |
| 39 | 128    | 60       | 4.4     | 81.91 | 82.09 |
| 40 | 128    | 60       | 4.5     | 82.00 | 82.04 |
| 41 | 128    | 60       | 4.6     | 82.34 | 82.13 |
| 42 | 128    | 60       | 4.8     | 83.23 | 82.20 |
| 43 | 128    | 60       | 5.0     | 83.47 | 82.26 |
| 44 | 128    | 60       | 5.2     | 85.27 | 82.30 |
| 45 | 128    | 60       | 5.4     | 82.43 | 82.34 |
| 46 | 128    | 60       | 5.5     | 82.33 | 82.35 |
| 47 | 128    | 60       | 5.6     | 82.26 | 82.36 |
| 48 | 128    | 60       | 5.8     | 82.10 | 82.36 |
| 49 | 128    | 60       | 6.0     | 82.36 | 82.36 |

利用 Matlab 程序中的 corrcoef 函数 $[R,P]=\text{corrcoef}(x,y)$ (其中  $x$  为实验值,  $y$  为回归值)求得实验值与模型回归值的相关系数  $R=0.998\ 8$ , 显著性水平  $P=0$ 。由于相关系数大于 0.9、显著性水平

小于 0.05 ( $P$  小于 0.05 就说明有统计学意义), 说明实验值与模型回归值吻合性较好, 即该模型的拟合度较高。

2.2 模型效果验证

任意设定 5 组染色温度、保温时间及染液 pH 值进行实验, 对涤纶织物分散染料染色工艺因素—上染率多因素模型进行验证, 即将实验值与模型预测值进行对比, 从而求出误差, 结果如表 2 所示。

表 2 涤纶织物分散染料染色工艺因素—上染率多因素模型预测结果

| 染色温度/℃ | 保温时间/min | 染液 pH 值 | 上染率/% |       | 误差/% |
|--------|----------|---------|-------|-------|------|
|        |          |         | 实验值   | 预测值   |      |
| 125    | 40       | 4.5     | 72.46 | 73.39 | 1.28 |
| 117    | 60       | 5.0     | 70.80 | 69.11 | 2.39 |
| 132    | 65       | 4.9     | 80.96 | 80.58 | 0.47 |
| 130    | 30       | 4.0     | 67.32 | 66.53 | 1.17 |
| 135    | 46       | 5.6     | 75.66 | 73.81 | 2.45 |

由表 2 可知, 利用涤纶织物分散染料染色工艺因素—上染率多因素模型计算的预测值与实验值较为接近, 误差均可控制在 3% 以内, 且实验值与模型预测值的相关系数  $R=0.977\ 8$ , 显著性水平  $P=0$ 。由于相关系数大于 0.9、显著性水平小于 0.05 ( $P$  小于 0.05 就说明有统计学意义), 说明该模型能精确预测涤纶织物分散染料染色过程中任何工艺条件下的上染率, 即该模型有效可行。

3 结 论

1) 建立了涤纶织物分散染料染色工艺因素—上染率多因素模型, 其实验值与模型回归值的相关系数  $R=0.998\ 8$ , 显著性水平  $P=0$ 。说明该数学模型的拟合度较高。

2) 通过对涤纶织物分散染料染色工艺因素—上染率多因素模型的验证, 实验值和模型预测值的误差在 3% 以内, 且实验值与模型预测值的相关系数  $R=0.9778$ , 显著性水平  $P=0$ 。说明该数学模型切实可行, 能预测涤纶织物分散染料染色的上染率结果, 对提高染色一次成功率具有理论价值和实际意义。

参考文献:

[1] Brent S. Dyebath monitoring and control: past, present, and future[J]. AATCC Review, 2007, 11: 36-41.  
[2] Lu Jun, Spiekermann Charles, McGregor Ralph, et al. A novel approach to modeling and controlling dyeing processes [J]. AATCC International, 1995, 27(3): 23-26.

- [3] 韩 颖. 基于 LS-SVM 的电炉钢水温度预测[J]. 沈阳大学学报, 2008, 20(2): 36-38.
- [4] 蒲宗耀, 黄玉华, 刘 庆. DMF 剥色法测定分散染料在细旦涤纶纤维上的固色率[J]. 印染, 1996, 22(12): 32-33.
- [5] 赵 涛. 染整工艺与原理: 下册[M]. 北京: 中国纺织出版社, 2010, 37-45, 176-180.
- [6] 景世磊. 基于 SVM 和 PSO 的烧结工况预报方法的研究与实现[D]. 沈阳: 东北大学, 2009: 25-32.
- [7] 刘 佳, 金福江. 基于 LS-SVM 的棉针织物染色上染率模型的研究[J]. 福州大学学报, 2010(6): 446-449.
- [8] 韩 敏, 王亚楠. 求解非线性回归问题的 Newton 算法[J]. 计算机学报, 2010(5): 841-846.
- [9] 邓乃扬, 田英杰. 数据挖掘中的新方法-支持向量机[M]. 北京: 科学出版社, 2004: 24-26.
- [10] 陈国良, 王煦法, 庄镇泉, 等. 遗传算法及其应用[M]. 北京: 人民邮电出版社, 2001: 3-10.
- [11] Chang Chih-chung, Lin Chi-jen. LIBSVM: a library for support vector machines[EB/OL]. (2010-01-10). [http: //www. csie. ntu. tw/~cjlin/libsvm](http://www.csie.ntu.tw/~cjlin/libsvm).

## Research on Modeling of Dye-Uptake Rate for Disperse Dyes on Polyester Fibers Based on SVM

WU Jing<sup>a</sup>, TIAN Yan-jie<sup>b</sup>, WANG Lan<sup>a</sup>, LING Jun-xiong<sup>b</sup>

(Zhejiang Sci-Tech University, a. The Key Laboratory of Advanced Textile Materials and Manufacturing Technology, Ministry of Education; b. School of Machinery and Automation, Hangzhou 310018, China)

**Abstract:** This paper takes the dyeing of polyester fabric at high temperature and pressure with Disperse Red FB for example, uses support vector machine (SVM) theory which shows many special advantages in solving small sample nonlinear and high dimensional pattern recognition as well as establishes the mathematical model of the relationship between dyeing process factors and dye-uptake rate. The experimental result shows that the correlation coefficient between predictive values and actual values reaches up to 0.977 8 and the errors are under 3%, which shows that the mathematical model has high accuracy and it can exactly reflect the dye-uptake rate in the dyeing process and meet the actual requirements of forecasting the dyeing results.

**Key words:** polyester fabric; disperse dyes; dye-uptake rate; support vector machine; multifactor models

(责任编辑: 许惠儿)