DOI: 10.16781/j.CN31-2187/R.20230333

·学术园地·

# 基于贝叶斯网络的随机森林优化填补算法

董鑫宇[1]，陈　琪[2]，杨志宇[1]，贺　佳[1,2*]
1. 同济大学医学院，上海 200092
2. 海军军医大学（第二军医大学）卫生勤务学系军队卫生统计学教研室，上海 200433

［摘要］ **目的**　评估并改进缺失数据处理方法，提升二分类结局预测模型性能。**方法**　模拟数据缺失场景，通过预测模型的ROC AUC及均方根误差（RMSE）共同评估直接剔除、均值填补、随机森林填补、多重填补对预测模型性能的影响，并将贝叶斯网络引入随机森林填补算法，利用变量间相关性进行填补方法的优化。**结果**　不同缺失占比下，通过AUC及RMSE均可得出贝叶斯网络优化随机森林填补算法效果最佳。此外，在缺失占比为10%～20%时，各种填补方法对预测模型的性能提升效果大体相同；当缺失占比为30%～40%时，相较于均值填补，除贝叶斯网络优化随机森林填补算法外，随机森林填补更好，其效果略优于多重填补；当缺失占比接近50%时，即使模型性能依旧较好，但填补数据逐渐偏离真实数据特征，模型的可用性下降。**结论**　贝叶斯网络优化随机森林填补算法总体效果较好，当随机缺失占比30%～40%时可优先考虑。

［关键词］ 预测模型；缺失填补；随机森林；贝叶斯网络
［引用本文］ 董鑫宇，陈琪，杨志宇，等. 基于贝叶斯网络的随机森林优化填补算法[J]. 海军军医大学学报，2025，46（2）：253-257. DOI: 10.16781/j.CN31-2187/R.20230333.

## Bayesian network optimized random forest imputation algorithm

DONG Xinyu[1], CHEN Qi[2], YANG Zhiyu[1], HE Jia[1,2*]
1. School of Medicine, Tongji University, Shanghai 200092, China
2. Department of Military Health Statistics, Faculty of Medical Sevices, Naval Medical University (Second Military Medical University), Shanghai 200433, China

［**Abstract**］ **Objective**　To evaluate and improve missing data imputation methods to enhance the performance of binary classification prediction model. **Methods**　By simulating data missing scenes, the effects of direct elimination, mean imputation, random forest (RF) imputation algorithm, and multiple imputation-random forest (MI-RF) on the performance of the prediction model were jointly evaluated by receiver operating characteristic area under curve (AUC) and root mean square error (RMSE). Bayesian Network was introduced for the random forest imputation algorithm to optimize the imputation method using the correlations between variables. **Results**　Under different missing proportions, both AUC and RMSE indicated that Bayesian network optimized random forest (BN-RF) imputation algorithm was better. In addition, when the missing proportion was 10%-20%, various imputation methods had roughly the same improvement effect for the prediction model; when the proportion of missing data was 30%-40%, compared to the mean imputation, except for the BN-RF, RF was more effective and its effect was slightly better than MI-RF; however, when the proportion of missing data was close to 50%, even if the model performance was still appropriate, the imputation data gradually deviated from the true data features, resulting in a decrease in the usability of the model. **Conclusion**　The overall effect of BN-RF is satisfactory, and it should be chosen when random missing was 30%-40%.

［**Key words**］ prediction model; missing data imputation; random forest; bayesian network
［**Citation**］ DONG X, CHEN Q,YANG Z, et al. Bayesian network optimized random forest imputation algorithm[J]. Acad J Naval Med Univ, 2025, 46(2): 253-257. DOI: 10.16781/j.CN31-2187/R.20230333.

［收稿日期］ 2023-06-14　［接受日期］ 2023-08-24
［基金项目］ 上海市卫生健康委员会新兴交叉领域研究专项（2022JC011），上海市产业协同创新项目（2021-cyxt1-kj10）. Supported by Emerging Interdisciplinary Research Project of Shanghai Municipal Health Commission (2022JC011) and Shanghai Industrial Collaborative Innovation Project (2021-cyxt1-kj10).
［作者简介］ 董鑫宇，硕士生. E-mail: 2131239@tongji.edu.cn
*通信作者（Corresponding author）. Tel: 021-81871441, E-mail: hejia63@yeah.net

在构建预测模型时，当前超过 40% 的研究从患者数据的描述到统计建模方法都忽略了对缺失数据处理方法的评估[1]。近乎半数的研究未报告缺失数据处理方法，其余研究中常用的处理方法是完全案例分析，采用多重填补方法者只占 10%[2-3]。目前对于构建预测模型时缺失问题的处理方法仍没有明确指导。

对于缺失数据的处理主要有剔除和填补 2 种思路，而直接剔除会降低数据样本量，导致有效信息流失。目前常用的缺失填补方法有均值填补、热卡填补、最大期望算法、回归填补、随机森林填补、多重填补，针对填补方法的评价大多基于数据处理的中间过程，即评价数据本身的预测准确性[4-6]，未见填补方法对预测模型性能影响的相关研究。

本研究提出并评估了一种新方法，即贝叶斯网络优化随机森林填补算法，该法基于贝叶斯网络筛选变量，利用缺失变量的相关变量进行填补，从而优化随机森林填补算法。本研究还对比了均值填补、随机森林填补、多重填补对预测模型性能的影响，旨在解决构建预测模型时的缺失数据问题。

## 1　方法和原理

1.1　*模拟研究*　数据模拟和统计分析使用 R 4.3.0 软件进行，涉及以下 4 个步骤。

1.1.1　*数据生成*　根据真实数据场景中变量类型及相关性模拟完整的数据集。模仿脓毒症患者的真实数据生成样本量为 1 000 例的完整数据集，因变量定义为二分类变量。20 个自变量包括基本信息（性别、年龄）、基础生命体征（体温、心率、呼吸频率、动脉收缩压、动脉舒张压）、实验室指标（血糖、血肌酐、血红蛋白、血钠、白细胞、血小板、24 h 尿量、血尿素）、病情分析变量（合并严重疾病、有疾病史、意识不清、序贯器官衰竭评分、ICU 体征评分）。应用 Kolmogorov-Smirnov 检验检测数据的正态性，正态分布的计量资料以 $\bar{x}\pm s$ 表示，两组间比较采用 $t$ 检验；计数资料以构成比表示，组间比较采用 $\chi^2$ 检验或 Fisher 精确概率法。检验水准（$\alpha$）为 0.05。

1.1.2　*模拟缺失*　基于完整数据集模拟不同缺失场景以生成缺失数据集。使用 R 4.3.0 软件的 simFrame 包对数据做缺失模拟，在模拟数据集中分别根据缺失变量名称（呼吸频率、动脉收缩压、血清钠、意识不清、ICU 体征评分）、缺失机制（随机缺失）和缺失比例（10%、20%、30%、40%、50%）模拟设计多变量混合缺失场景。大多数缺失数据方法都是为了解决随机缺失假设而设计的，因此在随后的填补方法研究中主要报告缺失数据为随机缺失的情况。

1.1.3　*缺失填补*　通过不同填补方法分别生成完整数据集，分别采用均值填补、随机森林填补、多重填补、贝叶斯网络优化随机森林填补这 4 种填补方法得到填补后的完整数据集。

1.1.4　*模型构建*　将填补后数据集用于构建预测模型并评估性能。分别用原始完整数据集和相应的填补数据集构建预测模型，分析各填补方法下相应预测模型的性能。为确保指标有意义，对参于二分类模型构建的因素进行分析。检验水准（$\alpha$）为 0.05。

1.2　*缺失填补方法*

1.2.1　*均值填补*　缺失的预测值由各自的平均值（或分类变量的众数）估算[7]。

1.2.2　*随机森林填补*　先使用均值代替缺失值拟合随机森林模型，再将该模型应用于缺失值的预测，并替换掉最初的均值，如此反复多次迭代[8]。

1.2.3　*多重填补*　对数据集中每个缺失值都构造 $m$（$m>1$）个填补值，综合 $m$ 次分析结果得到最终结果。可选参数包括填补次数及填补模型，为便于对比优化效果，本研究模型选用多重插补－随机森林（multiple imputation-random forest，MI-RF），填补次数为 5 次[9]。

1.2.4　*贝叶斯网络优化随机森林（Bayesian network optimized random forest，BN-RF）填补算法*　通过贝叶斯网络学习算法得到变量间的结构网络和条件概率表，提供变量间准确的相似度[10]。而基于随机森林的缺失数据填补 R 4.3.0 软件的 missForest，是一种存在复杂相互作用和非线性关系时不需要指定变量分布的填补算法，因此利用贝叶斯网络将缺失变量与其存在强相关的变量进行组合，再利用变量间相关性的最优组合进行随机森林填补，可更加灵活地推理出缺失数据的预测值[11-13]。

1.3　*预测模型构建方法*　Lasso-logistic 回归方法，主要通过 Lasso 方法解决变量选择问题，将用线性

回归得到的某连续数值结果转变成区间为（0，1）的概率值，进而处理分类问题[14]。

1.4　*模型评价方法*　采用内部验证方法。将数据集随机分为训练集和验证集，分别进行同样的缺失填补操作。本研究主要关注填补方法对预测模型性能的影响，采用预测模型的 ROC AUC 及均方根误差（root mean square error，RMSE）作为主要评估指标[15]。

## 2　结　果

2.1　*研究数据基线特征*　根据脓毒症患者的生存结局将患者分为生存组（$n$=694）和死亡组（$n$=306），分析脓毒症患者病情与各指标的关系。如表 1 所示，除血红蛋白、血钠及 ICU 体征评分这 3 个变量外，其余变量在生存组与死亡组间的差异均有统计学意义（均 $P$<0.05）。

**表 1　研究数据分布特征**

| 变量 | 所有患者 $N$=1 000 | 生存组 $N$=694 | 死亡组 $N$=306 | $P$ 值 |
|---|---|---|---|---|
| 男，$n$ (%) | 512 (51.2) | 372 (53.6) | 140 (45.8) | 0.022 |
| 年龄/岁，$\bar{x}\pm s$ | 51.5±18.8 | 49.4±18.3 | 56.2±19.0 | <0.001 |
| 基础生命体征，$\bar{x}\pm s$ | | | | |
| 体温/℃ | 38.1±1.6 | 37.9±1.5 | 38.7±1.7 | <0.001 |
| 心率/$min^{-1}$ | 95.9±18.7 | 92.2±17.6 | 104.3±18.6 | <0.001 |
| 呼吸频率/$min^{-1}$ | 19.3±1.9 | 18.9±1.9 | 20.1±1.6 | <0.001 |
| 收缩压/mmHg | 103.9±16.2 | 102.3±13.7 | 107.6±20.2 | <0.001 |
| 舒张压/mmHg | 77.2±10.0 | 77.7±9.6 | 76.2±10.8 | 0.022 |
| 实验室指标，$\bar{x}\pm s$ | | | | |
| 血糖/(mmol•$L^{-1}$) | 9.8±2.8 | 9.6±2.8 | 10.2±2.8 | 0.002 |
| 血肌酐/(μmol•$L^{-1}$) | 146.7±52.2 | 141.4±46.9 | 159.1±60.1 | <0.001 |
| 血红蛋白/(g•$L^{-1}$) | 132.7±30.8 | 132.1±30.8 | 134.0±31.9 | 0.378 |
| 血钠/(mmol•$L^{-1}$) | 139.5±29.4 | 139.3±30.1 | 140.0±27.8 | 0.711 |
| 白细胞计数/($L^{-1}$, ×$10^9$) | 15.3±2.3 | 15.3±1.5 | 16.7±2.9 | <0.001 |
| 血小板计数/($L^{-1}$, ×$10^9$) | 74.8±36.8 | 73.1±36.0 | 78.7±38.4 | 0.028 |
| 24 h 尿量/mL | 1 117.5±447.3 | 1 172.3±438.3 | 992.7±441.7 | <0.001 |
| 血尿素/(mmol•$L^{-1}$) | 5.6±1.9 | 5.4±1.7 | 6.0±2.1 | <0.001 |
| 病情分析变量 | | | | |
| 合并严重疾病，$n$ (%) | 459 (45.9) | 296 (42.7) | 163 (53.3) | 0.002 |
| 有疾病史，$n$ (%) | 475 (47.5) | 314 (45.2) | 161 (52.6) | 0.032 |
| 意识不清，$n$ (%) | 506 (50.6) | 383 (55.2) | 123 (40.2) | <0.001 |
| SOFA 评分/分，$\bar{x}\pm s$ | 7.3±3.1 | 6.9±3.0 | 8.1±3.2 | <0.001 |
| ICU 体征评分[a]，$n$ (%) | 555 (55.5) | 392 (56.5) | 163 (53.3) | 0.110 |

1 mmHg=0.133 kPa. [a]: 急性生理与慢性健康评分≥10 分. SOFA：序贯器官衰竭；ICU：重症监护病房.

2.2　*研究数据相关因素分析*　单因素分析结果显示，男性、舒张压、血钠、尿量、意识不清、ICU 体征评分均与死亡风险呈负相关（均 $P$<0.05），年龄、体温、呼吸频率、收缩压、血糖、血肌酐、白细胞计数、血小板计数、血尿素、合并严重疾病、有疾病史、SOFA 评分均与死亡风险呈正相关（均 $P$<0.05）。见表 2。

2.3　*不同填补方法对预测模型性能的影响*　如表 3，用原始完整数据集构建的预测模型 AUC 值为 0.892（95% *CI* 0.854～0.931），RMSE 值为 0.344 5；当缺失占比为 10% 时，直接剔除的 AUC 值为 0.897（95% *CI* 0.850～0.945），RMSE 值为 0.317 9；当缺失占比＞10% 时，则 4 种填补方法的 AUC 值均高于直接剔除、RMSE 值均低于直接剔除。

4 种填补方法比较结果提示，预测模型性能较好的是 BN-RF。当缺失占比为 10% 时，AUC 值为 0.891（95% *CI* 0.858～0.924），RMSE 值为 0.3481；当缺失占比为 20% 时，AUC 值为 0.887（95% *CI* 0.852～0.922），RMSE 值为 0.341 9；当缺失占比为 30% 时，AUC 值为 0.899（95% *CI* 0.868～0.931），RMSE 值为 0.339 3；当缺失占比为 40% 时，AUC 值为 0.881（95% *CI* 0.846～0.917），RMSE 值为 0.350 5；当缺失占比为 50% 时，AUC 值为 0.884（95% *CI* 0.850～0.919），RMSE 值为 0.352 6。

**表 2　各因素风险特征单因素分析结果**

| 因素 | *OR* ( 95% *CI*) | *P* 值 |
|---|---|---|
| 男 | 0.68 (0.47, 0.98) | 0.039 |
| 年龄 | 1.02 (1.01, 1.03) | 0.001 |
| 基础生命体征 | | |
| 体温 | 1.45 (1.28, 1.64) | ＜0.001 |
| 心率 | 1.02 (1.01, 1.04) | 0.002 |
| 呼吸频率 | 1.30 (1.11, 1.52) | ＜0.001 |
| 收缩压 | 1.03 (1.02, 1.05) | ＜0.001 |
| 舒张压 | 0.95 (0.93, 0.97) | ＜0.001 |
| 实验室指标 | | |
| 血糖 | 1.01 (1.00, 1.02) | 0.001 |
| 血肌酐 | 2.71 (1.94, 3.79) | ＜0.001 |
| 血红蛋白 | 1.07 (1.01, 1.14) | 0.031 |
| 血钠 | 0.98 (0.97, 1.00) | 0.003 |
| 白细胞计数 | 1.53 (1.40, 1.67) | ＜0.001 |
| 血小板计数 | 1.01 (1.00, 1.01) | 0.014 |
| 24 h 尿量 | 1.00 (0.99, 1.00) | ＜0.001 |
| 血尿素 | 1.03 (1.01, 1.05) | ＜0.001 |
| 程度分析变量 | | |
| 合并严重疾病 | 1.81 (1.24, 2.64) | 0.002 |
| 有疾病史 | 1.72 (1.18, 2.50) | 0.004 |
| 意识不清 | 0.51 (0.35, 0.74) | ＜0.001 |
| SOFA 评分 | 1.19 (1.12, 1.27) | ＜0.001 |
| ICU 体征评分 | 0.71 (0.57, 0.90) | 0.004 |

*OR*：比值比；*CI*：置信区间；SOFA：序贯器官衰竭；ICU：重症监护病房.

**表 3　基于完整数据集直接剔除及 4 种填补方法构建的预测模型性能**

| 模型构建条件 | AUC (95% *CI*) | RMSE |
|---|---|---|
| 完整数据集 | 0.892 (0.854, 0.931) | 0.344 5 |
| 缺失 10% | | |
| 直接剔除 | 0.897 (0.850, 0.945) | 0.317 9 |
| 均值填补 | 0.885 (0.851, 0.919) | 0.350 8 |
| RF 填补 | 0.885 (0.850, 0.920) | 0.349 1 |
| MI-RF 填补 | 0.881 (0.846, 0.916) | 0.354 0 |
| BN-RF 填补 | 0.891 (0.858, 0.924) | 0.348 1 |
| 缺失 20% | | |
| 直接剔除 | 0.877 (0.813, 0.940) | 0.361 6 |
| 均值填补 | 0.878 (0.841, 0.915) | 0.349 0 |
| RF 填补 | 0.883 (0.846, 0.920) | 0.344 9 |
| MI-RF 填补 | 0.881 (0.844, 0.917) | 0.348 3 |
| BN-RF 填补 | 0.887 (0.852, 0.922) | 0.341 9 |
| 缺失 30% | | |
| 直接剔除 | 0.759 (0.643, 0.875) | 0.448 6 |
| 均值填补 | 0.887 (0.853, 0.921) | 0.353 1 |
| RF 填补 | 0.896 (0.863, 0.929) | 0.341 8 |
| MI-RF 填补 | 0.891 (0.857, 0.924) | 0.345 1 |
| BN-RF 填补 | 0.899 (0.868, 0.931) | 0.339 3 |
| 缺失 40% | | |
| 直接剔除 | 0.740 (0.567, 0.914) | 0.612 4 |
| 均值填补 | 0.867 (0.829, 0.905) | 0.359 9 |
| RF 填补 | 0.872 (0.834, 0.910) | 0.355 7 |
| MI-RF 填补 | 0.870 (0.832, 0.908) | 0.355 2 |
| BN-RF 填补 | 0.881 (0.846, 0.917) | 0.350 5 |
| 缺失 50% | | |
| 直接剔除 | 0.554 (0.174, 0.935) | 0.685 5 |
| 均值填补 | 0.872 (0.835, 0.910) | 0.359 9 |
| RF 填补 | 0.877 (0.843, 0.911) | 0.353 1 |
| MI-RF 填补 | 0.874 (0.837, 0.911) | 0.353 8 |
| BN-RF 填补 | 0.884 (0.850, 0.919) | 0.352 6 |

AUC：曲线下面积；*CI*：置信区间；RMSE：均方根误差；RF：随机森林；MI-RF：多重插补-随机森林；BN-RF：贝叶斯网络优化随机森林.

## 3　讨　论

构建预测模型时，如忽略数据缺失问题会对结果造成一定影响，使用合适的填补策略不仅能保证样本量，还能够提升预测模型的性能。

本研究对比分析了直接剔除与均值、随机森林、MI-RF 及 BN-RF 填补 4 种方法对预测模型性能的影响。预测模型的效果受数据特征的影响，而缺失占比不同会改变原有数据特征，因此本研究根据缺失占比逐一对 4 种方法进行评估。结果显示，使用 BN-RF 填补能够在缺失占比 10%～40% 时保持较好的效果。

本研究的创新点在于将贝叶斯网络学习算法用于缺失数据的填补前处理，最大程度利用已知信息分析计算得到变量间相关性较强的组合，并利用变量间的强相关性进行缺失数据的填补。同时，随机森林填补也是利用已知信息进行填补，在其之前进行多一层变量筛选可避免无效变量带来的干扰，使填补过程更具有目的性，更有效地解决数据的缺失问题。

此外，对比其他 3 种填补方法发现，在缺失占比为 10%～20% 时，各种填补方法对预测模型的性能提升效果大体相同；当缺失占比为 30%～40% 时，相较于均值填补，除 BN-RF 填补外，随机森林填补和 MI-RF 填补是效果较好的，且随机森林填补的效果略优于 MI-RF 填补，这与杨弘等[4] 的结果一致；然而，从直接剔除后的模型性能指标可见，缺失占比为 50% 时所做的填补工作基本已改变原有数据特征，此时 4 种填补方法所得结果均无法得出有效结论。

综上所述，在缺失占比≤20% 时，采用均值填补可解决数据缺失问题；当缺失占比为 30%～40% 时，相较随机森林填补与 MI-RF 填补，BN-RF 填补在提升预测模型性能上有更好的效果；当缺失占比≥50% 时，与单纯评估填补方法的结果[4-6] 不一致，预测模型的性能依然很好，但有效数据或可利用数据过少，无论采用哪种填补方法均无法更好地体现原有数据特征。

本研究模拟缺失场景与实际缺失问题存在一定的差异，且主要关注的是多变量混合随机缺失的情况，后续将针对各种复杂缺失场景对预测模型性能的影响展开进一步研究。

## ［参 考 文 献］


［1］ TSVETANOVA A, SPERRIN M, PEEK N, et al. Missing data was handled inconsistently in UK prediction models: a review of method used[J]. J Clin Epidemiol, 2021, 140: 149-158. DOI: 10.1016/j.jclinepi.2021.09.008.

［2］ MASCONI K L, MATSHA T E, ECHOUFFO-TCHEUGUI J B, et al. Reporting and handling of missing data in predictive research for prevalent undiagnosed type 2 diabetes mellitus: a systematic review[J]. EPMA J, 2015, 6(1): 7. DOI: 10.1186/s13167-015-0028-0.

［3］ NIJMAN S, LEEUWENBERG A M, BEEKERS I, et al. Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review[J]. J Clin Epidemiol, 2022, 142: 218-229. DOI: 10.1016/j.jclinepi.2021.11.023.

［4］ 杨弘，田晶，王可，等. 混合型缺失数据填补方法比较与应用[J]. 中国卫生统计，2020，37(3)：395-399.

［5］ 郑智泉，陈妍，王孟孟，等. 不同缺失率下的数据填补算法稳定性研究[J]. 统计与决策，2023，39(8)：12-17. DOI: 10.13546/j.cnki.tjyjc.2023.08.002.

［6］ 宋亮，万建洲. 缺失数据插补方法的比较研究[J]. 统计与决策，2020，36(18)：10-14. DOI: 10.13546/j.cnki.tjyjc.2020.18.002.

［7］ ZHANG Z. Missing data imputation: focusing on single imputation[J]. Ann Transl Med, 2016, 4(1): 9. DOI: 10.3978/j.issn.2305-5839.2015.12.38.

［8］ TANG F, ISHWARAN H. Random forest missing data algorithms[J]. Stat Anal Data Min, 2017, 10(6): 363-377. DOI: 10.1002/sam.11348.

［9］ JAVADI S, BAHRAMPOUR A, SABER M M, et al. Evaluation of four multiple imputation methods for handling missing binary outcome data in the presence of an interaction between a dummy and a continuous variable[J]. J Probab Stat, 2021, 2021: 6668822. DOI: 10.1155/2021/6668822.

［10］ 蔡金成，孙浩军. 基于互信息与贝叶斯信念网络的关系层次距离混合聚类算法[J]. 汕头大学学报(自然科学版)，2018，33(2)：2，3-12.

［11］ 王旭春，宋伟梅，潘金花，等. MMPC-Tabu混合算法的贝叶斯网络模型在高脂血症相关因素研究中的应用[J]. 中国卫生统计，2022，39(3)：345-350，355. DOI: 10.3969/j.issn.1002-3674.2022.03.005.

［12］ HONG S, LYNN H S. Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction[J]. BMC Med Res Methodol, 2020, 20(1): 199. DOI: 10.1186/s12874-020-01080-1.

［13］ LI Y F, HUANG H Z, MI J, et al. Reliability analysis of multi-state systems with common cause failures based on Bayesian network and fuzzy probability[J]. Ann Oper Res, 2022, 311(1): 195-209. DOI: 10.1007/s10479-019-03247-6.

［14］ 秦瑶，韩红娟，陈杜荣，等. 基于LASSO logistic回归模型的轻度认知障碍逆转预测模型[J]. 中国卫生统计，2022，39(5)：653-658. DOI: 10.3969/j.issn.1002-3674.2022.05.003.

［15］ JANSSENS A C J W, MARTENS F K. Reflection on modern methods: Revisiting the area under the ROC Curve[J]. Int J Epidemiol, 2020, 49(4): 1397-1403. DOI: 10.1093/ije/dyz274.


［本文编辑］ 尹　茶