**Computer Science and Application** 计算机科学与应用, 2024, 14(5), 184-193
Published Online May 2024 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2024.145127

Hans汉斯

# 一种军事人才发展潜力评估模型构建方法

**孙二洋，李玉泉，王齐天，杨博文**

中国电子科技集团公司第二十八研究所，江苏 南京

收稿日期：2024年4月26日；录用日期：2024年5月24日；发布日期：2024年5月31日

## 摘　要

**针对军事背景下，可利用的数据量有限、人才发展潜力难以评估的问题，提出一种有效的军事人才发展潜力评估模型构建方法。为了有效提升模型训练可用数据量，采用了按照重要时间节点对关键数据进行分片的思想，成倍的增加数据集；在模型构建方面，为有效获得影响人才发展潜力的重要因素，使用随机森林算法进行模型训练。最后通过训练生成的模型与专家经验两种方式，针对同一批数据进行人才发展潜力评估，进行相互验证，证明了构建生成的模型有效性。**

## 关键词

**军事人才，发展潜力评估，数据有限，随机森林，评估模型构建**

# A Method for Constructing an Evaluation Model for Development Potential of Military Talents

**Eryang Sun, Yuquan Li, Qitian Wang, Bowen Yang**

The 28th Research Institute of China Electronics Technology Group Corporation, Nanjing Jiangsu

Received: Apr. 26th, 2024; accepted: May 24th, 2024; published: May 31st, 2024

## Abstract

**In the military context, aiming at the problem of the limited amount of available data and the difficulty in evaluating talent development potential, an effective method for constructing an evaluation model for the development potential of military talents is proposed. To effectively improve the amount of available data for model training, the idea of sharding key data according to time important nodes is adopted, which exponentially increases the dataset. In terms of model construction, in order to effectively identify important factors that affect talent development potential,**

文章引用: 孙二洋, 李玉泉, 王齐天, 杨博文. 一种军事人才发展潜力评估模型构建方法[J]. 计算机科学与应用, 2024, 14(5): 184-193. DOI: 10.12677/csa.2024.145127

**the random forest algorithm is used for model training. Finally, the validity of the constructed and generated model is proved by mutual validation through both the training-generated model and the expert's experience for the same batch of data for talent development potential assessment.**

## Keywords

**Military Talents, Talent Development Potential, Limited Data, Random Forest, Evaluation Model Construction**

Copyright © 2024 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

## 1. 引言

军事领域岗位体系是一个严格按照职称等级和职务等级划分的体系，人员分布呈现金字塔型结构，越是接近底层职级人员数目越多，越是接近顶层职级人员数目越少。且该金字塔结构不是一个静止的结构，而是一个动态结构，处在其中的每个人员在其军事领域任职生涯过程中，都会经过入职、培训、升职、调岗、离职五个环节，最终完成其在该领域服役的职业生涯。

鉴于军事人员分布上少下多的金字塔型结构，必然存在一部分人具备较高的潜力发展至较高的职务等级，影响因素可能包括受教育水平较高、参加实战演习经历丰富或发表著述专利产量丰富等。同理，也必然存在相当一部分人不具备高层级发展潜质，使其在发展到金字塔的某阶层时便退出了，可能是学历水平限制了其发展、也可能是其服役年限提前达到了上限、或是其某次年度考核结果未能获得优秀等次等。因此，本模型算法通过分析历史上那些潜力高低不同人员的履历数据和画像数据，构建人才发展潜力评估模型，深入挖掘表征人员发展潜力高低的重要特征因素，通过训练后的模型进行人员发展潜力预测。

## 2. 随机森林算法原理

### 2.1. 基本原理

随机森林算法[1]是一致机器学习算法组合算法，主要用于分类和回归领域。它的主要思想是用随机的方式建立一个森林，森林里包含多个决策树，每棵决策树进行分类和回归时保持独立，并通过每个决策树的结果进行投票给出一个最终结果。在进行分类时，每一棵决策树的训练样本都是从原始训练样本集 N 中有放回地重复随机抽取得到的。将多个决策树合并在一起，让每棵树的分布大致相同，每一棵树的分类能力和它们之间的相关性的大小就决定了随机森林总的分类误差，使决策树算法得到了改进。尽管单棵树的分类能力可能很小，但在随机产生大量的决策树后，由多个决策树组成的随机森林的分类能力得到很大提高。另外，训练样本和特征数选择的随机性使得随机森林可以很好地预防过学习问题。

### 2.2. 核心思想

随机森林的核心思想为分类回归树，分类回归树由分类树和回归树两部分组成；当最终结果是类别变量时用分类树，当最终结果是连续变量时则用回归树。在构造每一棵树时，随机森林算法随机地去选择特征对分类树的内部节点进行属性分裂。分类树常以基尼指数作为分类标准，该标准能够选出降低数据无序度的属性。在建立分类树时，每个分裂属性的选择是根据它对数据划分的优劣程度来进行的。

### 2.3. 训练集的抽取

Bagging (Bootstrap Aggregating)算法[2]的基础是自助抽样法，即从原始样本集 S 中有放回地随机抽取训练样本集 train。Bagging 算法的主要思想是：事先给顶一个元学习的算法和一个原始样本集 S，让该元学习算法通过上述的自助抽样法从原始样本集 S 中随机抽取的训练集训练多次，每个训练集构造一颗决策树模型。

## 3. 构建人才发展潜力评估模型

评估人员发展潜力，主要有两种技术路线：一种是基于先验知识的评估方法，通过人(专家)的经验聚焦于被评估对象的各项属性表征，最后进行综合；另一种是基于机器学习的评估方法，通过样本学习获得被评估对象的各项属性表征，最后由机器学习模型给出综合结果。本模型使用的是第二种技术路线，即通过有监督机器学习的方式挖掘数据规律，实现人员发展潜力预测。但在进行验证时，我们选取了相同的试验数据，经过相应专家按照其经验进行评估，以期实现人才发展潜力预测结果的相互印证。本模型的构建主要分为四个处理步骤，分别为数据梳理、构造样本、特征初选和模型训练，如图 1 所示。

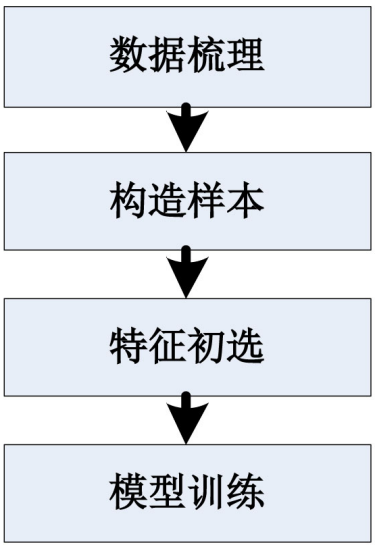


**Figure 1.** Model algorithm processing steps
**图 1.** 模型算法处理步骤

### 3.1. 数据梳理

数据梳理是后续模型训练的基础和前提，目的是对当前数据体系进行梳理，理清数据之间的关联关系，将数据对象转化为模型训练对象，并通过分片的思想扩充可用数据集。

人才发展潜力评估模型依赖的数据主要是人员岗位发生变化前后的人员的整体画像数据。因此我们在准备数据样本时，针对某个军事人员，按照时间轴，以职位发生变化的时间为分割点，获取其岗位发生变化前的各类因素数据，形成一个数据集，按照此类方法进行数据集生成，每个军事人员在每次发生岗位变动时，即可得到一个对应的数据集，从而扩充了可用数据量。数据拆分示例如图 2 所示，形成的数据集如图 3 所示。从图上可看出，通过梳理和分析，各类型数据被有机的关联在一起，形成一套完整的、维度多元的人员发展潜力数据用于后续模型训练和分析。

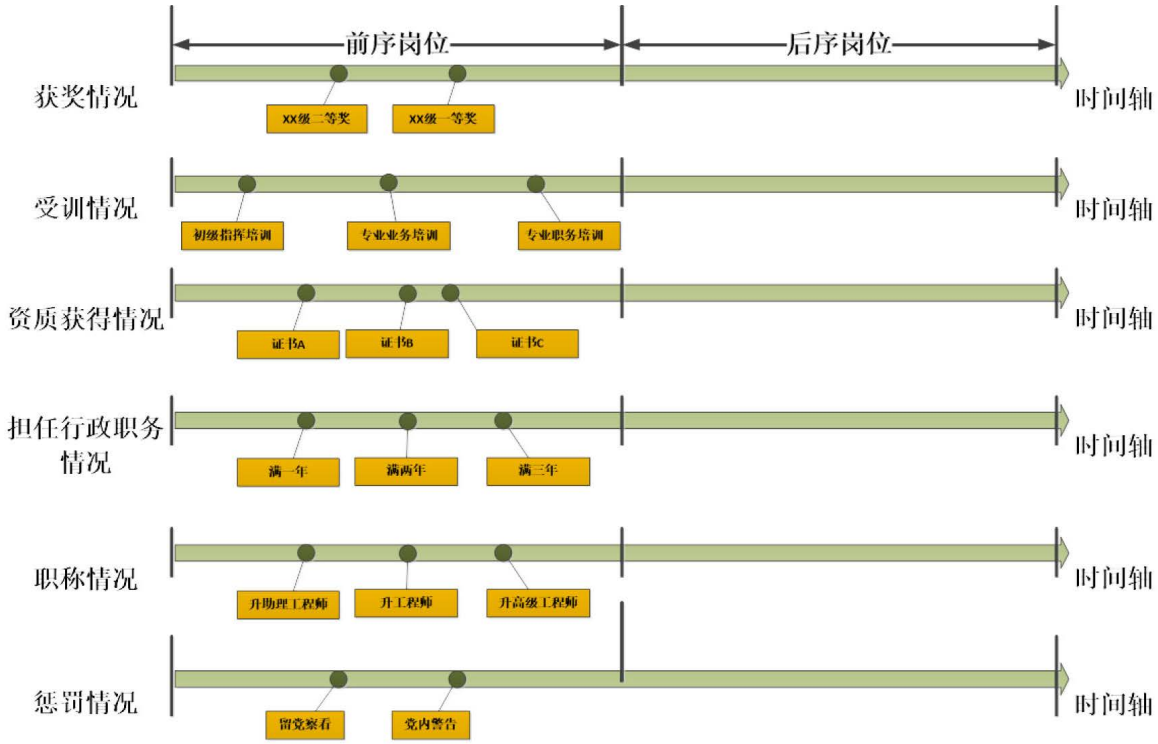


**Figure 2.** Work resume data
**图 2.** 人员部队工作履历数据


**Figure 3.** Cross temporal personnel profile data
**图 3.** 跨时间维度的人员画像数据

## 3.2. 构造样本

模型训练的基础是样本集合，通过上一步“数据梳理”可以将两万多条人员相关数据转化为近十九万条样本对象，下一步即可对这些对象进行正负样本的标注和区分。本模型采用自动标注为主，手动标注为辅的方式进行样本标注，如图 4 所示。标注规则为：

a) 正样本：历史上具备较高发展潜力的人员。

历史上具备较高发展潜力的人员特指在职业生涯的某一阶段其岗位职务等级发生了升级变动，即升

职了的那些人员。

b) 负样本：历史上不具备较高发展潜力的人员。

历史上不具备较高发展潜力的人员特指在职业生涯的某一阶段其岗位职务等级发生了降级变动、退伍退役等其他一切非升职类的状态改变。

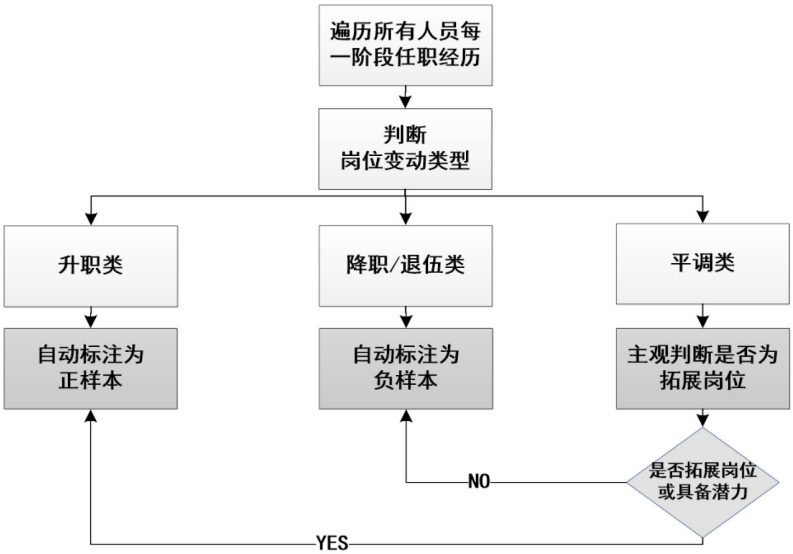


**Figure 4.** Sample labeling flowchart
**图 4.** 样本标注流程图

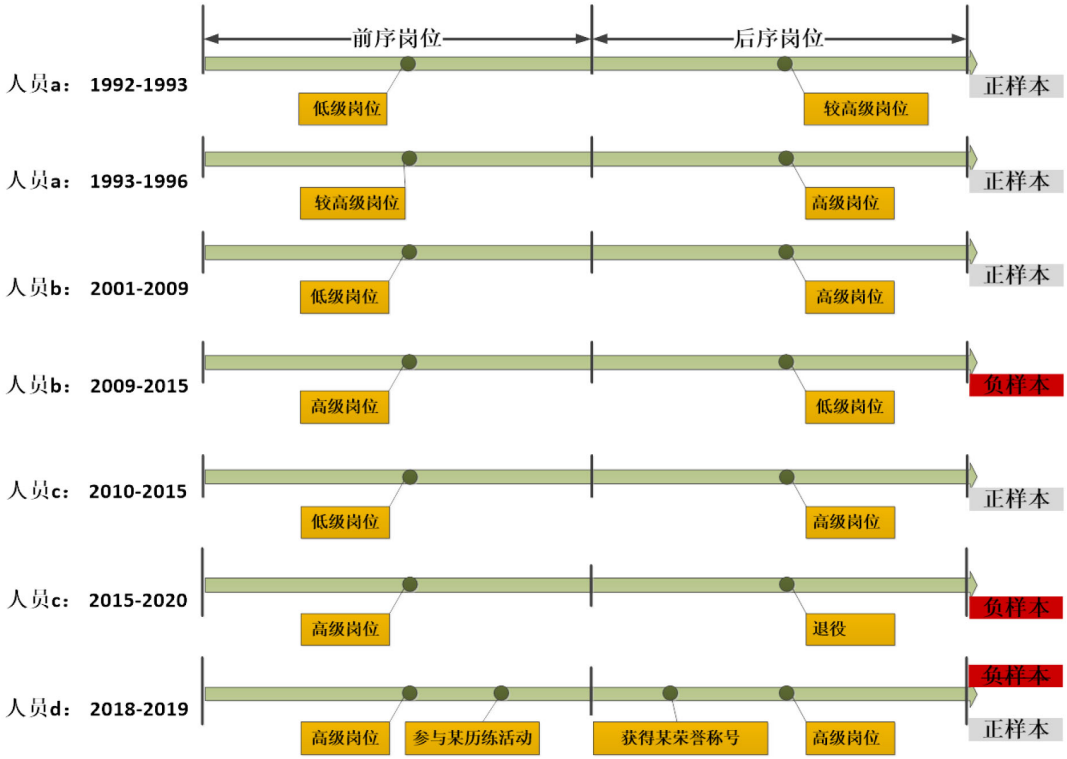


**Figure 5.** Sample and annotation diagram
**图 5.** 样本及其标注示意图

样本标注结果如图 5 所示，其中大部分人员在职业生涯一些阶段发生了升职变动而被程序自动标注为正样本，也存在部分人员虽未发生职务等级上的升级，但结合该人员平时的综合表现以及以历练为目的的岗位调用(拓展岗位)情况，我们判定该人员在该时期属于正样本范畴。

### 3.3. 特征初选

人员每一个指标项的变化在其职业生涯中都可映射为一个时间节点分布在时间轴上，同时这些指标项发生变化的时间节点也必然属于人员职业生涯发展的某一阶段或某一岗位期间，这些变化的因素极可能成为预测人员发展潜力的重要指标项，同时构成人员晋升任用的先决条件，因此在进行特征初选时，尽可能结合对业务知识的理解。本次模型构建过程中，选择的指标项如表 1 所示。

**Table 1.** Indicator description
**表 1.** 指标项说明

| 序号 | 特征名称 | 含义 |
|---|---|---|
| **1.** | 性别 | —— |
| **2.** | 入伍时间 | —— |
| **3.** | 入伍年限 | 工作经历命令时间减去入伍时间 |
| **4.** | 身份类别 | —— |
| **5.** | 来源类别 | —— |
| **6.** | 身份管理状态 | —— |
| **7.** | 政治面貌 | —— |
| **8.** | 入党时间/入团时间 | —— |
| **9.** | 学历 | —— |
| **10.** | 培训形式 | —— |
| **11.** | 毕业院校 | —— |
| **12.** | 入学时间 | —— |
| **13.** | 毕业时间 | —— |
| **14.** | 学位 | —— |
| **15.** | 全日学历 | —— |
| **16.** | 全日学位 | —— |
| **17.** | 人员受奖次数 | 当前职务阶段人员获奖次数 |
| **18.** | 人员累计受奖次数 | 计算工作经历时间内累计次数 |
| **19.** | 人员受惩次数 | 当前职务阶段内次数 |
| **20.** | 人员累计受惩次数 | 计算工作经历时间内累计次数 |
| **21.** | 德才表现次数 | 当前职务阶段内有德才表现的记录数 |
| **22.** | 德才累计表现次数 | 计算工作经历时间内有德才表现的累计记录数 |
| **23.** | 参加重要活动次数 | 当前职务阶段参加重要活动记录数 |
| **24.** | 参加重要活动累计次数 | 计算工作经历时间内参加重要活动的累计记录数 |
| **25.** | 立功次数 | 当前职务阶段完成任务并立功的次数 |
| **26.** | 累计立功次数 | 计算工作经历时间内完成任务并立功的次数 |
| **27.** | 获奖得分 | 三等奖得一分，二等奖得两分，一等奖得三分，阶段内得分累计相加 |
| **28.** | 奖累计得分 | 三等奖得一分，二等奖得两分，一等奖得三分，工作经历时间内得分累计相加 |
| **29.** | 人员科技奖励次数 | 计算阶段内人员获得科技奖励次数 |
| **30.** | 人员科技奖励累计次数 | 计算工作经历时间内人员获得科技奖励累计次数 |
| **31.** | 受训次数 | 计算本阶段内参与的培训次数 |
| **32.** | 受训累计次数 | 计算工作经历时间内参与的培训累计次数 |
| **33.** | 著述发表次数 | 计算本阶段内发表的著作次数 |

续表

| | | |
|---|---|---|
| **34.** | 著述发表累计次数 | 计算工作经历时间内发表的著作累计次数 |
| **35.** | 当前技能类职业技术等级 | —— |
| **36.** | 当前技师类职业技能等级 | —— |
| **37.** | 当前岗位时间 | 特指该人员在该任职阶段的任职时间 |
| **38.** | 调岗次数 | 特指该人员在职业生涯某阶段之前岗位变动累计次数 |
| **39.** | 升级次数 | 特指该人员在职业生涯某阶段之前职务等级累计升级次数 |
| **40.** | 职务变化次数 | 特指该人员在职业生涯某任职阶段升衔的次数 |

## 3.4. 模型训练

在上一步骤初步选出的数量众多且可能冗余的特征当中蕴含着表征人员发展潜力的重要特征信息，该信息可通过机器学习算法进行分析挖掘。本文选取随机森林模型算法进行模型训练。

随机森林模型是应用机器学习领域内一个强有力的数据挖掘模型，模型由多个决策树组成，在训练完成后，它可以指出哪些特征比较重要，在处理人员履历样本的多维度数据上相对其它算法有着很大的优势，表现良好。随机森林实际上是一种特殊的抽样集成(Bagging)方法，它将决策树用做 bagging 中的模型。其中 bagging 的名称来源于(Bootstrap Aggregating)，这种方法将训练样本集合分成 m 个新的训练样本集合，然后在每个新的训练集上构建一个模型，各自独立，最后预测时候再将这 m 个模型的结果进行整合得到最终结果。而随机森林就是用自助(Bootstrap)方法生成 m 个训练集，然后对于每个训练集，构造一颗决策树模型，在节点找特征进行分裂的时候，并不是对所有特征找到信息增益最大的，二是在特征中随机抽取一部分特征，在抽到的特征中找到最优解，应用于节点，进行分裂。随机森林的方法的思想所在，实际上相当于有了 Bagging，也就是集成于对样本和特征都进行了采样，所以可以避免过拟合。模型训练原理如图 6 所示：

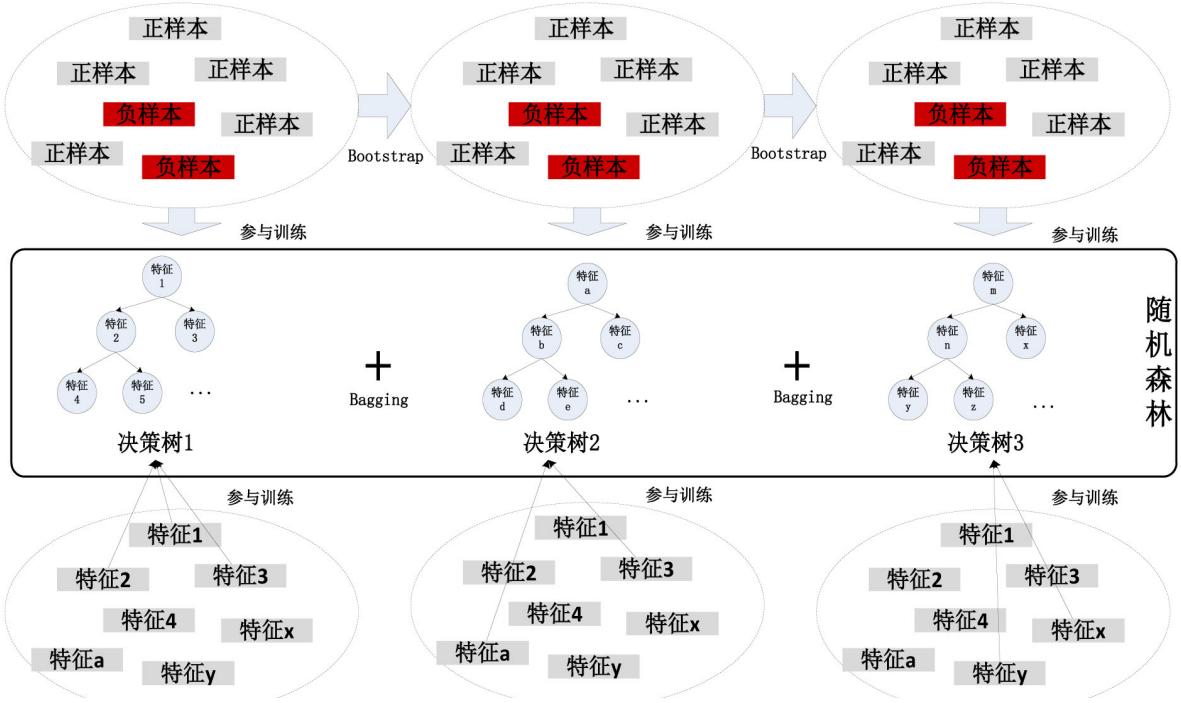


**Figure 6.** Principles of random forest training model
**图 6.** 随机森林训练模型原理

随机森林模型训练和潜力预测的处理流程如图 7 所示，本模型采用的 Bagging 方法为加权投票法，最终训练好的模型可以对人员给出发展潜力评分结果，同时随机森林各个决策树模型训练得到的具有较好分类能力的指标(决策树节点)将会被保存下来作为表征人才发展潜力的重要特征指标。

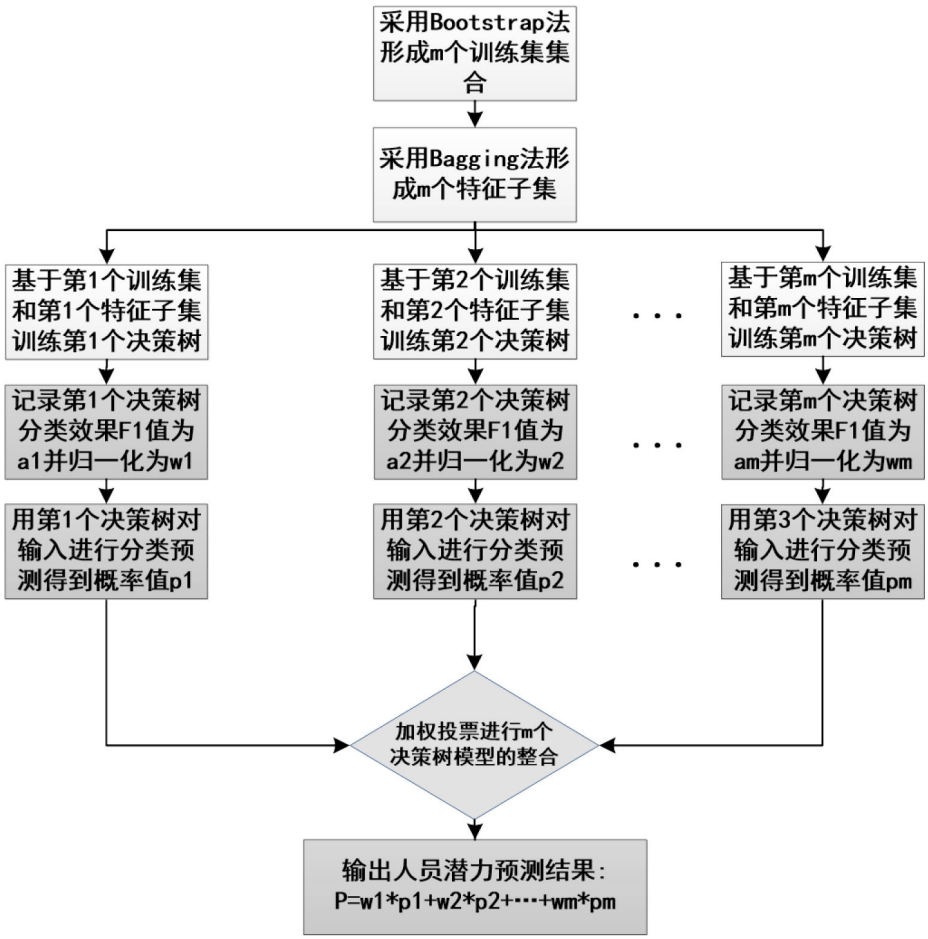


**Figure 7.** Process flow for training and potential prediction of random forest models
**图 7.** 随机森林模型训练和潜力预测处理流程

## 4. 模型检验

在训练及测试数据方面，参与模型训练的样本总计 16 万条，测试样本总计 3 万条，涉及人员近两万个。在模型评估方面，单纯靠准确率在评价模型训练效果往往不够科学全面，因此模型训练使用精确率、召回率、F1 值三个指标对模型分类效果进行评估。这些指标越高表示模型训练出的结果越好，其中召回率衡量了分类器对正例的识别能力，F1 值是对准确率和召回率的综合。

精确率：Precision = TP/(TP + FP)

召回率：Recall = TP/(TP + FN)

F1 值：F1-score = (2Recall * Precision)/(Recall + Precision)

其中：

TP：真正例，实际为正预测为正；

FP：假正例，实际为负但预测为正；

FN：假反例，实际为正但预测为负；

TN：真反例，实际为负预测为负。

模型训练结果为精确率 84%，召回率 84%，F1 值 84%，这些指标越高表示模型训练出的结果越好，对人员潜力的预测越是准确，后续随着数据质和量的提升，模型可逐步迭代并优化训练效果。模型最终训练得到的重要特征按照显著性排序如表 2、图 8 所示。

**Table 2.** Importance of characteristic factors
**表 2.** 特征因素重要性

| 序号 | 特征名称 | 相对重要性 |
|---|---|---|
| **1.** | 当前岗位任职时间 | 0.412 |
| **2.** | 调岗累计次数 | 0.186 |
| **3.** | 职务等级 | 0.0919 |
| **4.** | 职务升级累计次数 | 0.0819 |
| **5.** | 年龄 | 0.0753 |
| **6.** | 人员来源类别 | 0.0468 |
| **7.** | 职称等级 | 0.0463 |
| **8.** | 当前阶段职务变化次数 | 0.0206 |
| **9.** | 最高学历 | 0.0184 |
| **10.** | 奖励累计次数 | 0.00937 |
| **11.** | 当前阶段受训次数 | 0.00714 |
| **12.** | 德才表现累计次数 | 0.00194 |
| **13.** | 当前阶段惩罚次数 | 0.00143 |

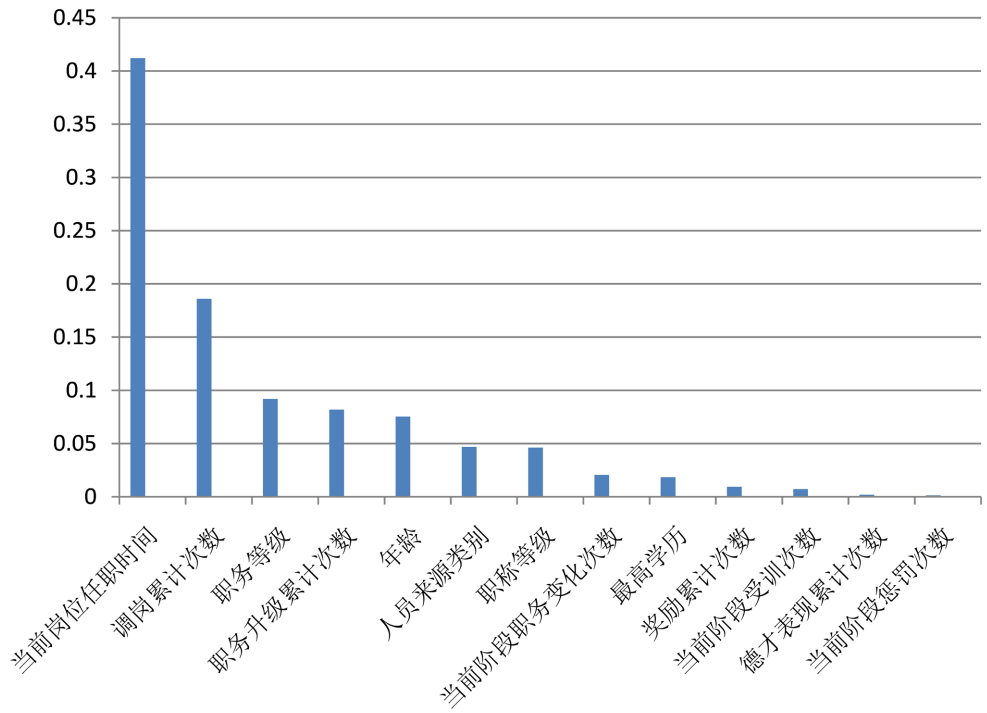


**Figure 8.** Importance of characteristic factors
**图 8.** 特征因素重要性

为对模型进行检验，本文从各个职务等级下选取典型军官，进行了人才发展潜力预测[3]，如表 3 所示。

**Table 3.** Prediction of personnel development potential
**表 3.** 人员发展潜力预测

| 序号 | 身份证号 | 年龄 | 调岗累计次数 | 职务等级 | 职称等级 | 学历水平 | 当前岗位任职时间 | 奖励累计次数 | 当前岗位培训情况 | 入伍时间 | 潜力评分 | 综合能力评分 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 90***79 | 46 | 6 | 一级专家 | 高级高程师 | 硕士研究生 | 2 年零 10 个月 | 三等奖 5 次先进个人 1 次 | 一次重要培训 | 1990 | 93.45 | 6.3 |
| 2 | 93***53 | 46 | 7 | 一级专家 | 高级高程师 | 硕士研究生 | 2 年零 11 个月 | 三等奖 5 次优秀党员 1 次 | 无 | 1993 | 93.62 | 6.2 |
| 3 | 97***76 | 41 | 4 | 二级专家 | 工程师 | 硕士研究生 | 2 年零 8 个月 | 三等奖 1 次 | — | 1997 | 89.33 | 4.7 |
| 4 | 02***56 | 41 | 5 | 二级专家 | 工程师 | 本科 | 2 年零 8 个月 | 二等奖 1 次三等奖 1 次一等奖 2 次 | 一次重要培训 | 2002 | 92.10 | 4.9 |

## 5. 结束语

本文主要从可获得使用的军事人员重要特征信息出发，结合随机森林算法，提出了一种军事人才发展潜力评估模型构建方法；为解决数据量小的问题，依据数据分片的思想，将数据成倍的增加，以满足模型训练的要求；为验证模型的准确性，与专家评估的方式进行比对，并给出了相对准确的指标，验证了方法的可行性。因目前人员特征属性及可用数据量均受相应限制，后续随着可用人员特征属性、人员数据量的增多，模型的准确性也会更加精准，更贴近具体的使用需求。

## 参考文献

[1] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. https://doi.org/10.1023/A:1010933404324
[2] Breiman, L. (1996) Bagging Predictors. *Machine Learning*, **24**, 123-140. https://doi.org/10.1007/BF00058655
[3] 陈静, 余建波, 李艳冰. 基于随机森林的用户流失预警研究[J]. 精密制造与自动化, 2021(2): 21-24, 51.