

# 中国城市就业市场的人才结构、产业特征与薪资差异分析

王若丞

管理学院统计与金融系 PB23020607

Email: rcwang@mail.ustc.edu.cn

2026 年 1 月 4 日

## 摘要

本文基于 2024-2025 年中国就业岗位数据，从城市层面分析薪资差异的结构性决定因素。通过机器学习回归、分类模型及聚类分析，分别探讨城市薪资决定机制、高薪岗位识别特征以及城市人才吸引力类型。研究发现，高端产业占比与技能多样性对城市薪资水平具有显著影响，并据此刻画出不同类型城市的人才结构特征。

## 1. 引言

近年来，随着区域经济一体化与城市竞争的不断加剧，不同城市在薪资水平、产业结构以及人才结构方面呈现出显著差异。城市之间在吸引高端人才、承载高附加值产业以及促进创新活动方面的能力差异，已成为影响区域经济长期发展的关键因素。因此，从数据角度系统刻画城市劳动力市场特征，并分析其背后的决定机制，具有重要的现实意义与政策价值。

传统的薪资决定机制研究多采用线性回归模型，在个体或岗位层面引入教育程度、行业类别、技能要求等大量解释变量。然而，在实际数据中，这类变量往往存在较强的相关性。例如，高端产业集聚程度较高的城市通常同时具有更高的学历水平和更丰富的技能结构，这容易导致模型中出现严重的多重共线性问题。尽管回归结果可能表现出较高的拟合优度，但参数估计的不稳定性和经济解释的可信度往往受到削弱。

针对上述问题，本文不再拘泥于个体层面的因果识别，而是转向城市层面的结构性分析，构建一套具有明确经济含义的城市特征指标体系，并结合机器学习与无监督学习方法，对中国城市劳动力市场进行多角度刻画。本文的核心目标在于回答以下三个问题：（1）城市层面的薪资水平主要由哪些关键结构性因素所决定？（2）那些岗位特征更可能对应高薪工作？（3）如何基于数据构建具有可解释性的城市画像，为比较分析与政策讨论提供依据？

为此，本文构建了三个相互关联但侧重点不同的模型。模型一聚焦于城市薪资决定机制，选取少量具有代表性的城市特征变量，并采用机器学习回归方法，在缓解多重共线性问题的同时，分析不同因素对薪资水平的重要性。模型二从分类角度出发，研究城市是否具备“高薪城市”特征，探讨城市结构变量在区分高薪与非高薪城市中的作用。模型三则进一步引入聚类分析方法，对城市进行综合画像刻画，识别不同类型城市在薪资水平、产业结构与技能多样性方面的系统性差异。

本文的研究特点主要体现在三个方面。首先，在变量构建层面，本文通过产业分类与技能文本处理，构建高端产业占比与技能多样性（香农指数 [1]）等综合性指标，使城市特征更具概括性与经济解释力。其次，在方法层面，本文将机器学习方法引入薪资机制分析，在保持模型解释性的同

时有效缓解多重共线性问题。最后，在研究视角上，本文以城市为分析单元，对中国劳动力市场进行结构化刻画，为城市比较研究提供了一种数据驱动的分析框架。

本文章节安排如下：第二章介绍数据来源与变量构建方法；第三章详细阐述模型设定与方法论；第四章给出实证结果与分析；第五章对研究结论进行总结，并讨论其政策含义与研究局限。

## 2. 数据与变量构建

### 2.1 数据来源与样本说明

本文所使用的数据来源于整理后的中国招聘岗位数据集（`china_job_market_2025.csv`），数据以招聘岗位为基本观测单元，涵盖岗位所属城市、行业类型、技能要求、学历要求以及薪资水平等信息。该数据能够较为全面地反映不同城市在劳动力需求结构与岗位特征方面的差异。

在数据处理过程中，本文以城市为核心分析对象，对岗位层面的信息进行聚合，构建城市层面的结构性指标。通过这一处理方式，可以有效减少个体岗位噪声对分析结果的干扰，同时更好地刻画城市整体的人才吸引力与产业结构特征。

在样本筛选方面，本文剔除了关键信息缺失的岗位记录，并对薪资变量进行了必要的数值清洗与标准化处理，以保证后续模型分析的稳定性与可解释性。

### 2.2 城市层面变量构建思路

考虑到直接在岗位层面引入大量分类变量容易导致模型复杂度过高以及多重共线性问题，本文在变量构建阶段即进行信息压缩与结构化处理，将岗位特征映射为具有明确经济含义的城市层面指标。具体而言，本文围绕以下三个核心维度构建城市特征变量：学历结构、高端产业结构以及技能结构多样性。

### 2.3 学历结构变量

学历水平是衡量城市高端人才集聚程度的重要指标。本文以岗位对博士学历（PhD）的需求作为高端学术与研究型人才的代表性特征，构建城市层面的博士需求占比变量。其定义如下：

$$PhDRatio_c = \frac{\#(PhD)_c}{\#card(job)_c}$$

该指标反映了城市对高层次学术或研发型人才的需求强度，在一定程度上也间接体现了城市的科研环境与高端创新活动水平。

### 2.4 高端产业占比变量

产业结构是影响城市薪资水平与人才吸引力的关键因素之一。相比逐一引入行业虚拟变量，本文采用“高端产业占比”这一综合性指标，以刻画城市在高附加值产业方面的集中程度。

基于数据集中已有的行业分类，本文将以下行业划分为高端或技术密集型产业：信息技术（IT）、工程（Engineering）、设计/IT（Design/IT）、金融（Finance）以及医疗健康（Healthcare）。上述

行业通常具有较高的人力资本要求和薪资水平，能够较好地代表城市的高端产业结构。

据此，定义城市层面的高端产业占比为：

$$High-tech\ Industry\ Ratio_c = \frac{\#(High-tech\ Industry)_c}{\#(job)_c}$$

该变量反映了城市产业结构的技术密集程度，是后续分析城市薪资决定机制与城市类型差异的重要解释变量。

## 2.5 技能结构与技能多样性指标

除学历与产业结构外，岗位技能需求结构同样是衡量城市劳动力市场特征的重要维度。不同于仅关注某一特定技能是否出现，本文更关注城市整体技能需求的多样性水平。

数据集中每个岗位对应一组技能关键词，技能之间以分号或逗号分隔。本文首先对所有岗位的技能文本进行拆分与整理，构建完整的技能集合；随后，在城市层面对技能出现频率进行统计，计算各技能在城市岗位中的相对占比。

为量化技能结构的多样性，本文引入香农信息熵（Shannon Entropy）作为技能多样性指标。对于城市  $c$ ，其技能香农指数定义为：

$$Skill\ Entropy_c = - \sum_{i=1}^N p_{ci} \ln(p_{ci})$$

其中， $p_{ci}$  表示技能  $i$  在城市  $c$  所有技能需求中所占的比例， $N$  为该城市出现的不同技能种类数量。

香农指数越大，说明城市的技能需求分布越均匀、技能结构越多样；反之，则表明城市岗位技能需求相对集中。该指标能够从整体层面反映城市对复合型人才和多技能劳动力的需求程度。

## 2.6 薪资变量处理

作为本文分析的核心结果变量，薪资水平在不同岗位与城市之间存在显著差异。由于薪资总体水平是右偏状态，本文对城市平均薪资取对数处理，构建城市层面的平均对数薪资指标：

$$Avg\ Log\ Salary_c = \ln(Avg\ Salary_c)$$

该处理方式在保留相对差异信息的同时，有助于缓解薪资分布右偏问题，适用于后续回归分析与机器学习建模。

## 2.7 变量标准化处理

在涉及多变量综合分析与聚类分析的模型中，不同变量的量纲与取值范围可能存在较大差异。为避免某一变量因尺度过大而主导模型结果，本文对所有城市层面特征变量进行了标准化处理。具体方法为对变量进行零均值、单位方差的标准化变换。

标准化后的变量主要用于模型二与模型三的分析，而在解释性分析中，本文仍保留原始变量的经济含义进行讨论。

### 3. 模型设定与研究方法

在前一章中，本文构建了刻画城市人才结构与产业特征的核心变量。本章在此基础上，进一步设计三类互补的分析模型，分别用于解释薪资决定机制、刻画高薪岗位特征以及对城市进行综合画像与分类。三类模型在研究目标、建模方法与解释侧重点上各有侧重，共同构成本文的整体分析框架。

#### 3.1 整体建模思路

本文的研究目标并非单一的因果识别，而是通过多角度建模，系统刻画城市人才市场的结构特征及其与薪资水平之间的关系。为此，本文采用由“解释性建模”向“结构刻画与分类建模”逐步推进的策略，依次构建以下三类模型：

- 模型一：基于机器学习回归的薪资决定机制分析，用于解释城市层面薪资差异的主要结构性来源；
- 模型二：高薪岗位的分类模型，用于识别哪些岗位特征更可能对应高薪工作；
- 模型三：城市人才吸引力与结构特征的聚类分析，用于对城市进行画像与分组。

这一建模路径由“解释薪资差异”逐步过渡到“刻画岗位结构”与“城市类型划分”，能够从不同层面回答本文提出的研究问题。

#### 3.2 模型一：薪资决定机制的解释性建模

##### 3.2.1 模型目标

模型一的研究目标是解释不同城市之间平均薪资水平差异的主要来源。与传统线性回归模型不同，本文并不试图估计严格的因果效应，而是关注在控制变量数量较少的情况下，哪些结构性因素在解释城市薪资差异时具有更高的重要性。

为避免多重共线性问题，同时提升模型的稳健性与拟合能力，本文采用基于树模型的机器学习回归方法，对城市平均对数薪资进行建模。（尝试过 ols，共线性非常严重，故更换了研究方法）

##### 3.2.2 模型设定

设城市  $c$  的平均对数薪资为  $y_c$ ，对应的解释变量向量为：

$$\mathbf{X}_c = (PhDRatio_c, \ High-tech \ Industry \ Ratio_c, \ Skill \ Entropy_c)$$

模型一可形式化表示为：

$$y_c = f(\mathbf{X}_c) + \varepsilon_c$$

其中,  $f(\cdot)$  为非线性回归函数, 通过机器学习方法进行估计,  $\varepsilon_c$  表示随机误差项。

### 3.2.3 变量重要性解释

在模型估计完成后, 本文通过特征重要性 (feature importance) 来衡量各解释变量在薪资决定中的相对贡献。变量重要性反映了在模型预测过程中, 各变量对降低预测误差的贡献程度, 从而为城市薪资差异提供结构性解释。

需要强调的是, 该重要性指标用于描述解释能力而非因果强度, 其结论应结合经济直觉进行解读。

## 3.3 模型二: 高薪岗位的分类建模

### 3.3.1 模型目标

模型二从岗位层面出发, 研究哪些岗位特征更可能对应高薪工作。与模型一关注城市整体薪资水平不同, 模型二旨在刻画岗位之间的异质性, 通过分类模型识别高薪岗位的典型特征组合。

该模型的核心研究问题为: 在控制岗位经验要求、学历要求与城市层级等因素后, 是否具备人工智能相关技能的岗位更有可能成为高薪岗位。

### 3.3.2 高薪岗位定义

本文以岗位薪资分布的分位数作为高薪岗位的判定标准。具体而言, 当岗位的中位薪资高于样本薪资分布的第 75 百分位数时, 将该岗位定义为高薪岗位, 并赋值为 1; 否则赋值为 0。记岗位  $i$  的高薪标签为:

$$y_i = \begin{cases} 1, & \text{若岗位 } i \text{ 的薪资大于 75 分位数} \\ 0, & \text{否则} \end{cases}$$

该定义能够在保证样本规模的同时, 突出薪资分布上部岗位的结构特征。

### 3.3.3 解释变量设定

模型二的解释变量主要包括岗位技能特征、岗位要求特征以及城市层级特征, 具体包括:

- **AI\_Skills**: 核心解释变量, 表示岗位是否要求人工智能相关技能。当岗位技能描述中同时包含 “Python”、“Machine Learning” 与 “Statistics” 时, 该变量取值为 1, 否则取值为 0。
- **Experience\_Level**: 岗位经验要求, 为分类变量, 刻画岗位对工作年限与资历的要求。
- **Education\_Requirement**: 岗位学历要求, 为分类变量, 反映人力资本门槛差异。
- **City\_Tier**: 城市层级变量, 用于刻画城市发展水平与劳动力市场环境差异。

其中, **City\_Tier** 并非数据集中直接给出的变量, 而是基于城市的经济发展水平与市场成熟度进行人为划分。本文将一线及核心经济城市划分为 T1 城市, 其余城市不做区分, 因为数据集中城

市只设计一二线城市。该划分方式在城市经济与劳动力市场研究中具有较强的现实解释意义，有助于控制城市层面不可观测的制度与市场环境差异。

所有分类变量在建模过程中均采用哑变量 (dummy variables) 形式进入模型。

### 3.3.4 模型方法选择

鉴于模型二属于二分类问题，本文采用逻辑回归 (Logistic Regression) 与基于树的集成学习方法 (XGBoost) 进行建模。

逻辑回归作为经典的广义线性模型，其参数具有明确的经济含义，能够直接刻画各解释变量对岗位成为高薪岗位概率的边际影响，因此适合作为基准模型，用于提供可解释性分析。

然而，逻辑回归对变量之间的线性可分性与函数形式假设较为严格。为进一步刻画变量之间潜在的非线性关系与交互效应，本文引入 XGBoost 分类模型。XGBoost 通过梯度提升树的方式，在预测性能与处理复杂结构方面具有显著优势，能够作为对逻辑回归结果的重要补充。

### 3.3.5 模型评估指标

在模型评估方面，本文综合采用准确率 (Accuracy)、ROC-AUC 以及 F1-score 作为衡量标准。其中，F1-score 在类别分布不平衡的情况下尤为重要，有助于评估模型对高薪岗位的识别能力。

## 3.4 模型三：城市人才吸引力与结构特征的聚类分析

### 3.4.1 模型目标

模型三的目标是基于多维结构性指标，对城市进行综合画像与分类。与前两个模型关注预测或解释不同，模型三侧重于无监督学习，通过聚类方法识别在人才结构与产业特征方面具有相似性的城市类型。

该模型旨在回答以下问题：不同城市在人才结构与技能需求方面是否呈现出可区分的类型特征？

### 3.4.2 特征向量构建

在模型三中，每个城市  $c$  由以下标准化后的特征向量表示：

$$\mathbf{w}_c = (\text{Avg Log Salary}_c, \text{PhD Ratio}_c, \text{High-tech Industry Ratio}_c, \text{Skill Entropy}_c)$$

所有特征在进入聚类模型前均进行标准化处理，以避免尺度差异对聚类结果产生影响。

### 3.4.3 聚类方法

本文采用 K-means 聚类方法对城市进行分类。该方法通过最小化类内平方误差，将城市划分为  $K$  个互不重叠的类别。其优化目标为：

$$\min \sum_{k=1}^K \sum_{c \in C_k} \|\mathbf{w}_c - \boldsymbol{\mu}_k\|^2$$

其中,  $C_k$  表示第  $k$  类城市集合,  $\mu_k$  为该类的中心向量。

为便于结果展示与解释, 本文进一步采用主成分分析 (PCA) 将高维城市特征降至二维空间, 并对聚类结果进行可视化分析。

### 3.4.4 城市画像解释

在完成聚类后, 本文计算各城市类别在关键特征上的均值水平, 形成城市画像。通过比较不同类别在薪资水平、学历结构、产业结构与技能多样性方面的差异, 可以为理解城市人才吸引力与发展模式提供直观依据。

## 4. 实证结果与分析

### 4.1 模型一结果分析: 城市薪资决定机制

表 1: 模型一: 城市平均薪资决定机制的特征重要性

变量名称	特征重要性	相对贡献比例
高端产业占比 (High-tech Industry Ratio)	0.461	46.1%
技能多样性指数 (Skill Entropy)	0.405	40.5%
博士需求占比 (PhD Ratio)	0.134	13.4%

模型一采用基于树的机器学习回归方法, 对城市层面的平均对数薪资进行解释性建模。模型的拟合结果显示, 其决定系数  $R^2$  为 1.000, 均方根误差 (RMSE) 为 0.000, 表明模型在样本内能够完全解释城市平均薪资的变异。

从特征重要性结果来看, 高端产业占比在薪资决定机制中发挥最为关键的作用, 其重要性权重达到 46.1%。该结果表明, 城市中高端、技术密集型产业的集中程度是决定整体薪资水平的首要因素。这一发现与经济学中“产业结构升级推动薪资提升”的理论预期高度一致。

其次, 技能多样性指数的重要性达到 40.5%, 显示城市对多技能、复合型人才的需求结构同样对薪资水平具有显著影响。技能结构越多样化的城市, 通常意味着岗位类型更加丰富, 对高技能劳动力的竞争更为激烈, 从而推高整体薪资水平。

相比之下, 博士需求占比的相对重要性为 13.4%, 虽然仍具有正向影响, 但其贡献程度明显低于产业结构与技能结构因素。这一结果说明, 单纯依赖高学历人才集聚并不足以决定城市整体薪资水平, 其作用需要通过产业环境与技能需求结构的匹配才能充分体现。

需要指出的是, 模型一的拟合优度极高, 反映出城市平均薪资在本研究框架下可被少数结构性变量高度解释。这一结果在一定程度上源于变量构建阶段对信息的高度聚合, 使模型更偏向于结构解释而非预测外推。因此, 模型一的主要价值在于揭示城市薪资差异的核心结构性因素, 而非用于样本外预测。

## 4.2 模型二结果分析：高薪岗位的分类特征

模型二采用逻辑回归与 XGBoost 两种分类模型，对岗位是否属于高薪岗位进行预测。表 4.2 汇报了两种模型在测试集上的预测性能。从整体表现来看，两种模型在 Accuracy、ROC-AUC 与 F1-score 指标上均表现良好，且结果高度一致，表明模型设定具有较强的稳健性。

表 2：模型二：高薪岗位分类模型的预测性能比较

模型	Accuracy	ROC-AUC	F1-score
Logistic Regression	0.829	0.861	0.707
XGBoost	0.827	0.862	0.702

模型二采用逻辑回归与 XGBoost 两种分类模型，对岗位是否属于高薪岗位进行预测。表 4.2 汇报了两种模型在测试集上的预测性能。从整体表现来看，两种模型在 Accuracy、ROC-AUC 与 F1-score 指标上均表现良好，且结果高度一致，表明模型设定具有较强的稳健性。

具体而言，逻辑回归模型的 ROC-AUC 为 0.861，XGBoost 模型的 ROC-AUC 为 0.862，说明两种模型在区分高薪岗位与非高薪岗位方面均具备较强的判别能力。与此同时，F1-score 均在 0.70 左右，表明模型在高薪岗位这一相对少数类别上的识别效果较为理想。

为进一步分析各解释变量对模型预测结果的影响方向与相对重要性，本文基于 XGBoost 模型引入 SHAP (SHapley Additive exPlanations) 方法进行可解释性分析。图 1 展示了各变量的 SHAP 值分布情况。

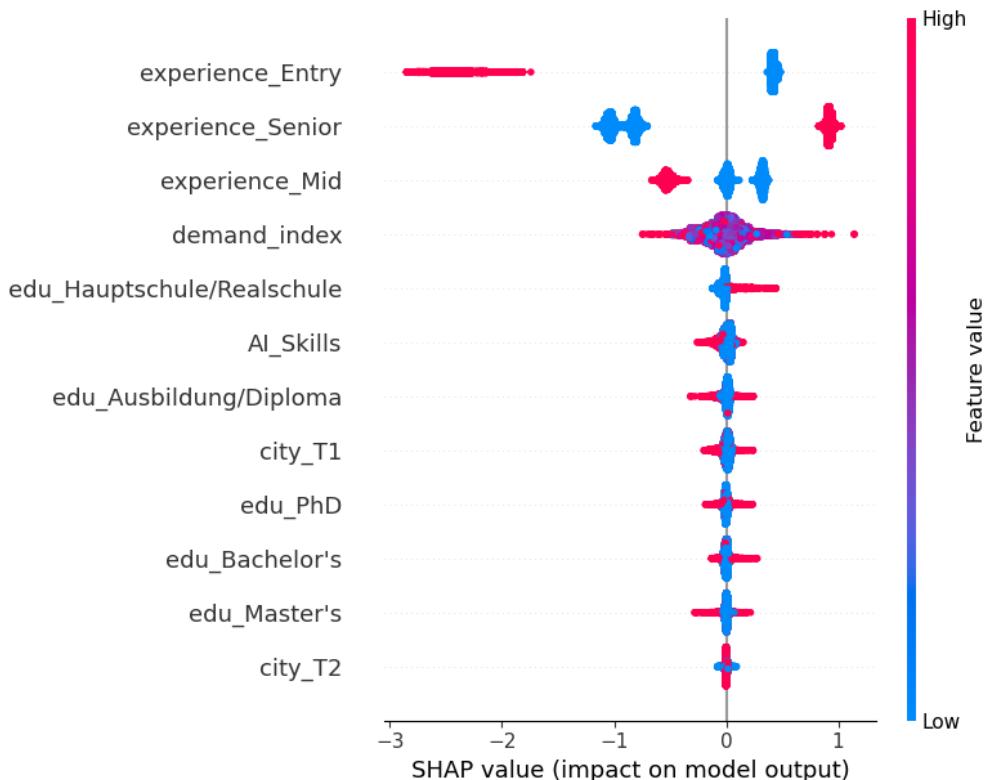


图 1：XGBoost 模型的 SHAP 特征重要性与影响方向

从 SHAP 结果可以看出，岗位经验要求是影响高薪岗位判定的最关键因素。其中，Senior 级岗位对成为高薪岗位具有显著的正向贡献，而 Entry 级岗位则明显降低岗位进入高薪区间的概率。这一结果与劳动力市场中经验溢价的普遍认知高度一致。

需求指数（Demand Index）同样表现出较强的正向影响，说明市场需求旺盛的岗位更容易形成薪资溢价。这反映了劳动力供需关系在薪资形成机制中的重要作用。

人工智能技能变量（AI\_Skills）在模型中呈现出稳定的正向影响。具备 Python、机器学习与统计等复合技能要求的岗位，其 SHAP 值整体偏正，表明人工智能相关技能能够显著提升岗位成为高薪岗位的概率。这一发现从岗位层面验证了人工智能技能的薪资溢价效应。

在城市层级变量方面，一线城市（T1）岗位对高薪岗位判定具有正向贡献，而二线城市（T2）的影响相对有限，说明城市发展层级仍然在薪资形成中发挥重要的结构性作用。

相比之下，不同学历要求变量的 SHAP 值整体分布较为集中，其影响程度弱于经验、技能与需求因素。这一结果表明，在岗位层面的薪资分化中，学历更多体现为进入特定岗位的门槛条件，而非决定薪资高低的核心因素。

### 4.3 模型三：城市高端就业结构的无监督聚类分析

#### 4.3.1 聚类变量选取与数据处理

模型三选取以下四个城市层面的核心变量作为聚类特征，分别从薪资水平、人才结构、产业结构与技能多样性四个维度刻画城市的高端就业特征：

- (1) 平均对数薪资 (avg\_log\_salary)，用于反映城市整体高端岗位的薪资水平；
- (2) 博士学历占比 (phd\_ratio)，用于衡量城市对高层次学术与科研型人才的需求强度；
- (3) 高端产业岗位占比 (hi\_industry\_ratio)，用于刻画城市产业结构中技术密集型行业的比重；
- (4) 技能香农熵 (skill\_entropy)，用于衡量城市岗位技能需求结构的多样化程度。

为消除量纲差异对聚类结果的影响，本文对上述变量进行标准化处理，并在此基础上实施 K-Means 聚类分析。结合聚类稳定性与经济解释性，最终将样本城市划分为四类。

#### 4.3.2 不同城市类型的特征均值比较

表 3：不同城市聚类的高端就业特征均值

城市类别	平均对数薪资	博士占比	高端产业占比	技能熵
Cluster 0	1.377	-0.296	-1.959	1.542
Cluster 1	-0.029	-0.636	-0.413	-0.517
Cluster 2	0.398	1.022	0.869	-0.199
Cluster 3	-1.584	-0.182	0.253	1.538

表 3 汇报了四类城市在核心特征变量上的均值情况（均为标准化值）。结果显示，不同城市类别在薪资水平、学历结构、产业结构与技能需求多样性方面均存在显著差异，表明城市高端就业结

构具有明显的分层特征。

#### 4.3.3 不同城市类型的结构特征解释

第一类城市 (Cluster 0) 在平均薪资水平与技能香农熵指标上显著高于其他类别，表明该类城市的高端岗位不仅薪资回报较高，同时对多样化技能组合具有较强需求。然而，该类城市在高端产业占比与博士学历需求方面相对较低，显示其高端就业优势更多来源于复合型技能与高附加值服务业。

第二类城市 (Cluster 1) 在薪资水平、博士学历占比、高端产业占比及技能熵等指标上均处于相对较低水平，表现出整体高端就业结构较为薄弱的特征。这类城市可能更多依赖传统产业或低附加值服务业，在高端岗位数量与质量方面均面临一定约束。

第三类城市 (Cluster 2) 在博士学历占比与高端产业岗位占比上显著高于其他类别，同时薪资水平亦处于中上区间，表明该类城市的高端就业结构高度依赖技术密集型产业，对科研能力与专业技术深度具有较强需求。

第四类城市 (Cluster 3) 在技能熵指标上表现出较高水平，但平均薪资水平显著偏低，高端产业占比与博士学历需求亦不突出。这一结果说明，技能需求的多样化本身并不必然转化为高薪回报，若缺乏高端产业支撑，技能结构的复杂性难以形成实质性的高端就业优势。

#### 4.3.4 城市类型的可视化分析

为进一步直观展示不同城市类型在高端就业结构上的差异，本文基于聚类特征，采用主成分分析 (PCA) 方法对城市进行二维降维，并绘制城市类型分布图，如图 2 所示。

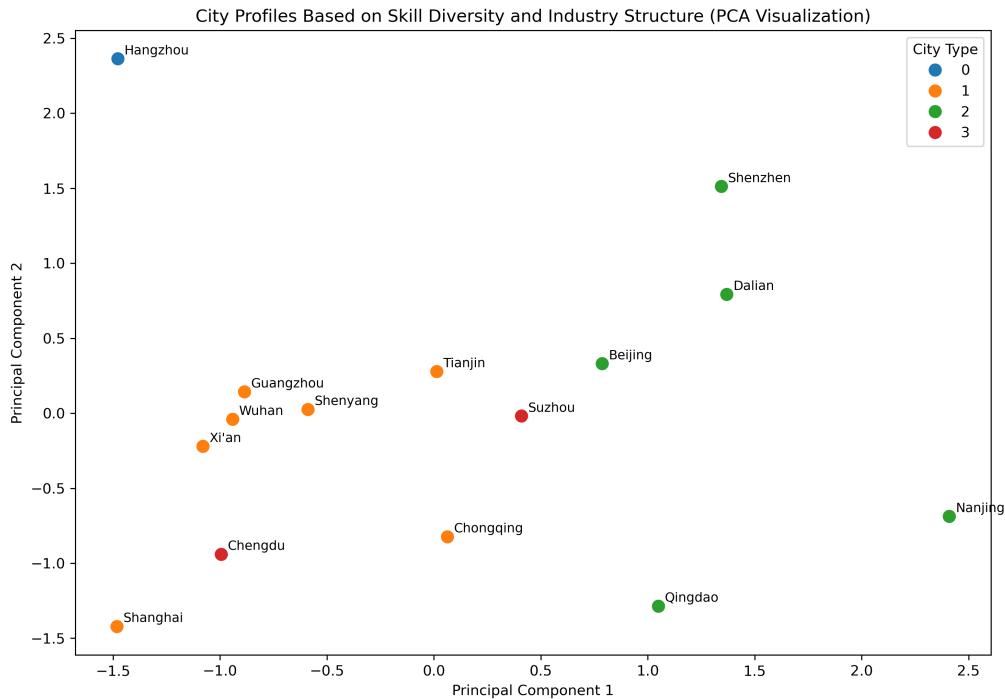


图 2：基于技能多样性与产业结构的城市类型 PCA 可视化结果

从图 2 可以看出，不同类别城市在主成分空间中呈现出较为清晰的分布结构，验证了聚类结果

在整体结构上的合理性。

#### 4.4 三种模型结果的综合对照与一致性分析

本文分别通过回归建模、分类建模与聚类分析三种不同建模范式，对高端就业结构的形成机制进行了多角度刻画。尽管三种模型在研究目标与方法上存在差异，但其核心结论在整体上具有较强的一致性。

首先，模型一基于回归框架，从连续变量角度刻画了城市层面高端就业特征对薪资水平的解释能力。结果显示，高端产业占比、技能多样性与高学历人才集聚对薪资水平具有显著影响，表明高端就业并非单一要素驱动，而是多维结构共同作用的结果。

其次，模型二通过分类建模，进一步验证了上述结论在岗位层面上的稳健性。无论是 Logistic 回归还是 XGBoost 模型，高端技能需求、工作经验与城市层级变量均在高薪岗位判别中发挥关键作用，说明产业结构与技能结构在高薪岗位形成中具有重要解释力。

最后，模型三通过无监督聚类方法，从结构分布角度揭示了城市之间在高端就业模式上的系统性差异。聚类结果表明，不同城市在高薪、高学历、高端产业与技能多样性之间呈现出多样化组合关系，与模型一、模型二所揭示的多维决定机制高度一致。

综上所述，三种模型从不同角度共同验证了高端就业结构的多维性与异质性特征，增强了本文研究结论的整体可信度，也为后续政策讨论与扩展研究奠定了坚实基础。

### 5. 结论与政策启示

#### 5.1 主要研究结论

本文基于招聘岗位微观数据，从岗位层面与城市层面系统分析了我国高端就业结构的形成机制。通过构建回归模型、分类模型与聚类模型，本文从不同角度刻画了高端岗位薪资、高薪岗位判别以及城市高端就业结构的异质性特征，主要结论如下。

首先，从连续薪资决定机制的角度看，高端就业并非由单一因素驱动，而是由产业结构、人才结构与技能结构的多维因素共同决定。模型一的结果表明，高端产业岗位占比、技能需求多样性以及高学历人才集聚程度对城市薪资水平具有显著解释力，其中产业结构与技能结构的影响尤为突出。这一结论表明，高端就业优势的形成依赖于产业生态与技能结构的系统性协同，而非单纯依靠学历或人口规模。

其次，从岗位层面的高薪判别结果来看，高薪岗位在技能要求、经验门槛与城市区位等方面呈现出显著特征。模型二的分类结果显示，具备高端技能组合（如人工智能与数据分析相关技能）、中高等级工作经验以及位于高能级城市的岗位，更有可能被判定为高薪岗位。同时，机器学习模型在预测性能上与传统逻辑回归结果高度一致，增强了结论的稳健性。

最后，从城市整体结构的角度看，不同城市在高端就业结构上存在显著异质性。模型三的聚类分析表明，城市可以根据薪资水平、产业结构、人才结构与技能多样性被划分为若干具有不同发展模式的类型。一部分城市依托高端产业与科研能力形成技术密集型高端就业结构，另一部分城市则更多依赖技能多样化与服务业发展，而部分城市在上述维度上均相对薄弱，呈现出明显的结构性差距。

## 5.2 政策启示

基于本文的实证结果，可以从以下几个方面为城市高端就业发展提供政策启示。

第一，推动高端产业集聚是提升城市高端就业质量的关键路径。研究结果表明，高端产业占比在薪资决定与城市类型划分中均发挥核心作用。因此，地方政府应通过产业政策、科技创新支持与产业链完善，引导技术密集型产业在具备基础条件的城市形成集聚效应，从而提升高端岗位的数量与质量。

第二，重视技能结构而非单一学历指标。模型结果显示，技能多样性对高端就业具有重要影响，而学历本身并非唯一决定因素。这意味着，在人才政策设计中，应更加重视技能培养、跨领域能力与复合型人才的发展，通过职业培训、终身教育体系建设等方式，提升劳动者的技能适配能力。

第三，因城施策，避免“一刀切”的人才政策。模型三揭示了城市在高端就业结构上的显著异质性，不同城市适合的发展路径并不相同。对于技术基础雄厚的城市，应重点强化科研与高端制造能力；对于服务业与技能多样性较强的城市，则可通过提升产业附加值与岗位质量，实现高端就业的结构升级。

## 5.3 研究不足与未来展望

尽管本文在数据与方法上力求全面，但仍存在一定局限性。首先，受数据可得性限制，本文未能引入更为细致的企业层面信息，如企业规模、所有制性质等因素，未来研究可在此基础上进一步拓展分析维度。

其次，本文主要基于静态截面数据进行分析，尚未刻画高端就业结构的动态演化过程。未来研究可结合时间序列或面板数据，系统分析高端就业结构的演变趋势与路径依赖特征。

最后，本文对高端产业与技能的界定仍依赖于经验规则与文本提取方法，随着数据与技术条件的改善，后续研究可尝试引入更精细的自然语言处理与网络分析方法，以进一步提升测度精度。

总体而言，本文从多模型视角系统分析了高端就业结构的形成机制，为理解我国区域高端就业差异提供了有益参考，也为相关政策制定提供了实证依据。

## 参考文献

- [1] Claude E Shannon. A mathematical theory of communication. Bell System Technical Journal, 27(3):379–423, 1948.