大数据与人工智能时代下我国农户多维相对贫困的动态演变及影响因素研究

李倩倩

（湖南师范大学商学院，湖南长沙，410000）

摘要：为解决相对贫困的深层次问题，探讨我国农户多维相对贫困的动态演变及其影响因素，以服务精准扶贫目标。利用 Python 爬虫和文本挖掘梳理研究热点，结合中国家庭追踪调查（CHARLS）数据，采用 A-F 多维贫困测度法测算多维贫困指数，并通过随机森林和 Logistic 回归模型分析关键影响因素。结果显示，我国农户多维相对贫困程度较高，教育和收入维度贡献显著，家庭医疗支出负担是主要致贫因素，因病致贫和因灾致贫尤为突出。政策建议包括：加强教育与技能培训，完善医疗保障体系，强化防灾减灾能力，降低多维贫困风险。

关键词：多维相对贫困；A-F 多维贫困测度方法；随机森林；Logistic 回归模型；因病因灾返贫

中图分类号： F8　　**文献标识码：**A

一、引言

脱贫攻坚一直是党中央治国理政的重要议题，也是全面建成小康社会的重要基础。经过全党全国各族人民的共同努力，我国于 2020 年全面消除了绝对贫困，取得了历史性胜利，标志着减贫事业迈入新的阶段。然而，绝对贫困的消除并不意味着减贫任务的结束，相对贫困问题正逐渐凸显。新时代的减贫工作已从聚焦绝对贫困转向相对贫困，这既是对全面脱贫成果的巩固，也是实现共同富裕的重要步骤。相对贫困的实质是收入差距问题，其主要表现为个体收入与社会平均收入的显著差距，尤其在城乡发展不平衡的背景下，相对贫困问题在农村地区更为突出，且呈现出散点分布高、人口流动性强、贫困人群特征复杂（如老弱病残占比高）等新特点。这使得相对贫困的识别更加复杂、治理更加困难，需要开辟新的路径和方法。

传统的贫困研究方法多依赖于人口普查和问卷访谈等社会经济统计数据，虽然为贫困测量提供了基础支持，但由于时效性差、覆盖范围有限、空间分辨率低以及获取成本高等问题，难以支撑对复杂相对贫困的成因分析与动态变化的研究。在人工智能和大数据技术快速发展的背景下，利用智能化技术对海量数据进行分析，已成为解决相对贫困问题的新方向。人工智能技术具备自我学习和优化算法的能力，通过机器学习、深度学习和数据挖掘等方法，不仅能够精准识别相对贫困，还能动态跟踪其演变过程，有效避免传统方法中的主观偏差，提高研究效率和科学性。国务院于 2017 年颁布的《新一代人工智能发展规划》，为推动人工智

能在社会经济领域的应用提供了政策支持。

基于此背景，本研究旨在利用人工智能技术和大数据分析方法，建立科学的多维相对贫困研究体系，以提高贫困识别的准确性和政策干预的针对性。通过探讨农户多维相对贫困的动态演变及其影响因素，本研究为精准扶贫提供科学依据，助力我国减贫事业向着共同富裕的目标迈进。

二、网络数据爬取与文本挖掘分析

（一） 数据采集

为了能够更加了解我国多维相对贫困的研究现状，本文对知网上相关信息进行文本挖掘，利用 python 爬虫技术，得到知网上关于我国多维相对贫困标题、摘要、关键词三个部分数据。对得到的文本数据进行词频统计，然后绘制相应的词云图，展示出学者对我国多维相对贫困研究的“热点词”。

以“多维相对贫困”为关键词，在知网上进行搜索。对该搜索页面上的所有文献的详情进行数据爬取，重点关注在文献的标题、摘要、关键词三个部分。通过这三部分，分析得出学者研究的关注点，为后续指标的构建和方法的采用提供参考。采集到知网上上共 4102 篇文献的信息，其中关键词共 16408 条。其中对收集到的数据结构化，以表格形式储存，为后续处理与分析做铺垫。具体详见附件。

（二）数据清洗与处理

在数据采集过程中采集到的原始数据，往往包含大量不完整、不一致或者异常的数据等情况，为了提高数据分析效率和数据可用性，本研究对采集到的原始数据进行了清洗，针对原始数据中、空白数据和重复数据予以删除，使得数据更符合分析要求。在本研究收集到的评论数据中，含有大量长句。为了得到长句评论中的信息，本研究对原始评论数据进行了分词处理。同时，本研究将分词后中部分无效词语删除。

表 1 分词结果示例

分词前文本	分词后文本
2020 年底,现行标准下我国农村贫困人口全部脱贫,贫困县全部摘帽,贫困村全部出	我国 农村 贫困人口 全部脱贫 贫困县 摘帽 贫困村 区域性 整体 消除 绝对

列,区域性整体贫困得到解决,完成了消除绝对贫困的艰巨任务。绝对贫困的消除并不意味着贫困的全部消灭,相对贫困将会凸显并长期存在。这意味...	贫困 艰巨任务 贫困 消灭 相对贫困 凸显 长期 存在
“相对贫困”是后扶贫时代我国反贫困事业的重要问题。虽然2020年在现行标准下的9899万农村贫困人口全面脱贫,但是由于发展历史、地域环境和资源禀赋等原因,黄河中下游五省(区)相对贫困...	相对贫困 后扶贫时代 我国 反贫困 事业 重要 问题 现行标准 全面 脱贫 发展 历史 地域 环境 资源 禀赋 等 原因 黄河 中下游 五 省(区)

（三）词频分析

根据分词后的结果，本研究进行了词频统计，绘制了相应词云图。结合词云图可以发现，在知网文献的标题中，多维相对贫困是图中最大的词，表明它是研究中的核心主题，研究主要聚焦于探索和理解多维相对贫困的概念和特性。影响因素这个词也相对较大，说明文献常常探讨导致多维相对贫困的各种因素，如经济、社会、文化等因素如何共同作用导致贫困状态。扶贫、政策、测量这些词的出现突出了研究的实用方向，即如何通过政策干预和准确测量来缓解或消除多维相对贫困。教育、健康、农村这些词反映了研究中关注的特定领域或群体。教育和健康是影响贫困的重要维度，而农村则是多维贫困研究的重要地理焦点。持续发展、社会保障这些概念的出现表明研究中不仅关注短期扶贫措施，也关注长期的可持续发展和社会保障系统的构建。贫困线、城乡差异这些词显示了研究的具体焦点，比如贫困线的设定、城乡之间的贫困差异等，这些都是制定有效扶贫政策的关键考虑因素。整体而言，这张词云图提供了对多维相对贫困研究领域的一个直观总览，显示了该领域研究的热点词汇和主题，有助于了解当前研究的重点和研究趋势。

图 2 标题词云图

图 1 标题词云图

在摘要词云图中，可以看到图中最显著的词是农村，说明研究主要集中在农村相关的议题上。农民、农业、农村经济这些词频繁出现，表明研究聚焦于农民的生活状态、农业的发展以及农村经济的整体状况。发展、改革、政策这些词的突出显示，指出了政策和改革措施在推动农村发展中的重要性。扶贫、收入、就业这些与经济相关的词汇突出了农村地区扶贫工作的重点，如提高农民收入和增加就业机会。教育和健康服务的提及说明这些也是研究的重要方面，关注农村居民的生活质量和福利。资源、土地、可持续这些词指向农村资源管理和土地使用的可持续性问题，反映了环境保护和可持续发展的研究趋势。技术、信息、网络表明科技在农村发展中起着越来越重要的角色，特别是信息技术和网络的普及。市场、产业、合作社这些词反映了农村市场经济的发展动向，以及农村合作社和产业组织的研究关注。

图 2 摘要词云图

关键词词云图中多维贫困指数（MPI）显著放在中心位置，表明它是研究中的核心主题。这一指标用于评估和量化贫困的多个维度，是现代贫困研究中的重要工具。Alkire 这个词指的是 Sabina Alkire，她是多维贫困指数的共同开发者之一，表明她的工作对这一领域的研究具有重大影响。还有一些方法论关键词如 DEA（数据包络分析），是一种评价生产效率的方法，可能用于评估政策或项目在不同维度上的效率，特别是在多维贫困背景下；PSM（倾向得分匹配），是一种统计方法，用于处理观察性研究中的因果推断问题，说明研究者可能利用它来评估某些干预对多维贫困的影响；DID（双重差分法），这是一种用于评估政策干预效果的经济学方法，说明其在多维贫困研究中的应用。对于应用领域的关键词则有健康、教育、住房，这些都是构成多维贫困指数的关键维度，反映了贫困研究关注的领域；CiteSpace，这是一种科学文献可视化工具，表明研究者可能使用该工具来分析和可视化多维贫困研究的文献网络。而在统计与评估工具方面的主要关键词则为随机森林、Logistic 回归，这两个是数据分析中常用的统计方法，用于分析和预测数据模型，特别是在处理大量数据和复杂模型时。这张词云图提供了关于多维贫困研究的广泛视角，包括研究方法、关键人物、常用的统计工具以及研究的主要领域。通过这些关键词，我们可以看到多维贫困研究是一个多学科、方法多样化的领域，涉及从理论到实践的多个方面。

图 3 关键词词云图

三、研究方法与数据概述

（一）研究方法概述

1. A-F 多维贫困测度方法

A-F 多维贫困测度方法，即 Alkire-Foster 方法，是由 Sabina Alkire 和 James Foster 提出的。这种方法不仅考虑了收入贫困，还综合考虑了健康、教育、生活水平等多个维度的剥夺情况，是一种更全面反映贫困状态的测度方法。A-F 方法被广泛应用于国际和国家层面的贫困分析，如联合国开发计划署（UNDP）的多维贫困指数（MPI）。

A-F 多维贫困测度方法主要包括三个核心概念：剥夺（Deprivation）、多维贫困发生率（H）和多维贫困指数（M）。剥夺（Deprivation）：指在某个特定维度上的贫困情况。每个维度都有特定的剥夺标准，当一个家庭或个体在某个维度上未达到标准时，就被认为在该维度上受到了剥夺。多维贫困发生率（Headcount Ratio, H）：指在所有调查对象中，被识别为多维贫困的家庭或个体所占的比例。这个指标反映了多维贫困的广度。多维贫困指数（Multidimensional Poverty Index, M）：结合了贫困发生率（H）和平均剥夺份额（A），综合反映了多维贫困的深度和广度。计算公式为：

$$H(k) = \frac{q(k)}{n} \tag{1}$$

$$A(k) = \sum \frac{c_i(k)}{q(k)} \tag{2}$$

$$A(k) = \sum \frac{c_i(k)}{q(k)} \tag{3}$$

其中 A（k）表示多维贫困者在被剥夺的各个维度上的平均剥夺份额。

2. 核密度估计

①核密度估计（Kernel Density Estimation, KDE）

核密度估计是一种非参数统计方法，用于估计随机变量的概率密度函数。与直方图不同，核密度估计能提供更加平滑和连续的密度曲线，常用于数据分析中的分布形状探索和模式识别。

②基本概念

核密度估计的核心思想是利用核函数（kernel function）对数据点进行平滑，从而估计出整体数据的密度分布。核密度估计的公式如下：

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{4}$$

$\hat{f}(x)$是在点 x 处的密度估计值，n 是样本大小，h 是带宽（bandwidth），决定平滑程度，k 是核函数，常见的核函数有高斯核、均匀核、三角核等。x_i 是第 i 个样本点。

3. 随机森林

在机器学习中，随机森林是一个基于 Bagging 算法同时利用多个决策树的分类模型。该模型选取了大量的决策树模型，各个决策树独立的做出学习并进行分类，最后将分类结合为一个最终的结果，其效果优于单个决策树所做出的分类结果。所以，随机森林作为一种有监督的分类模型，其所具有的准确性和灵活性使得其被广泛的应用。其中，随机森林模型的示意图如下：

（二）数据来源与变量说明

1. 数据来源

本研究使用的数据来自 CHARLS 数据库，涵盖了 2015 年、2018 年和 2020 年的调查数

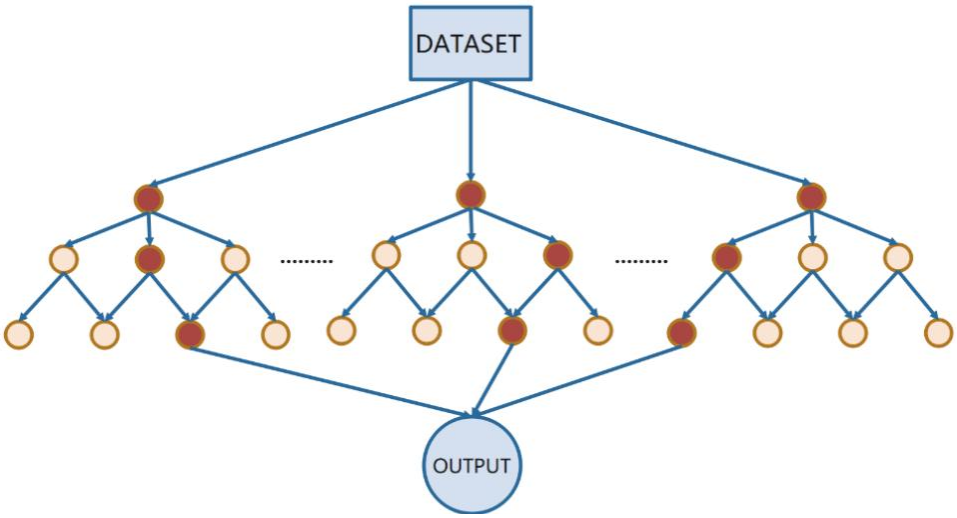

图 4 随机森林模型示意图

据。CHARLS 是由北京大学国家发展研究院组织实施的一项全国性、代表性、以家庭为单位的纵向调查，旨在收集一套代表中国 45 岁及以上中老年人家庭和个人的高质量微观数据，用以分析我国人口老龄化问题，推动老龄化问题的跨学科研究。CHARLS 全国基线调查于 2011 年开展，覆盖 150 个县级单位，450 个村级单位，约 1 万户家庭中的 1.7 万人。这些样本以后每两到三年追踪一次，调查结束一年后，数据将对学术界展开。通过对中老年人及其家庭的经济、社会和健康状况进行详细跟踪，提供关于中国老龄化及其社会经济影响的高质量数据。

2. 变量选取

本文旨在测算我国农户的多维相对贫困动态演变及其影响因素，参考已有研究，构建如下多维相对贫困指标体系。

表 2 多维相对贫困指标体系

维度	指标	指标说明	权重
经济维度	家庭总收入	总收入低于样本家庭中位数的 50%=1，否则=0	1/5
健康维度	健康状况	存在至少 1 名家庭成员自评为“不健康”=1，否则=0	1/15
	医疗保障	家庭成员中无人参加任何形式的医疗保险=1，否则=0	1/15
	医疗支出	实际医疗支出超过样本家庭的中位数=1，否则=0	1/15
教育维度	教育程度	受教育年限低于样本家庭中位数=1，否则=0	1/10
	入学青少年	当前家庭有未就读大学（高职、大专）的适龄青少年或其受教育年限小于等于 12 年=1，否则=0	1/10
就业维度	工作保障	家庭中至少有 1 名成年成员的工作未签订劳动合同=1，否则=0	1/10
	工作满意度	家庭中至少有 1 名成年成员对现有工作表示不满意=1，否则=0	1/10
生活水平维度	烹饪燃料	家庭使用的燃料为非清洁能源，如木材、农作物残留物=1，否则=0	1/20
	自有住房	没有自有住房=1，否则=0	1/20
	饮用水水源	饮用水非自来水=1，否则=0	1/20
	通电情况	家庭未通电=1，否则=0	1/20

四、我国多维相对贫困的测度

（一）2015 年全国及城乡多维相对贫困识别结果

表中展示了 2015 年全国及城乡样本在不同剥夺数量下的多维相对贫困识别结果。总体

来看，我国不存在 12 个指标同时被剥夺的情况，被剥夺的数量在 2015 年的最大值为 11。无论是多维相对贫困发生率（H）、平均多维剥夺份额（A）还是多维相对贫困指数（M），我国农村地区的相对贫困均显著高于城镇地区。然而，随着剥夺数量的增加，城乡之间的差距逐渐缩小。多维相对贫困发生率（H）和多维相对贫困指数（M）：随着剥夺数量的增加，H 和 M 呈现出下降趋势，尤其当剥夺数量达到 10 个及以上时，H 和 M 已经非常小，趋于稳定。这在一定程度上反映了我国多维相对贫困的顽固性，即总有极少数人处于极度贫困中。A 随着剥夺数量的增加逐渐上升，尤其当剥夺数量超过 10 时，A 接近于 100%。

具体来说当剥夺数量为 1 时，全国范围内有 99.4%的家庭至少被剥夺了 1 个指标，城镇和农村分别为 98.9%和 99.7%。此时的平均多维剥夺份额（A）全国为 36.5%，城镇为 29.1%，农村为 40.9%，多维相对贫困指数（M）全国为 36.3%，城镇为 28.7%，农村为 40.7%。当剥夺数量增加到 6 时，全国的多维相对贫困发生率（H）降至 23.8%，城镇和农村分别为 11.8%和 30.9%。此时的平均多维剥夺份额（A）全国为 60.3%，城镇为 53.0%，农村为 62.4%，多维相对贫困指数（M）全国为 14.7%，城镇为 7.1%，农村为 19.3%。当剥夺数量达到 10 时，全国的多维相对贫困发生率（H）降至 0.7%，城镇和农村分别为 0.1%和 1.3%。此时的平均多维剥夺份额（A）全国为 88.0%，城镇为 88.7%，农村为 87.6%，多维相对贫困指数（M）全国为 0.6%，城镇为 0.1%，农村为 1.1%。在所有剥夺数量范围内，农村地区的多维相对贫困发生率（H）和多维相对贫困指数（M）均显著高于城镇地区。特别是在剥夺数量为 8 时，农村地区的 H 和 M 分别是城镇的 1.83 倍和 1.82 倍。通过以上分析，我们可以看到，尽管我国在 2015 年的多维相对贫困问题依然存在，但随着剥夺指标数量的增加，贫困程度逐渐减轻。

表 3 2015 年全国及城乡多维相对贫困识别结果

剥夺数量	多维相对贫困发生率 H			平均多维剥夺份额 A			多维相对贫困指数 M		
	全国	城镇	农村	全国	城镇	农村	全国	城镇	农村
1/12	0.994	0.989	0.997	0.365	0.291	0.409	0.363	0.287	0.407
2/12	0.809	0.669	0.997	0.418	0.366	0.409	0.338	0.245	0.407
3/12	0.705	0.538	0.805	0.448	0.403	0.466	0.316	0.217	0.375
4/12	0.545	0.353	0.659	0.497	0.465	0.507	0.271	0.164	0.334
5/12	0.332	0.175	0.425	0.575	0.556	0.58	0.191	0.098	0.246
6/12	0.238	0.118	0.309	0.62	0.603	0.624	0.147	0.071	0.193
7/12	0.14	0.066	0.183	0.682	0.661	0.687	0.095	0.043	0.126

8/12	0.06	0.02	0.083	0.76	0.741	0.763	0.045	0.015	0.063
9/12	0.03	0.007	0.044	0.806	0.792	0.807	0.024	0.006	0.035
10/12	0.007	0.001	0.01	0.88	0.887	0.88	0.006	0.001	0.009
11/12	0.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000
12/12	0.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000

（二）2018 年全国及城乡多维相对贫困识别结果

在 2018 年，不存在 12 个指标同时被剥夺的情况，最大剥夺数量为 11，与 2015 年的情况相似，显示了一定的连续性和稳定性在贫困问题的处理中。多维相对贫困发生率（H）、平均多维剥夺份额（A）以及多维相对贫困指数（M）均显示农村地区的相对贫困水平显著高于城镇地区。随着剥夺指标的数量增加，城乡间的差异有所减小，但在剥夺程度较高的指标数量区间，城乡差异依然存在。随着指标剥夺数量的增加，H 与 M 呈现下降趋势，而 A 呈现上升趋势。当指标剥夺数量达到较高水平时（超过 10），H 与 M 值接近于 0，显示出贫困在高剥夺水平下的集中性和顽固性。

从 2018 年的 H、A 与 M 的变化趋势来看，当指标剥夺数量≤4/12 时，H 与 M 变化较小，下降速度较慢，而 A 保持在 35%到 45%的区间波动，M 处于 29%到 41%的区间，表明在较低剥夺水平下，多维相对贫困问题依然普遍存在。这一阶段，接近 100%的家庭至少存在 1 个指标被剥夺，超过 90%的家庭至少存在 2 个以上的指标被剥夺。当 5/12≤指标剥夺数量≤7/12 时，此时 A 的波动范围上升至 40%到 60%，M 处于 9%到 33%的区间，此时 H 与 M 下降速度加快，相对较低。这一阶段，超过 60%的家庭至少存在 5 个指标被剥夺，表明在中等剥夺水平下，较大比例的家庭面临多维贫困问题。当指标剥夺数量≥8/12 时，我国整体上的 H 与 M 迅速下降并趋近于 0，显示出在高剥夺水平下，遭受剥夺的家庭数量相对较少，而这些家庭的平均剥夺份额 A 接近或达到 100%，反映出这些家庭遭受的贫困剥夺极其严重。

表 4 2018 年全国及城乡多维相对贫困识别结果

剥夺数量	多维相对贫困发生率 H			平均多维剥夺份额 A			多维相对贫困指数 M		
	全国	城镇	农村	全国	城镇	农村	全国	城镇	农村
1/12	0.992	0.982	0.999	0.36	0.312	0.391	0.357	0.306	0.391
2/12	0.924	0.857	0.969	0.377	0.339	0.399	0.348	0.29	0.387
3/12	0.752	0.623	0.838	0.416	0.388	0.429	0.312	0.242	0.359
4/12	0.495	0.352	0.589	0.477	0.46	0.484	0.236	0.162	0.285
5/12	0.285	0.183	0.353	0.547	0.534	0.552	0.156	0.098	0.195
6/12	0.184	0.111	0.233	0.595	0.58	0.6	0.11	0.064	0.14
7/12	0.104	0.055	0.136	0.644	0.634	0.647	0.067	0.035	0.088

8/12	0.031	0.016	0.022	0.719	0.706	0.722	0.022	0.011	0.719
9/12	0.008	0.003	0.01	0.772	0.753	0.776	0.006	0.002	0.008
10/12	0.001	0.000	0.001	0.854	1.000	0.854	0.001	0.000	0.001
11/12	0.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000
12/12	0.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000

（三）2020 年全国及城乡多维相对贫困识别结果

2020 年，延续了之前年度的趋势，表明即使在极端贫困情况下，也没有出现所有指标同时被剥夺的现象。无论是多维相对贫困发生率（H）、平均多维剥夺份额（A）还是多维相对贫困指数（M），我国农村地区的相对贫困水平依然显著高于城镇地区。然而，随着剥夺数量的增加，城乡之间的差距有所减小，但在高剥夺数量区间，城乡差异依然显著。随着指标剥夺数量的增加，H 与 M 保持下降趋势，而 A 则呈现逐渐上升的趋势。当剥夺数量达到较高水平（超过 10）时，H 与 M 值接近于 0，显示出贫困在高剥夺水平下的集中性和顽固性。

从 2020 年的 H、A 与 M 的变化趋势来看，当指标剥夺数量≤4/12 时，H 与 M 变化较小，下降速度较慢，而 A 保持在 30%到 50%的区间波动，M 处于 29%到 41%的区间，表明在较低剥夺水平下，多维相对贫困问题依然普遍存在。在这一阶段，接近 100%的家庭至少存在 1 个指标被剥夺，超过 80%的家庭至少存在 2 个以上的指标被剥夺。当 5/12≤指标剥夺数量≤7/12 时，A 的波动范围上升至 50%到 70%，M 处于 15%到 30%的区间，此时 H 与 M 下降速度加快，相对较低。在这一阶段，超过 30%的家庭至少存在 5 个指标被剥夺，24%的家庭至少存在 6 个指标被剥夺，14%的家庭至少存在 7 个指标被剥夺。当指标剥夺数量≥8/12 时，我国整体上的 H 与 M 迅速下降并趋近于 0，显示出在高剥夺水平下，遭受剥夺的家庭数量相对较少，而这些家庭的平均剥夺份额 A 接近或达到 100%，反映出这些家庭遭受的贫困剥夺极其严重。此外，在指标剥夺数量为 1~12 时，农村地区的 H 与 M 对城镇地区的平均倍数分别是 1.45 与 1.42，其中在指标剥夺数量为 7 时的倍数最高，为 1.82 与 1.83 倍。这表明，农村地区在多维贫困的广度和深度上都显著高于城镇地区。

总的来说，2015-2020 年的数据表明，我国的多维相对贫困问题依然存在，尽管总体上贫困发生率有所下降，但贫困的深度依然严峻，尤其是在农村地区。随着剥夺指标数量的增加，贫困发生率和贫困指数迅速下降，但剥夺份额的增加表明贫困的深度加剧。未来的扶贫工作需要持续关注和改善贫困群体的具体状况，特别是那些处于高剥夺水平的家庭。这些数据为评估和调整扶贫政策提供了重要的实证依据，有助于精准施策和提高扶贫效率。

表 5 2020 年全国及城乡多维相对贫困识别结果

剥夺数量	多维相对贫困发生率 H			平均多维剥夺份额 A			多维相对贫困指数 M		
	全国	城镇	农村	全国	城镇	农村	全国	城镇	农村
1/12	0.996	0.994	0.998	0.396	0.35	0.429	0.395	0.348	0.428
2/12	0.949	0.906	0.979	0.41	0.372	0.435	0.389	0.337	0.426
3/12	0.831	0.74	0.896	0.438	0.408	0.455	0.364	0.302	0.408
4/12	0.597	0.465	0.69	0.492	0.473	0.501	0.294	0.22	0.346
5/12	0.364	0.246	0.447	0.56	0.553	0.563	0.204	0.136	0.252
6/12	0.244	0.155	0.307	0.606	0.603	0.607	0.148	0.094	0.186
7/12	0.15	0.096	0.189	0.653	0.646	0.655	0.098	0.062	0.124
8/12	0.051	0.029	0.067	0.727	0.728	0.727	0.037	0.021	0.049
9/12	0.019	0.011	0.025	0.774	0.781	0.772	0.015	0.009	0.02
10/12	0.002	0.001	0.002	0.856	0.875	0.85	0.002	0.001	0.002
11/12	0.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000
12/12	0.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000

五、多维相对贫困的动态演变分析

通过 A-F 多维贫困测度方法对我国农户的多维贫困指数进行测算，并在时间维度的基础上，利用核密度估计函数分析我国农户多维相对指数的密度分布及动态演变情况。2015 年、2018 年和 2020 年我国农户多维相对贫困指数的核密度估计如图所示。

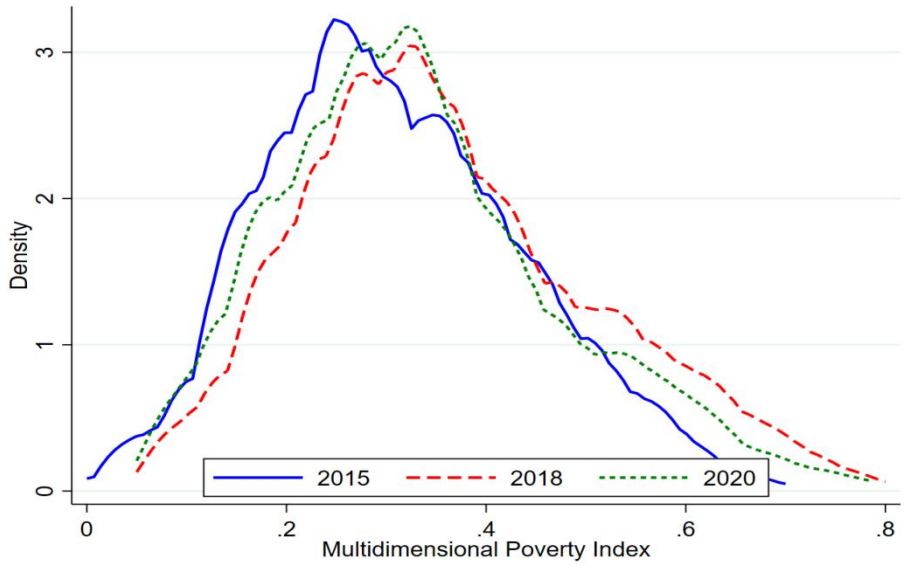

图 5 2015-2018 年我国农户多维相对贫困指数核密度估计图

图中显示了 2015 年、2018 年和 2020 年三个年份的多维贫困指数的核密度估计结果，分别用蓝色实线、红色虚线和绿色短虚线表示。每条曲线代表该年份所有农户的多维贫困指

数的分布情况。从图中可以观察到，2015 年的多维贫困指数在 0.2 到 0.3 之间达到峰值，密度值最高，这表明在 2015 年，大多数农户的多维贫困指数集中在这一范围内。2018 年的峰值略微向右移动，表明农户的贫困状况有所改善。2020 年的峰值进一步向右移动，显示出农户贫困状况的持续改善。这种峰值的右移表明随着时间的推移，整体的贫困程度在减轻。

2015 年和 2018 年的曲线较为接近，表明在这段时间内，农户的贫困状况变化相对较小。然而，2020 年的曲线相对于前两个年份的曲线明显向右平移，并且曲线的形状变得更加平缓。这表明贫困指数的分布变得更加均匀，农户之间的贫困程度差异有所减少。这也可能反映出国家扶贫政策的广泛覆盖和整体效果。2020 年的曲线尾部比 2015 年和 2018 年的曲线更为延展，表明尽管整体贫困状况有所改善，仍有一部分农户的贫困指数较高，处于较为严重的贫困状态。尾部的延展提示我们，虽然大多数农户的贫困状况有所改善，但仍有一些群体未能充分受益，需要进一步关注和支持。

这种变化趋势可能与国家在这段时间内实施的一系列精准扶贫政策有关。自 2015 年以来，中国政府通过多种措施改善农户的生活条件，提高教育水平，改善医疗卫生服务和住房条件，这些措施对整体贫困状况的改善起到了积极作用。通过对比不同年份的密度曲线，可以直观地看到扶贫政策在不同时间段的效果。2015 年曲线的高峰和陡峭表明当时贫困集中度较高，而 2020 年曲线的平缓和右移反映了扶贫工作的显著成效。这种对比分析有助于评估政策的有效性，并为未来的扶贫策略提供调整依据。

总体来看，图中的核密度估计结果展示了我国农户多维贫困指数在 2015 年至 2020 年期间的动态变化趋势。随着时间的推移，农户的贫困状况有所改善，贫困指数的分布更加均匀，贫困集中度有所减少。然而，仍有部分农户处于较为严重的贫困状态，未来的扶贫工作需要进一步聚焦这些群体，确保扶贫政策的全面覆盖和持续效果。

六、我国农户多维贫困的影响因素研究

（一）基于随机森林的农户陷入多维相对贫困特征重要性分析

1. 变量选择

通过文献研究及调查结果，我们将性别、年龄、教育支出、家庭总收入、健康状况、医疗保障、医疗支出、工作保障和生活条件等作为随机森林模型的特征变量，然后将是否处于多维相对贫困（被剥夺的维度数值>指标总数的 1/4）作为分类的标准。然后，将上述数据

进行编码，利用 matlab 编程实现模型。

2. 训练集与测试集的划分

我们采用交叉验证法，也就是 Cross-validation 对训练集与测试集进行划分。交叉验证法是一种统计学上将数据样本切割成较小的子集的方法。其主要目的是先在一个子集上做分析，然后在其他子集上做确认及验证。这样就可以减少过拟合和选择偏差等问题。

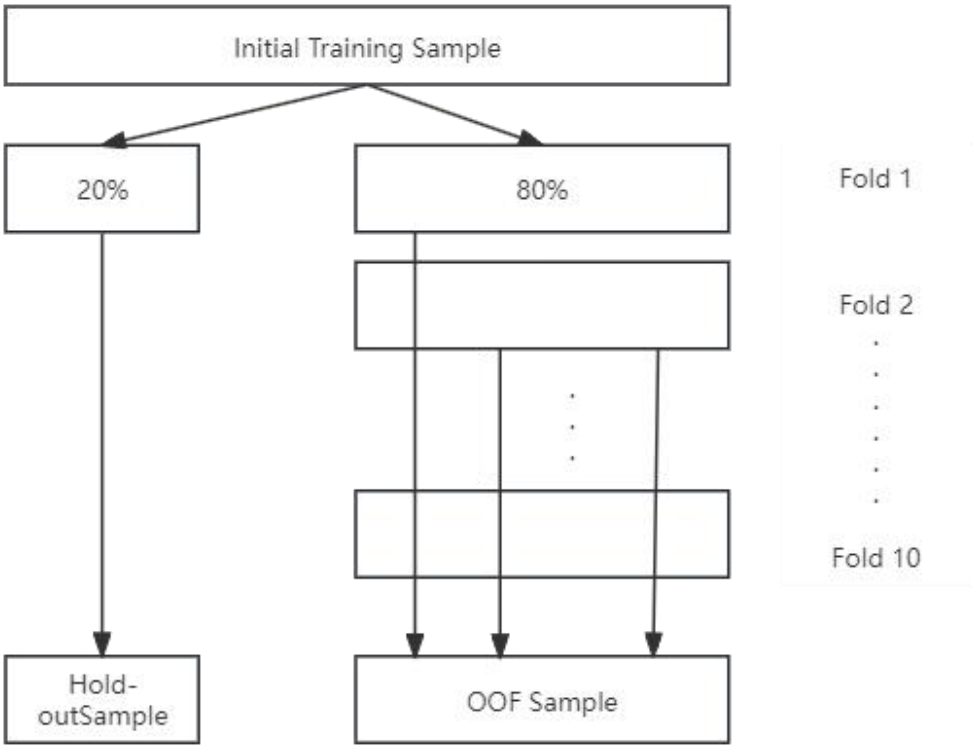

图 6 10-fold cross-validation 示意图

然后，在我们所建立的随机森林模型中，我们所使用的是十折交叉验证，也就是 10-fold cross-validation，将训练集分割成 10 个子样本，一个单独的子样本被保留作为验证模型的数据，其他 9 个样本用来训练。交叉验证重复 10 次，每个子样本验证一次，平均 10 次的结果或者使用其它结合方式，最终得到一个单一估测。

3. 特征重要性

通过利用 matlab 计算随机森林分类模型中各个特征的重要性(Permutation Importance)，其结果如下图所示：

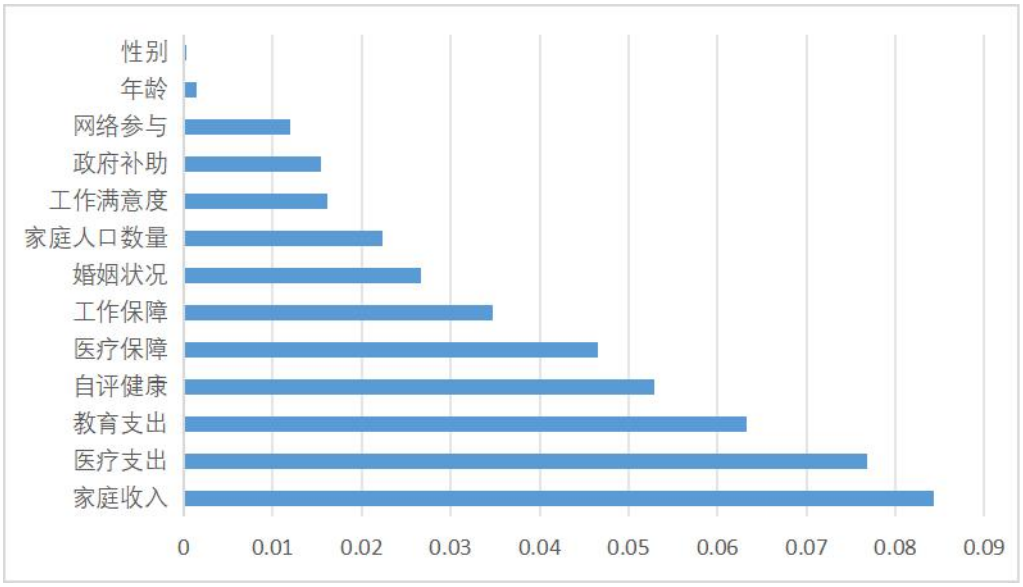

图 7 随机森林模型中的特征重要性图

从上图中，可以发现在随机森林分类模型中，对于影响农户是否处于相对贫困的特征变量里面，特征重要性排在前三名的分别是：家庭收入、医疗支出、教育支出。而特征重要性排在后三名的是：性别、年龄、网络参与。

（二）我国农户多维相对贫困的影响因素分析

1. 变量选取

为了深入探讨农村家庭多维相对贫困的成因并解析其致贫机制，本研究选择了农户的多维相对贫困状态、户主的个人特征、家庭特征以及农户所在村落的特征作为分析对象进行实证研究。

①农户多维相对贫困状态的定义与测量。

本文借鉴已有的研究方法，采用了 K=3（即权重临界值为 1/3）作为判断农户是否处于多维相对贫困状态的标准。根据这一标准，如果农户在临界值以上，则判定为处于多维相对贫困状态，赋值为 1；如果不在临界值以上，则赋值为 0。因此，本研究中的多维相对贫困状态被定义为一个二值变量。这种方法使得多维贫困的测定更为明确，便于后续的数据分析和结果解释。

②户主特征变量。

户主特征包括以下几个方面：年龄，年龄可能影响家庭的经济决策和资源获取能力；性别，性别差异可能导致不同的家庭角色分配和经济活动参与程度；婚姻状态，婚姻状态可以影响家庭的经济合作与支持系统；自评健康，户主的健康状况直接影响到劳动力的质量与数量，进而影响家庭的经济状况和生活质量。

③家庭特征变量。

家庭特征涵盖以下内容：家庭人口数量，反映家庭的劳动力供给和经济负担，是重要的经济条件指标；社会网络，使用人情礼支出的对数值作为代理变量，反映家庭在社会中的关系网络和互助能力，通常礼金数额的大小能够显现家庭的社会资本规模；个体私营和农林牧副渔活动，是否从事这些活动可能影响家庭的收入来源和经济稳定性；政府补助和集体土地，这两项资源的拥有情况可以显著提升家庭收入，减轻贫困状况；家庭医疗支出和教育支出负担，医疗和教育支出是重大的家庭经济负担，高支出可能导致贫困或加剧已有的贫困状况。

2. 模型设定

由于本研究的被解释变量为农户的多维相对贫困状态，是一个二值变量，因此本文根据研究需要将模型设定为二元 Logistic 模型。同时为了加强实证分析结果的稳健性，也将使用二元 Probit 模型进行估计，以更科学地估计多维相对贫困的影响因素。

$$MRPK_i = \beta_0 + \sum_j \beta_j X_{ij} + \varepsilon_i \quad (5)$$

其中，$MRPK_i$为第 i 个农户的多维相对贫困状态，处于多维相对贫困状态为 1，否则为 0，β_0是截距项，X 代表农户的户主特征变量、家庭特征变量和村居特征变量，ε_i为随机扰动项。

表 6 回归结果

变量	Logit 模型		Probit 模型	
	回归系数	边际效应	回归系数	边际效应
年龄	0.037***	0.007***	0.023***	0.006***
	(0.010)	(0.002)	(0.005)	(0.001)
性别	-0.350	-0.068	-0.212	-0.067
	(0.253)	(0.046)	(0.153)	(0.045)
婚姻状态	-0.814**	-0.155**	-0.502**	-0.157**
	(0.397)	(0.078)	(0.237)	(0.075)
自评健康	-0.237**	-0.046**	-0.147**	-0.047**
	(0.099)	(0.014)	(0.054)	(0.014)
家庭人口	0.072	0.014	0.047	0.012
	(0.074)	(0.012)	(0.045)	(0.015)
网络参与	-0.053	-0.011	-0.033	-0.012
	(0.046)	(0.009)	(0.025)	(0.009)
是否从事个体私营	-0.093	-0.016	-0.047	-0.016
	(0.477)	(0.088)	(0.277)	(0.086)
是否从事农林牧副渔	-0.097	-0.015	-0.071	-0.023
	(0.281)	(0.051)	(0.176)	(0.055)
是否享有政府补助	0.201	0.035	0.130	0.040
	(0.251)	(0.046)	(0.148)	(0.045)
是否拥有集体土地	-0.489	-0.090	-0.301	-0.095
	(0.419)	(0.077)	(0.251)	(0.076)
医疗支出负担	2.746***	0.512***	1.664***	0.516***
	(0.760)	(0.139)	(0.441)	(0.133)

教育支出负担	-2.013*** (0.736)	-0.376*** (0.133)	-1.213*** (0.445)	-0.379*** (0.137)
截距项	-0.683 (0.841)	-	-0.401 (0.507)	-
-				
观测者	2278	2278	2278	2278
Pseudo R2	0.201		0.202	
Wald chi2(12)	85.21		100.23	
P-value	0.000		0.000	

注***、**、*分别代表在 1%、5%、10%的显著性水平下显著;括号内为稳健标准误。

分析表明，户主年龄与多维相对贫困之间存在正相关关系，即随着户主年龄的增加，农户陷入多维相对贫困的可能性增加。这可能是因为随着年龄增长，劳动力减弱，收入降低，导致贫困风险增加。婚姻状态和健康水平这两个变量的回归系数和边际效应均为负值，表明婚姻稳定和良好的健康状况能显著减少多维相对贫困的风险。这可能因为婚姻带来的经济和社会支持以及良好的健康状况维持了稳定的劳动能力。户主性别变量的回归系数不显著，说明在本研究样本中，户主性别对多维相对贫困状态的影响不明显。

家庭人口数量、社会网络等因素这些因素的回归系数不显著，表明它们对多维相对贫困状态的直接影响有限。家庭医疗支出负担回归系数和边际效应均为正值，说明高额的家庭医疗支出是导致多维相对贫困的重要因素。医疗费用的增加可能迅速耗尽家庭资源，加剧贫困状况。家庭教育支出的回归系数和边际效应为负值，表明增加教育投入能显著降低贫困风险。这突显了教育在提高家庭经济状况和社会地位方面的重要作用，符合人力资本理论的预期。

七、研究结论与政策建议

本研究通过 A-F 双界法对我国农户家庭的多维相对贫困状态进行了深入的度量，并明确了贫困的主要维度和驱动因素。我们的分析揭示了以下几个关键发现：首先，我国农村地区在教育和收入方面面临显著的多维相对贫困问题，这两个因素对整体贫困指数的影响最为重大。此外，我们发现河北省农户中 16 岁及以上成员的平均受教育年限较低，自评健康状况不佳，以及现金和储蓄较少，这些因素都严重限制了他们的生活质量和经济机会。其次，较高的家庭医疗支出显著增加了农户陷入多维相对贫困的风险。这些因素表明，因病和因灾是导致贫困的主要外部因素。

鉴于上述发现，我们建议河北省采取以下策略以缓解农村地区的多维贫困问题：第一，应加强对农村多维贫困问题的系统治理，特别是在教育和经济发展方面，通过提供教育机会和促进内生发展动力，而非仅依靠外部补助。通过教育培养和技能提升，激发农户的自我发

展潜能，从根本上改善其生活和经济条件。第二，应提升农村地区的教育资源和设施，普及成人及儿童教育，并改善教育质量，确保农村儿童能够获得平等的教育机会。加强职业和技能教育，以提高农民的就业能力和生产效率。第三，增强农户的抗灾抗病能力，通过提高医疗保险覆盖率和报销比例，减轻医疗经济负担，同时加强防灾减灾设施和教育，提高农户的灾害应对能力。最后，改善和加强农村基础设施，特别是交通和物流系统，缩短农村与县城之间的距离和成本，促进农产品的流通和农民的经济活动，从而提高农村地区的经济发展潜力和居民的生活水平。通过这些策略，可以有效地减轻农村地区的多维相对贫困问题，促进农村的全面发展和农户福祉的提升。

参考文献

[1]祝振华,张红丽,李洁艳.城乡相对贫困的动态识别与量化分解——兼论相对贫困线的设定[J].农业经济与管理,2023,(05):83-94.

[2]王晶,高艳云,高燕.中国多维相对贫困的空间格局与动态演进研究[J].统计与决策,2023,39(11):55-60.

[3]周常春,李文会.共同富裕视角下农村数字化与农户多维相对贫困：影响分析与作用机制[J].农林经济管理学报,2023,22(04):397-405.

[4]甘晓成,蔡瑶瑶,肖鸿波.中国多维相对贫困测度及其分布动态演进[J].统计与决策,2023,39(06):50-55.

[5]田祥宇.乡村振兴驱动共同富裕：逻辑、特征与政策保障[J].山西财经大学学报,2023,45(01):1-12.

[6]陆汉文,杨永伟.从脱贫攻坚到相对贫困治理：变化与创新[J].新疆师范大学学报(哲学社会科学版),2020,41(05):86-94+2.

[7]左停,苏武峥.乡村振兴背景下中国相对贫困治理的战略指向与政策选择[J].新疆师范大学学报(哲学社会科学版),2020,41(04):88-96.

[8]叶兴庆,殷浩栋.从消除绝对贫困到缓解相对贫困：中国减贫历程与2020年后的减贫战略[J].改革,2019,(12):5-15.

[9]高强,孔祥智.论相对贫困的内涵、特点难点及应对之策[J].新疆师范大学学报(哲学社会科学版),2020,41(03):120-128+2.

[10]孙久文,夏添.中国扶贫战略与2020年后相对贫困线划定——基于理论、政策和数据的分析[J].中国农村经济,2019,(10):98-113.

[11]汪三贵.汪三贵.当代中国扶贫[M].中国人民大学出版社:201908.188.

[12]周强,张全红.中国家庭长期多维贫困状态转化及教育因素研究[J].数量经济技术经济研究,2017,34(04):3-19.

[13]牟秋菊.农村金融扶贫供给侧结构性改革初探——基于尤努斯的小额信贷扶贫实践反思[J].新金

融,2016,(11):28-31.

[14]马瑜,李政宵,马敏.中国老年多维贫困的测度和致贫因素——基于社区和家庭的分层研究[J].经济问题,2016,(10):27-33.

[15]杨慧敏,罗庆,李小建,高更和.生态敏感区农户多维贫困测度及影响因素分析——以河南省淅川县 3 个村为例[J].经济地理,2016,36(10):137-144.

[16]揭子平,丁士军.农户多维贫困测度及反贫困对策研究——基于湖北省恩施市的农户调研数据[J].农村经济,2016,(04):40-44.

[17]高帅,毕洁颖.农村人口动态多维贫困:状态持续与转变[J].中国人口·资源与环境,2016,26(02):76-83.

[18]祝树民.发挥政策性金融扶贫主导作用全力支持精准扶贫[J].农业发展与金融,2015,(12):11-12.

[19]张全红,周强.中国农村多维贫困的动态变化:1991—2011[J].财贸研究,2015,26(06):22-29.

[20]刘艳华,徐勇.中国农村多维贫困地理识别及类型划分[J].地理学报,2015,70(06):993-1007.

[21]周孟亮,彭雅婷.我国连片特困地区金融扶贫体系构建研究[J].当代经济管理,2015,37(04):85-90.

[22]向玲凛,邓翔.西南少数民族地区反贫困政策绩效研究[J].三峡大学学报(人文社会科学版),2014,36(04):65-68.

[23]高帅.农村人口食物安全测度与动因——基于家庭微观面板数据的分析[J].上海财经大学学报,2014,16(02):36-42.

[24]高艳云,马瑜.多维框架下中国家庭贫困的动态识别[J].统计研究,2013,30(12):89-94.

[25]石智雷,邹蔚然.库区农户的多维贫困及致贫机理分析[J].农业经济问题,2013,34(06):61-69+111.

[26]王素霞,王小林.中国多维贫困测量[J].中国农业大学学报(社会科学版),2013,30(02):129-136.

[27]高艳云.中国城乡多维贫困的测度及比较[J].统计研究,2012,29(11):61-66.

[28]张青.相对贫困标准及相对贫困人口比率[J].统计与决策,2012,(06):87-88.

[29]郭建宇,吴国宝.基于不同指标及权重选择的多维贫困测量——以山西省贫困县为例[J].中国农村经济,2012,(02):12-20.

[30]邹薇,方迎风.关于中国贫困的动态多维度研究[J].中国人口科学,2011,(06):49-59+111.

[31]王小林,Sabina Alkire.中国多维贫困测量:估计和政策含义[J].中国农村经济,2009,(12):4-10+23.

[32]李丹.北京市城镇居民贫困测度及亲贫困增长判定[J].统计教育,2008,(06):13-18+36.

[33]李春雨,刘志彪.区域经济发展与趋势预测研究[J].统计研究,2002,(11):51-53.

[34]Terzi S. How to integrate macro and micro perspectives: An example on human development and multidimensional poverty[J]. Social indicators research, 2013, 114: 935-945.

[35]Wagle U. Rethinking poverty: definition and measurement[J]. International Social Science Journal, 2002, 54(171): 155-165.

[36]Townsend P. Poverty in the United Kingdom: a survey of household resources and standards of living[M]. Univ of California Press, 1979.

[37]Townsend,P., 1954. Measuring poverty［J］.British journal of sociology,1954,5 (2) : 130－137.

[38]Sen A. Issues in the Measurement of Poverty[M]//Measurement in public choice. London: Palgrave Macmillan UK, 1981: 144-166.

[39]Sen A. Capabilities, lists, and public reason: continuing the conversation[J]. Feminist economics, 2004, 10(3): 77-80.

[40]Rowntree B S. Poverty: A study of town life[M]. Macmillan, 1902.

[41]Strobel C D. Recent Developments on the Fringe Benefit and T&E Front[J]. Journal of Corporate Accounting & Finance (Wiley), 1996, 8(1).

[42]Taylor A M. Latin America and foreign capital in the twentieth century: economics, politics, and institutional change[J]. 1999.

[43]Alkire S, Kanagaratnam U, Suppa N. The global multidimensional poverty index (MPI) 2021[J]. 2021.

[44]Wang X. On the relationship between income poverty and multidimensional poverty in China[M]//Multidimensional Poverty Measurement: Theory and Methodology. Singapore: Springer Nature Singapore, 2022: 85-106.

[45]Bárcena-Martín E, Pérez-Moreno S, Rodríguez-Díaz B. Rethinking multidimensional poverty through a multi-criteria analysis[J]. Economic Modelling, 2020, 91: 313-325.

[46]Alkire S, Kanagaratnam U. Revisions of the global multidimensional poverty index: indicator options and their empirical assessment[J]. Oxford Development Studies, 2021, 49(2): 169-183.

[47]Aguilar G R, Sumner A. Who are the world ' s poor? A new profile of global multidimensional poverty[J]. World Development, 2020, 126: 104716.

Research on the Dynamic Evolution of Multidimensional Relative Poverty and Influencing Factors of Farming Households in China in the Era of Big Data and Artificial Intelligence

Li Qianqian

(Business School of Hunan Normal University, Changsha / Hunan, 410000)

Abstract： In order to solve the deep-rooted problems of relative poverty, we explore the dynamic evolution of multidimensional relative poverty and its influencing factors of China's farm households, so as to serve the goal of precise poverty alleviation. Using Python crawler and text mining to sort out the research hotspots, combining with the data of China Household Tracking Survey (CHARLS), the multidimensional poverty index is measured by A-F multidimensional poverty measure, and the key influencing factors are analyzed by Random Forest and Logistic regression models. The results show that the multidimensional relative poverty of China's farm households is high, with significant contributions from the education and income dimensions, the burden of household medical expenditures is the main poverty-causing factor, and poverty due to illness and disaster is particularly prominent. Policy recommendations include: strengthening education and skills training, improving the medical security system, strengthening disaster prevention and mitigation capacity, and reducing the risk of multidimensional poverty.

Keywords: Multidimensional relative poverty; A-F multidimensional poverty measures; random forests; logistic regression models; poverty due to illness and disaster