# 浅谈税务数据挖掘与深度利用

来源：网络 作者：逝水流年 更新时间：2024-03-03

*\" 中国入世对政府部门的职能提出了严峻挑战，实现税务信息化是税务部门急待解决的问题，而数据挖掘是实现信息化的必由之路。；金税三期展开前后，税务部门都比较关注税务数据的深度利用和数据挖掘问题本文浅谈对税务数据深度利用的理解，并在此基础上提出...*

\"

中国入世对政府部门的职能提出了严峻挑战，实现税务信息化是税务部门急待解决的问题，而数据挖掘是实现信息化的必由之路。；金税三期展开前后，税务部门都比较关注税务数据的深度利用和数据挖掘问题本文浅谈对税务数据深度利用的理解，并在此基础上提出应该考虑的核心工作，最后介绍在税务数据利用平台建设方面的一些心得体会。

一、对税务数据深度利用的理解

长期以来，税收工作中数据利用比较常见的形式有：报表浏览、简单查询、复杂查询、税源分析、税负分析、收入预测、过程监控等，多数专家认为，目前税务数据应用的一般特征是基于汇总、分类、简单计算基础之上的原始税收数据的“复制式”展现和对税收现象的“陈列式”描述。

随着经济、社会的发展，税收数据的般利用已经不能满足税收信息化深化和税收管理现代化的内在需求，为了加强税收征管、规范税收秩序，国务院于1994年开始实施“金税工程”。“金税工程”初期以“增值税监管”为主要目标；二期时，内容已拓宽为增值税防伪税控开票系统、防伪税控认证系统、增值税计算机交叉稽核系统、发票协查信息管理系统的四个系统；到了三期，其目标已经成为：在二期基础上，建立七个子系统(管理子系统、征收子系统、稽查子系统、处罚子系统、执行子系统、救济子系统、监控子系统)，35个模块。依据美国学者Richard．L-Nolan的理论(对于任何行业，信息化大体要经历初始、蔓延、控制、集成、数据管理和成熟这样几个发展阶段，这是信息化发展的般规律。)和Mische的补充(他认为集成和数据管理是密不可分的，因此信息化发展的必然路径是起步、增长、成熟和更新四个阶段)，目前，税务信息化的发展阶段已开始向成熟阶段过渡。于是税收数据的深度利用便提上日程。我们可以从税收管理战略和税收政策分析两方面来看这种需求的提出。

从税务管理战略来看，在纳税前如何综合评价简化管理制度(法律)及照章纳税宣传的相对效果；在纳税中如何核算税收结构和管理程度的实际资源成本(管理、照章纳税、效率、逃税)，以及纳税后对税收差距的衡量(包括潜在税收与申报税收的差距、申报税收与实收税收的差距、实收税收与送达国库税收的差距)，都涉及到税收数据的深度利用问题。

从税收政策分析的角度来看，税收经济的和谐发展度量、税制改革方案分析、税收减免和优惠的成本和政策收益、税收政策的经济影响等等也涉及到税收数据的深度利用和挖掘问题。

这些问题都从以下两方面引发了我们对税务数据深度利用的理解和思考：一方面提出了我们需要全面检视拥有的税务数据信息的需求。在各国税务数据信息深度利用的经验当中，提出过一些全面检视的标准，例如按照税基到税收收入的实现途径，可以检视：税基的规模，包括真实税基和潜在税基；税收管理资源使用方向的详细分类；管理资源使用的效用；税收管理的效果，例如收到税款的多少，处理案件的数量等。

另一方面，更为关键的是，提出了如何科学利用、深度利用的问题。总结以上两方面，我们认为税收数据的深度利用是指：在数据集中和系统整合的基础上，建立全面的税务数据信息，既包括税务系统内部数据，也包括其他政府部门、企业、居民等外部数据，并且进一步在各种模型的帮助下，发现数据的内在规律。就目前而言，重点任务是在税务管理方面提出适用中国实践的模型并且应用，同时初步探索在税收经济方面能够刻画符合我国国情的模型。\"

二、构建模型是数据深度利用的切入口

如前所述，税务数据深度利用和挖掘的关键在于模型的应用，下面我们就来讨论模型是什么?我们为什么需要模型?我们需要什么样的模型?就税收数据深度利用的模型而言，大致可以分为两类：以科学化管理、定量化管理、精细化管理为内在思想的管理工具和手段所形成的模型；以研究税收经济关系协调发展为目的的税收经济模型。

在基本认识了税务数据深度利用中的模型是什么之后，虽然我们达成了一种共识，我们需要模型，但是如果我们思考过为什么需要模型?显然会对模型应用更能得心应手。我们认为模型所发挥的作用无外乎以下三种：

首先，刻画税收经济关系。一般而言，我们经常提到的是模型在刻画税收经济关系当中所起的作用，即采用代数形式的定量分析将税收经济理论模型化，然后适当根据实践情况把理论模型予以修正，并将相关数据应用到修正模型中，对模型结果进行经验分析。这种利用的过程是阶段性的，是从初级到高级的过程，是一个水平不断提高、效果不断改进的发展过程。

其次，归纳税收管理实践。模型起到的作用是将复杂的税收征纳活动通过数字化的形式总结归纳，将税收征纳的每一个过程精细化、每一个结果数据化，并且建立起投入到产出之间的对应关系。最后，数据组织的导向性作用。这种导向性作用的发挥是通过模型应用过程当中对各类数据提出的要求实现的，通过该作用，随着时间的发展，数据集中的有效性与目的性不断加强，反之，模型应用空间不断扩展。在这一方面，美国个人所得税模型应用为我们提供了很好的启迪。

结合目前的税务数据基础及其发展趋势来看，金税三期将成为税收数据深度利用的良好契机，构建相应模型是我们形成税务数据深度利用良好局面的切入口。

对于“我们需要什么样的模型”的回答，是一个不断结合实际进行摸索的过程，但是就现阶段而言，从可操作性的角度出发，我们还是需要给所应用的模型框定一个边界：数据可利用性，如果没有数据的支持，模型应用将无从谈起；可计算性，模型应用迅速发展的基石之一就是现代计算技术的发展，没有计算工具的支持，具有庞大计算量的各种税收模型的完成无法想象，支持税收模型应用的计算工具包括硬件具备的计算能力和软件具有的算法能力两种。

三、数据深度利用平台建设的体厶

数据深度利用和挖掘最终必须落实到具体计算平台上，否则纵然有大量的数据积累，仍然摆脱不了研究与实践部门脱节的窘态。虽然目前我们拥有大量的计算软件平台，然而，总感觉到这些应用平台离我们的实际需要有一定距离。

目前我们正在参与完成一个国家自然科学基金研究项目：税收政策分析模型支持系统的实现及其在税制改革中的应用研究，其主要内容和实质就是探索构建一个有利干数据深度利用的计算平台。从该平台的构建来看，有几点体会：首先，平台的构建必须结合具体的研究问题展开。通用性的平台虽然很好，但是由于前面所提到的数据可利用性和可计算性的原因，加上实际工作的紧迫性需求，往往使得通用性平台的规划会落空，甚至于进一步影响数据深度利用工作本身。在该问题上，我们的平台研究就结合了增值税转型的测算问题，利用了CGE平台进行实证性的应用。\"

其次，考虑针对具体问题研究的通用性拓展。虽然实用为先，然而要做到持续性的数据利用，必然要考虑拓展的问题。在这个问题上，我们的平台通过税制表示方法、税收政策分析模型描述语言中国税收政策分析模型支持系统等方法进行尝试。

更为重要的是，对适合中国国情的税收经济模型的提出。由于长期以来的数据缺少原因，在我国模型建设方面没有进一步的探索。这种缺陷在海量的数据突然呈现在我们面前的时候更加突出。我们正在尝试提出适用干中国的税收经济模型，虽然肯定会比较艰难，但是这是一条必经之路。

最后，平台的研究必须以应用为导向。将平台应用到实际部门，可以最终检验平台的正确性，带来数据深度利用平台的可持续发展空间，体现出促进税收工作的真实效用，避免科研部门“孤芳自赏”的局面。

当然，在数据深度利用和挖掘过程当中，我们也碰到了其它一些问题，例如模型应用的技术性问题，包括数据调整、不同模型的评估和模型应用的不确定性等，还有如数据积累和维护体系、模型使用反馈机制、结果使用反馈机制等等问题。这些问题都需要我们在数据深度利用过程中加以关注和解决。

本文档由028GTXX.CN范文网提供，海量范文请访问 https://www.028gtxx.cn