

第 1 章 绪 论

1.1 研究背景和意义

根据世界旅行&旅游业委员会 (WTTC) 统计, 至 2021 年, 中国的旅游经济将处于世界第三位, 旅游需求的增长将位于世界第二位。旅游需求的增长直接带动了旅游客流量的急剧增加和旅游经济的快速发展, 旅游成为我国国民经济的重要来源之一。

同时, 随着科技的不断发展, 以信息和物联网为技术支撑的“智慧景区”成为未来旅游业发展的新趋势。“智慧景区”的建成对旅游行业来说, 有利于其建立完善的旅游安全警报体系, 有效地配置旅游资源, 降低未来决策的风险。“智慧景区”建设是一项复杂的系统工程, 其中旅游短期客流量预测是“智慧景区”信息管理平台建设的核心内容之一。如何实现不同时期客流量预测, 并通过智慧景区预测系统及应用平台及时获取预测结果对旅游景区至关重要。但是, 在我国由于四季自然、气候等条件各异, 这些因素对不同地理区域的景区影响程度各不相同。对受自然、气候等因素影响较大的旅游景区, 如黄山、九寨沟、华山等风景区, 旅游客流量的增加并不是呈现均匀分布态势, 加之节假日、旅游突发事件等多种因素影响, 使得客流量在不同时期的不均衡状态更加显著。一方面, 在以节假日为主体的旅游旺季, 游客消费集中释放, 大量游客在同一时间快速集中于同一地点。据 2011 年 2 月全国假日旅游部际协调会议办公室发布的《2011 年春节、十一黄金周旅游统计报告》显示, 2011 年春节期间旅游景区共计接待游客 1.53 亿人次, 比 2010 年春节黄金周增长 22.7%; 国庆节期间接待游客 3.02 亿人次, 同比增长 18.8%, 居民节

假日旅游不仅比往年更加拥挤，而且也更为集中。客流量在短时间内急剧增长，超过景区最大承载量，造成游客超载、旅游资源被过度消费以及由此带来的游客被困、乘车站点拥堵及车辆调度不畅等问题，不仅对旅游景区生态环境造成破坏，而且极易诱发安全事故，降低游客游览质量。近几年，由于旅游景区客流量超载等问题造成的安全事故频发，如华山、九寨沟、张家界及黄山等景区拥堵事件，直接导致了游客与旅游景区的正面冲突，给旅游景区及旅游行业造成了极大的负面影响，也给游客的生命安全造成严重的威胁。另一方面，在以平常日（除节假日外）为基础的旅游淡季，旅游景区、酒店宾馆及相关旅行社接待游客数量相对不足，旅游设施及资源闲置现象存在，造成不必要的人、财、物的浪费。除此之外，旅游景区还面临着另外一种不确定的状况：各种旅游突发事件的冲击。由于旅游突发事件具有产生的瞬间性、爆发点的偶然性和对社会的危害性等特征，会给旅游客流量在短期内产生极大的波动，呈现高度的不确定性和随机性。客流量在不同情况下表现的不均衡状况给旅游景区内部资源管理、调度和科学决策带来很大的挑战。

传统的旅游客流量预测大都建立在感性的管理经验及景区宏观预测基础上。一直以来，由于缺乏清晰的预测方法，导致旅游客流量预测值和真实值之间存在较大偏差，更无法建立客流量与其影响因素如历史客流量、自然气候、节假日及旅游突发事件等因素之间的关系的方法模型。这些预测方法满足不了智慧景区对客流量预测的要求，因此也不能为旅游决策者带来更多的参考和建议。

因此，如何在智慧景区背景下研究旅游客流量影响因素、特点及分类，并利用更有效的方法实现不同时期客流量的准确预测，已成为智慧景区预测系统及应用平台能够建立并实施成功的前提之一，也是旅游企业未来成功发展的一个重要基础，更是旅游需求预测研究中的一个关键科学问题。

然而，与解决上述问题的迫切需求相比，相关的关于旅游客流量预测理论研究在国内尚处于起步阶段。虽然近年来有关旅游需求预测的研究逐渐增加，但是具体到对不同时期旅游客流量预测的研究尚未发现。本研究以智慧景区为研究背景，以旅游客流量不同影响因素研究为切入

口，以不同时期客流量准确预测为主要目标，并通过数据研究、实证分析、智能算法、神经网络和统计学习理论等对智慧景区背景下影响客流量的主要因素进行系统分析，探讨不同时期客流量的特点及分类，探析不同时期客流量的预测方法，并以此为基础，建立智慧景区旅游短期客流量预测方法系统以便为智慧景区建设提供理论支持和实践指导，从而实现中长期预测难以达到的效果。其意义重大，主要体现在以下几个方面：

第一，对智慧景区来说，可以使景区管理部门能够实时根据系统监控情况对景区开放时间、游览方式及游览路径等提前做出规划，为旅游景区在不同时期科学调度、资源统一管理、分流游客等提供直接信息，为旅游管理者在面临各种复杂的环境下进行科学决策提供主要依据；

第二，对游客来说，旅游景区能够实现发布未来客流量情况，可为游客提供出行参考，避免拥堵，从而提高其旅游的质量和满意度；

第三，对酒店、宾馆及交通运输等与旅游相关的行业具有重要的参考价值。

因此，旅游短期客流量预测对旅游行业、旅游景区、游客乃至旅游相关行业都有着至关重要的意义。

由于旅游短期客流量是一个主要由游客参与的、复杂的、不确定的非线性系统，受到诸如自然气候、节假日及旅游突发事件等多种因素的影响，在不同时期呈现不同的特点，因此给短期客流量的预测带来了极大的难度，目前很难用一个统一的方法去实现不同时期的短期客流量预测。

目前，我国多数旅游景区已经逐步开始重视短期客流量的预测工作，但由于信息化起步较晚，有记录的客流量数据样本小，对短期客流量如日客流量的预测都是建立在感性的管理经验及景区宏观预测上，其过程由政府相关统计部门根据每日汽车、火车、飞机等流量信息，参考经济增长、历史数据推算出较为宏观的游客数量增长率；部分历年游客资料较完善的旅游景区可以利用历年来同期同日游客数量，配合天气、政府宏观预测等因素，大致推算出每日游客的数量；少数信息化较早的景区还可以利用门票系统的预售票、每日门票销售情况，结合多年的管理经验，在每日早上推算出当日接待游客数。但是由于客流量与节假日、天气、重大活动、历史客流、突发事件等诸多要素相关，以上推算一般存

在较大的误差，对旅游景区来说可操作性较差，甚至可能会出现因为措施制订失误造成严重的浪费。一直以来，由于缺乏清晰的数学模型，导致短期客流量预测值和真实值之间存在较大偏差，更无法建立客流量与历史客流量、自然气候、节假日等因素之间的定量关系。

黄山风景区作为我国著名的 5A 级山岳风景区，由于其独特的地形地貌、植被和人文景观成为游客热点旅游目的地，而且黄山风景区四季自然、气候等条件分明，客流量常年处于极度不均衡状态，因此本研究将代表性地选取黄山风景区为应用背景，研究以日为主的旅游短期客流量预测问题，通过对不同影响因素及不同时期客流量特点的分析，拟建立准确的、能反映与历史客流量、自然气候、节假日、旅游突发事件等要素有定量关系的不同时期的短期客流量预测模型。

1.2 国内外旅游需求研究方法分析

国际上从 20 世纪 60 年代初期就对旅游需求进行了研究，通过建立不同的预测模型对旅游客流量进行预测，从最初的传统时间序列方法到计量经济学模型，发展到后来的人工智能预测方法，从理论到方法逐渐趋已成熟，并且取得大量的研究成果。

1.2.1 经典时间序列预测方法

经典时间序列预测方法是根据系统观测得到的时间序列数据，发现其历史的趋势和模式（如季节性），并依据这种历史的趋势和模式，通过曲线拟合和参数估计来建立数学模型的理论和方法。

1. 研究现状

经典时间序列预测方法主要以指数平滑模型 (exponential smoothing, ES) 及 Box-Jenkins 预测等方法为主，其中自回归模型 (auto regressive,

AR)、滑动平均模型 (moving average, MA), 以及在此基础上提出的自回归滑动平均模型 (autoregressive moving average, ARMA)、自回归求和滑动平均模型 (autoregressive integrated moving average, ARIMA) 均为 Box-Jenkins 方法的不同形式, ARIMA 模型后来成为 Box-Jenkins 方法的代名词。

初期的旅游客流量时间序列预测是关于一些预测技术的应用及预测准确性的分析, 预测方法以单变量的 Box-Jenkins 预测方法和 ES 等最为常见。Geurts 等人用 Box-Jenkins 预测方法对夏威夷旅游市场每月客流量进行预测, 并同指数平滑方法做比较, 结果发现 Box-Jenkins 预测方法效果较好, 平均误差达到 3.5%, 平均绝对误差达到 8.3%; Liepa 用 Box-Jenkins 预测方法预测美国到加拿大的每月客流量, 认为该方法能够实现较准确的预测。Fritz 等人用 ARIMA 方法与计量经济学模型天真法等结合, 通过游客停留的时间、支出等来预测通过航空途径到达佛罗里达州游客的数量。Witt 等人用指数平滑方法预测拉斯维加斯旅游客流量, 在与天真 1 方法的比较中, 指数平滑 MAPE 值最小。Chu 用了六种不同的预测方法对来自亚太地区的十个不同国家如新加坡、泰国、韩国等到达中国台湾的客流量进行预测, 结果显示以 ARIMA 为基础的模式 MAPE 值最小, 预测效果最好。

进入 21 世纪后, 学者们关于旅游客流量预测的文献有超过三分之二使用的是 ARMA、ARIMA 模型及以它们为基础的季节 ARIMA 模型、ARMAX 模型、广义自回归条件异方差模型 (generalized autoregressive conditional heteroskedasticity, GARCH) 等。这些方法进一步丰富了经典时间序列预测方法, 并在旅游需求预测方面得到极大应用并发挥了重要的作用。

Gustavsson 和 Nordstrom 利用 ARMA 对不同类型瑞典入境游客流量进行预测, 取得很好的预测效果; Lim 和 McAleer 分别使用 ES 模型和 ARMA 模型预测中国香港、马来西亚、新加坡到澳大利亚的月客流量和季客流量; Tideswell 等人认为 ES 模型能够将较大的权数放在最近的数据上, 因此预测的结果更能反映实际的趋势; Cho 用 ARIMA 方法预测美国、英国、新加坡、日本、中国台湾和韩国到达中国香港的客流量; Du

等人预测四个欧洲国家：法国、德国、意大利和英国到 Seychelles 岛的客流量，结果发现 ARIMA 预测效果优于单变量和多变量状态空间模型；KongOh 和 Morzuch、Pai 和 Hong 使用 ARIMA 预测旅游客流量；由于季节性是旅游行业的突出特征，Goh 和 Law、Papatheodorou 和 Song (2005)、雷可为和陈瑛 (2007)、王丽英和刘后平 (2008) 使用季节 ARIMA 模型预测旅游客流量，都取得了很好的预测效果。考虑到 ARIMA、SARIMA 等模型存在预测性能的易变性，学者们努力将单变量的时间序列预测模型向多变量转变，Akai 通过 ARMAX 模型预测土耳其的国际游客到达量；Chan 等人利用三种不同类型的多变量 GARCH 模型预测日本、新西兰、英国和美国到澳大利亚的月客流量，为旅游部门提供了很好的预测和分析工具。其他的一些时间序列预测方法如简单的自回归模型、天真 1、天真 2 也经常被应用于旅游客流量预测，但是这些模型在更多的时候是作为一种衡量预测准确性的对比模型出现的。

2. 存在的问题

在利用经典时间序列方法进行旅游客流量预测时，对历史数据准确性有较高要求，坏数据对其预测效果影响很大，需要严格处理脏数据；同时还要求被预测变量在过去、现在和将来的各种客观条件基本保持不变，历史数据解释的规律可以延续到未来，预测变量的发展过程是渐变的，而不是跳跃式的或大起大落的，因此经典时间序列预测多以线性方法为主，重在线性时间趋势的外推，对旅游客流量生成与影响因素的内在作用机理分析不够，没有考虑其他对旅游需求影响的各种干扰因素，模型中没有体现影响因素和预测值之间的因果关系，对中长期尤其是长期的旅游客流量趋势预测能取得较好的预测效果。旅游短期客流量由于受到众多因素的影响，尤其当遇到一些节假日、突发事件等因素而导致客流量呈现巨大波动性时，非线性性、随机性不确定趋势明显，经典时间序列预测模型很难把握短时客流量的特征，往往难以实现复杂的非线性的短期客流量，因此不能为旅游决策者带来更多的参考和建议。

1.2.2 计量经济学预测方法

计量经济学方法是揭示经济活动中各个因素之间的定量关系,通过对与预测对象有联系的经济现象变动趋势的分析,推算预测对象未来状态数量表现,用随机性的数学方程加以描述的一种预测方法。计量经济学模型在旅游客流量预测方面应用广泛,其主要优点是能够反映出旅游客流量和它的影响因素之间的因果关系,能够从经济角度解释旅游客流量的变化情况,而这一点是经典时间序列预测方法无法做到的。

1. 研究现状

近年来,计量经济学方法关于旅游需求方面的预测主要以误差修正模型(the error correction model, ECM)、向量自回归模型(the vector autoregressive, VAR)、时变参数模型(the time varying parameter models, TVP)、自回归分布滞后模型(the autoregressive distributed lag model, ADLM)和近似理想需求体系模型(the almost ideal demand system model, AIDS)为主。

Kulendran 和 Wilson 等人用 ECM 预测澳大利亚四个重要的贸易伙伴到澳大利亚商务游的客流量数目, Kulendran 和 Witt 用 ECM 预测英国到六个主要目的地的出境游,所用的变量为汇率,研究结果表明 ECM 更适合做中长期预测; Lim 和 McAleer 用 ECM 模型预测新加坡和中国香港到澳大利亚的客流量,使用收入和旅游价格作为影响因素,研究结果同样表明 ECM 更适合做中长期预测。Dritsakis 和 Athanasiadis、Ismail 等人用 ADLM 模型分别预测希腊市场的旅游客流量和日本到关岛的客流量; Song、Witt 和 Jensen 将 ADLM 模型用在丹麦的入境客流量预测上, Song、Wong 和 Chon 也将其用在中国香港的入境客流量预测上。Shan 和 Wilson 使用 VAR 模型研究国际贸易和旅游客流量的关系; Witt 等人通过旅游支出来实现客流量预测,他们认为 VAR 能达到最好的预测效果; Song 和 Witt (2006) 利用三种不同类型 VAR 模型预测六个客源国和地区:中国内地、中国香港、中国台湾、日本、韩国和菲律宾到中国澳门的客流量。Wong 等人用贝叶斯向量自回归模型(the bayesian vector autoregressive, BVAR) 预测客源国:澳大利亚、加拿大、法国、德国、英国和美国到中国香港的客流量。作为一种后来发展起来的计量模型, AIDS 具有更强的

经济理论基础,近年来被用于旅游客流量分析, Durbarry 和 Sinclair、De 等人、Li 等人用 AIDS 方法分别对客流量进行了预测。

近年来,ECM 模型常常也和其他模型结合在一起使用,如 ECM-TVP 模型,有文献将 TVP 和 ECM 优点结合,通过人均旅游消费预测英国到法国、西班牙、意大利、希腊和葡萄牙等国的客流量;有文献将 ECM-AIDS 方法用于旅游客流量相关预测,取得了较单个模型更好的效果。曾忠禄、郑勇(2009)曾通过中国内地人均可支配收入、消费者物价指数、国内生产总值,以及澳门与内地之间商品进出口总额、港币与人民币汇率等变量,建立计量经济学模型来估计内地赴澳门游客客流量。计量经济学模型在国外研究中应用较多,而在国内用于旅游需求研究的较少。

2. 存在的问题

计量经济学方法主要通过游客收入、目的地相对于客源地的旅游价格、竞争目的地的替代价格以及汇率等作为解释变量,通过这些解释变量来反映它们和旅游客流量之间的因果关系,克服了经典时间序列预测模型的缺点。但该类方法主要研究的是以目的地—客源地的客流量预测为主,存在耗时耗费大、影响因子难以确定等问题;而且由于不同目的地—客源地影响客流量的宏观因素各不相同,很难建立一个对所有目的地—客源地都适合的计量方法;同时该类方法所用的数据大多为间接数据,多通过经济变量来间接预测旅游客流量,真正来源于旅游风景区的直接数据相对较少,因此对预测的准确性造成一定的影响。

1.2.3 人工神经网络预测方法

除了经典时间序列预测方法和计量经济学模型外,一些新的定量预测方法被用于旅游客流量预测。人工神经网络(artificial neural network, ANN)方法是其中的一种。作为一种模拟人脑神经性能的数据处理技术,它是由大量处理单元互联组成的非线性、自适应信息处理系统。ANN 被分为两种:前馈神经网络(feedforward networks)和反馈神经网络(feedback networks)。ANN 由于其具有的自学习、对非线性数据和不完

备数据的较强的处理能力、高速寻优能力，已经被广泛地应用于旅游需求预测领域。

1. 研究现状

ANN方法虽然已经被广泛地应用于许多领域，但是具体应用到旅游预测领域的时间不长。神经网络最早应用在旅游预测领域是在1999年，Law等人用前馈神经网络预测日本到中国香港的游客数量，取得了较多元回归、指数平滑等方法较精确的结果。2000年以后学者们才开始将ANN方法应用于该领域，并取得很大的成功。

有文献在前馈神经网络延伸的基础上，用反向传播神经网络（back-propagation neural network, BPNN）预测中国台湾到中国香港的客流量，取得了良好的预测效果，准确性优于时间序列预测方法。有文献研究显示，同天真1、ES、ARIMA等方法对旅游客流量进行预测比较时，ANN是表现最好的方法；同样的结论也被Burger（2001）等人和Cho（2003）证实：他们用ANN方法分别预测美国到南非德班的客流量，美国、英国、日本等国到中国香港的客流量，结果均显示该方法优于ARIMA、ES等方法；Kon和Turner（2005）用ANN预测新加坡的游客到达量时，结果显示ANN预测效果明显好于天真1等方法。Pai和Hong（2005）、Palmer等（2006）、鲍青青等（2008）利用ANN预测旅游客流量，焦淑华等人用BPNN对中国的入境游人数进行预测，张彬等人用BPNN对美国、英国、澳大利亚三个客源国的入境游客进行预测，均取得一定的成果。

2. 存在的问题

ANN对非线性数据有着很强的处理能力，它不依赖于变量之间的关系，对数据分布的特点无特别要求，但也有其自身的缺陷。首先，作为一种计算方法，和传统时间序列模型相比，它缺少理论支撑和系统的建模过程，参数的选取往往要通过反复试验的方法获得，学习过程通常较慢，整个过程耗时长，容易陷入局部最优；其次，ANN方法需要大量的数据进行训练，在国外的旅游客流量预测中，ANN显示出预测的优越性，但是，这一点在我国往往却难以实现，因为我国的旅游景区信息化起步

比较晚，真正完整的保存下来的旅游数据样本小，数量有限，因此在旅游客流量预测方面受到了一定的限制；最后，ANN 技术解释性不强，无法从经济理论视角很好地解释旅游客流量，对客流量短期波动性和季节性不能进行很好的分析和预测，而且由于神经网络的学习过程通常较慢，对突发事件的适应性差，因而不能为政策的评估和决定提供更多的建议与帮助，这也限制了 ANN 在旅游客流量预测方面应用的广度和深度。

1.2.4 支持向量回归预测方法

基于统计学习理论的支持向量回归 (support vector regression, SVR) 方法以其良好的泛化能力、处理小样本、非线性数据的独特能力为非线性时间序列预测提供了一种新的理论指导。近年来，SVR 在预测领域的应用发展迅速，主要集中在金融、电力、交通、旅游等方面的广泛应用。在这些应用中，SVR 主要同其他方法技术结合，实现了非线性预测的目的。

1. 研究现状

将 SVR 应用于旅游需求预测是近几年相对比较新而又富有挑战性的领域。有文献针对多数预测模型解决单因素预测问题，提出一种多因素的 SVR (multifactor SVR) 用于旅游需求预测，结果显示，该模型优于 BPNN。有文献提出将 SVR 和 GA 结合 (GA-SVR) 预测旅游需求，分别预测中国入境游客流量和陕西旅游客流量，用 BPNN 和 ARIMA 作为对比，结果均显示，GA-SVR 明显优于 BPNN 和 ARIMA 方法。有文献提出将线性和非线性模型结合起来预测旅游需求：天真法 (Naïve method)、ES、ARIMA、BPNN、SVR，组成六种混合模型，分别为 Naïve-BPNN、ES-BPNN、ARIMA-BPNN、Naïve-SVR、ES-SVR、ARIMA-SVR，并将六种混合模型分别用于中国台湾出境游的六个地区的年度客流量预测：中国、东北亚、东南亚、美洲、欧洲、大洋洲，结果显示，Naïve-SVR、ES-SVR、ARIMA-SVR 在准确性和转折点的预测上明显好于其他三种方法。Hong 等人将混沌 GA 和 SVR 结合预测 Barbados 入境游，取得了较 GA-SVR 准确的效果。