

第 1 章 绪 论

1.1 研究背景及意义

1.1.1 研究背景

自工业革命以来,由于科学技术的进步、医疗条件的改善和生活水平的提高,全球人口迅猛增长。据美国人口调查局统计,在 18 世纪中期即第一次工业革命期间,全球人口不到 8 亿;然而,截止到 2013 年 1 月,全球人口已达 70.57 亿。这期间,中国人口数量从 2 亿左右增加到了近 14 亿,成为世界上人口增加最快的国家之一。人口的急剧增加给全球的经济发展、社会安定和人们的生活带来诸多问题,比如贫困问题、资源短缺、城市无序发展、交通拥堵、环境破坏等。尤其在包括中国在内的一些发展中国家和经济不发达的国家或地区,这些问题显得尤为严重。因此,人口问题已成为全球的突出问题,是当代人类面临的重大挑战之一。

北京作为中国的首都,是全国的政治、文化、教育和对外交往中心,集中了中央部委、党政机关等大量政府机构和各类大中型企业的总部,经济发达,医疗条件、教育环境、社会保障体系、公共交通系统等社会资源丰富,凭借这些独特的政治、经济和社会条件,北京市对全国乃至全世界的人口流动都有着巨大的吸引力,人口问题更加严重。据 2010 年第六次全国人口普查数据显示:北京市登记常住人口为 1961.2 万人;与 2000 年相比,10 年间增长 604.3 万人,提前 10 年突破北京市 2020 年人口控制规模“1800 万”的红线,人口呈现急剧增长的态势。而且,随着我国人口总规模的不断增长和城市化进程的加快,北京市人口在未来的几十年间还有愈加膨胀的趋势。我国人口发展研究战

略课题组发布的研究报告指出：“到 21 世纪中叶，我国人口峰值控制在 15 亿左右，之后，人口总量缓慢下降。”我国目前人口总数为 13.4 亿，距峰值尚有约 1.6 亿的差距，据此，我国人口总规模还会在一定时期内继续增长；此外，我国处于城市化进程快速发展期，中国社会科学院发布的《2012 年社会蓝皮书》显示，“2011 年中国城镇人口占总人口比重已经突破 50%”，现有研究表明，一国城市化率为 70% 时，标志其进入现代化，社会发展最为稳定^[1]。中国的城市化若从目前的 50% 提高到 70%，意味着 15 亿人中将会有 11 亿人生活在城市。目前，中国城市人口仅为 6.7 亿，与 11 亿之间存在约 4.3 亿的缺口，人口继续向城市迁徙是大势所趋。在这种背景下，作为中国首都的北京，由于其在政治、经济、文化等领域所处的中心地位，对全国人口必然会形成较强和持续的吸引力，北京市人口规模在未来一段时期内依然会呈现持续增长的趋势。

北京市人口的急剧增长，给北京市现有水、电、煤气、住房、交通等资源的供给造成极大压力，引发了北京市能源外部依赖性增强、土地紧张、交通拥堵、基础设施超负荷运转、垃圾处理困难、城市管理难度增大等一系列相关问题。外来人口大量涌入的同时，北京的环境资源承载能力显得捉襟见肘。有统计数据表明，北京人均土地面积不足全国平均水平的 1/6，人均水资源占有量不足全国平均水平的 1/10、世界的 1/35；目前北京 100% 的天然气、100% 的石油、95% 的煤炭、64% 的电力、55% 的成品油均需从外地调入^[2]。中共北京市委十一届三次全会上，北京市市委书记郭金龙提到，人口无序过快增长、大气污染、交通拥堵、部分地区环境脏乱、违法建设等问题，已严重影响到北京的可持续发展^[3]。北京市市长王安顺则表示，坚决控制人口过快增长是解决北京交通、环境等许多问题的关键。而北京市政府副秘书长、市政府研究室主任周立云则强调，北京要治理“城市病”，必须首先控制人口，同时要保证区域协调发展^[4]。面对北京市人口膨胀所引发的的问题，北京市“十二五”规划纲要明确提出：“遏制人口快速增长，逐步构建适合北京发展的人口格局。”

因此，如何在全国人口规模不断增长和城市化进程不断加快的大背景下合理、有效地控制北京市人口增长速度，使之与北京市可利用资源的供给相匹配，减缓北京市能源、交通、土地、环境、垃圾处理和城市管理等方方面面的压力，成为北京市亟待解决的重点和难点问题之一。

1.1.2 研究意义

如何综合运用智能化数据挖掘技术和知识挖掘技术对影响预测精度的数量化因素和定性类知识性因素进行复合挖掘,并经恰当的预处理后代入预测模型,提高预测精度,是预测理论发展到今天所要解决的前沿问题和难点问题。我们针对北京市人口膨胀问题,利用复合知识挖掘技术对影响北京市人口膨胀的结构化的数据因素和非结构化的文本类因素进行系统深入的复合挖掘,建立基于复合知识挖掘的人口预测模型,并利用建立的模型对北京市人口未来增长趋势进行预测。希望我们的研究可以突破性地提高北京市人口增长预测精度,为人口预测领域提供新的思路和方法,为人口研究的发展做出贡献,同时,也为北京市有效调控人口规模、合理调动和规划资源供给、缓解人口膨胀带来的资源压力提供重要的参考依据。

要想合理、有效地控制北京市人口增长速度,首先要对促进北京市人口增长的各种因素进行全面和深入的挖掘,只有挖掘到那些对北京市人口膨胀有重要影响的因素,才能做到有的放矢,针对影响因素进行调控,最终实现北京市“十二五”规划纲要中的“遏制人口快速增长”的目的。影响北京市人口增长的因素很多,很复杂,其中有许多是定性的知识性文本类因素,比如城市化进程、经济增长速度、北京人口政策、环境因素,等等。恰恰是这类知识性文本类因素对北京市人口增长起到重要影响作用,如果在预测中忽略这些因素,无论你用何种预测模型都得不到好的预测结果。而预测理论发展到今天,面临的重大难题就是如何处理这类知识性文本因素。为了解决上述问题,我们研究运用智能化的知识挖掘技术,通过对影响北京市人口增长的知识性文本类因素进行知识发现、知识分类、知识清洗、知识提取、知识预处理等工作,挖掘出对北京市人口膨胀具有重要影响的知识性因素,希望能为北京市人口的有效调控提供重要参考。

要想破解北京市由于人口增长而引发的一系列相关问题,就必须对北京市人口增长规模做出准确预测。只有这样,才能对北京市未来资源的供给与需求做出科学、合理的调度与规划,才能据此为北京市未来的城市发展规划提供合理、真实的依据,缓解由于人口增长带来的资源及城市管理等方面的压力,为北京市安全、正常、健康运转提供保障。为了提高对北京市人口预测的精确度,我们研究构建智能优化神经网络预测模型对北京市人口增长进行预测,通

过对神经网络进行各种算法的优化、比较、选择,提高神经网络的运算速度、寻优能力和预测精度;通过在神经网络中嵌入智能决策树技术,解决知识性文本类影响因素的输入和处理问题,从而创建一个基于复合知识挖掘的智能优化神经网络人口增长预测模型。希望通过运用该模型,极大地提高北京市人口预测精度,为北京市制订资源供给和城市发展规划提供重要的参考依据。

智能化信息技术的不断发展,为我们下一步的攻关与研究提供了支持,我们这些新的预测方法的研究,也可以为其他领域在复杂多类型知识性影响因素条件下的预测工作提供借鉴与参考。

1.2 国内外研究现状

目前,国内外研究人口问题的预测方法主要分为两大类,一类是传统的统计类预测方法,主要包括 Malthus 人口增长模型、Logistic 预测模型、Keyfitz 模型、回归预测模型、Leslie 模型等;另一类是创新的智能化的预测方法,包括灰色预测方法、神经网络预测方法等。我国的汪寿阳、张世英、唐小我、邓聚龙等教授以及一批管理科学的专家学者,为我国的预测理论发展做出了突出贡献。

传统的统计类预测方法的最大局限,就是难以处理大量的、非线性的影响因素,难以处理定性类知识性影响因素,导致在对知识环境中复杂影响因素作用下的预测目标进行预测时,预测精度难以令人满意。Malthus 模型假设人口按指数规律无限增长,参见文献 [5] — [8]; Logistic 模型考虑了自然资源、环境等因素对人口增长产生的阻滞,认为人口增长呈 Logistic 函数形态,是被应用较多的人口预测模型之一,参见文献 [9] — [12]; Keyfitz 模型属于双曲线增长模型,可以被看成是 Logistic 模型的极端形式,参见文献 [13] — [14]。上述三种预测模型都属于基于确定性的数学模型,无法利用其对受复杂的、非确定性因素影响的预测目标进行预测。Leslie 模型属于随机预测模型,它是在对历史数据考察的基础上,拟合出带有随机性的平稳时间序列,进而对未来的各种变量进行估计。该模型尽管在一定程度上考虑到了非确定性因素,但仍然无法处理大量、复杂的非线性影响因素,参见文献 [15] — [17]。国内外学者也曾尝试用各类回归模型对人口增长进行预测,参见文献

[18]—[22],但回归类模型与上述模型一样,依然解决不了预测中的复杂的非线性影响因素和定性类知识性因素的问题。

创新的智能化预测方法,包括灰色预测方法和神经网络预测方法等,由于具有一定的智能性,可以在一定程度上解决非线性影响因素的问题,但在预测中也存在着不足。灰色预测方法仅适用于小样本、波动小的非线性数据,对于指数增长形状的历史数据曲线具有较好的预测效果,对于灰度强化、白度弱化的历史数据序列预测效果容易出现偏差,参见文献[23]—[26]。普通神经网络人口增长预测模型具有可考虑数据型影响因素、无须识别变化规律、可模拟任意非线性复杂映射、可处理大量复杂的非线性历史数据的一定智能性等优点,但是仍不能处理定性知识因素的影响,模型本身的优化也存在需要研究的问题,参见文献[27]—[31]。

以上研究方法及研究工作者为人口增长预测工作奠定了基础,在目前的人口预测工作中发挥了重要作用。但是,随着我国经济不断发展和城市化进程不断加快,大城市人口膨胀问题严重,在影响这些大城市人口增长的因素中,很多是非数量化的、定性类的知识性因素。由于以往的相关预测中很少考虑这些重要影响因素,导致预测精度不高。因此,如何将这类知识性影响因素代入预测模型,进行数量化影响因素和非数量化知识性因素的复合知识挖掘,是提高人口预测精度的关键技术,因此成为人口预测领域亟待解决和必须深入研究的问题。

知识挖掘是在数据库知识发现基础上发展起来的,其定义是由 Usama M. Feldman 在 1995 年首次给出的:从知识资料集中识别有效的、新颖的、潜在有用的,以及最终可理解的模式的高级处理过程,参见文献[32]。其目的是将大量的数据和知识融合成有序的、分层次的、易于理解的信息,并进一步转化成可用于干预预测和决策的知识,是一个智能化、自动化的过程,是统计学、管理科学、计算机科学、模式识别、人工智能、机器学习与其他学科相结合的产物。其知识资料集包含显性与隐性、定性与定量、经验与推理等类型知识,能对原有的数据、知识进行高度自动分析和归纳推理,从中挖掘出潜在的模式和规律,为预测决策、模式识别、故障诊断、生产过程优化等的实现提供知识服务。

国际上,1995年在加拿大蒙特利尔召开了第一届知识发现和数据挖掘国际会议,引起了学术界和工程界的极大关注,由此开始了数据挖掘的发展。而

知识挖掘理论的研究,近十年来才开始具备研究条件并逐步得到重视,被认为是寻找内在规律的新方向。我国的王众托、黄梯云等教授是知识挖掘研究领域的开拓者。

文献[33]中给出了知识工程的建设框架,文献[34]中 Wendi Bukowitz 给出了知识管理过程的框架,文献[35]中给出了智能知识管理模型框架,文献[36]对基于知识管理的决策支持系统进行了研究,文献[37]在聚类分析及 Apriori 关联规则分析的基础之上,建立了一种复合知识挖掘方法。文献[38]将数据挖掘和知识挖掘技术结合在一起,建立了一个以分类算法为基础的信息融合系统,并采用一组高风险状态下的样本数据进行分析,发现该模型能有效地侦测出现实生活中的欺诈事件。文献[39]描述了一个名为“文本域识别和关联挖掘”(DATAM)的决策支持系统,该系统将基于知识挖掘的关联分析和本体结合在一起,能够精确地检测到样本数据集中的异常点,为改善汽车行业的售后服务提供决策支持。文献[40]将文本分析工具和一种名为“Arizona”金融文本的预测系统结合起来,对金融新闻中的作者的措辞、语气和股价变化之间的关系进行了分析,研究表明,金融新闻中作者的措辞和语气会在一定程度上对股价造成影响。文献[41]以新浪体育论坛中 31 个不同的论坛和 220 053 个公告中的文本信息为样本,利用知识挖掘技术将其归类到各个集合并赋予适当的数值,每个集合都代表当前的热点话题,以此来发掘当年最受热议的话题,并对本年度的热议话题进行预测。文献[42]提出了一种可以自动分析金融文本信息的知识挖掘方法,这种方法可以根据金融文本信息自动侦测金融事件,利用该方法对诈骗和破产两种金融事件进行实证分析后,发现模型具有很高的精确性。文献[43]建立了一种以知识挖掘为基础并结合了粗糙集理论和支持向量机的模型,这种模型能通过优化、特征选择和掌握台湾股市非系统性的新闻,进而准确地预测新闻发布一小时后股价的浮动趋势。文献[44]利用知识挖掘来预测纳米技术在国内的发展趋势,进而帮助政府制订对纳米技术产业的投资政策,对该产业在国内的发展现状、潜在风险进行分析,并预测其未来在国际上的竞争优势。文献[45]—[57]是我国学者在不同领域对知识管理和知识挖掘展开的一系列研究。

根据北京市人口增长预测的需要和特点,我们提出“复合知识挖掘”的概念:复合知识挖掘是指针对预测目标的各项知识特征,通过对数据库中的历史数据及对应的各类影响因素的文本分库、推理规则分库、经验知识分库等进

行同时、综合、协调的知识挖掘,用知识类文本因素动态修正数据类因素预测中产生的误差,自动提取具有高度相似性复合知识特征的同类历史数据,克服以往的预测方法中存在的不足,使预测精度得到突破性的提高。

这种“复合知识挖掘”和普通知识挖掘相比有两点不同,第一,挖掘资料集由单一的知识库改变为同时挖掘各类数据库、文本分库、经验知识分库和推理规则分库;第二,挖掘的方法由对单一类型知识的检索、搜寻、提取改变为针对各类知识按照总体关联程度进行的综合的、协调的、同时的知识挖掘及发现。

目前,在国内外尚未见到将复合知识挖掘技术与人口预测相结合进行的研究工作及其成果。我们希望在优化神经网络的基础上,运用复合知识挖掘技术,建立基于复合知识挖掘的智能优化神经网络人口预测模型,对影响北京市人口增长的复杂的非线性的数量化因素和定性类知识性因素进行全面综合考虑,克服以往预测研究中对知识性因素考虑不足的缺陷,对北京市人口膨胀趋势进行预测,希望能较大幅度地提高预测精度。

由于我们在预测问题的研究方面具有一定的基础,经过学习和参考国内外的研究工作,认为有必要将基于复合知识挖掘的人口增长智能预测的一系列新的理论方法应用到北京市人口增长趋势的预测中,希望在北京自然科学基金的资助下深入研究,为人口预测理论,为北京市人口预测、人口调控做出贡献。

1.3 研究内容

本书主要研究了人口预测模型的理论方法,建立了基于复合知识挖掘的北京市人口膨胀趋势预测模型,并对北京市未来人口增长趋势进行预测。本书对北京市目前常住人口发展的现状进行了总结,介绍了北京市人口增长的特点,分析了影响北京市人口增长变化的主要因素;介绍了目前人口预测的理论方法,包括传统的基于统计学原理的预测方法和创新类的智能化的预测方法,并对两类预测方法进行了比较分析,指出各自的优缺点和适用范围,并在此基础上总结出人口预测模型应具备的功能与特点;对我国目前人口预测的现状进行了分析,并指出其存在的问题及相应的解决对策;介绍神经网络的各种群智优化算法,并运用经过遗传算法(GA)、粒子群算法(PSO)和微分进化算法

(DE) 优化了的 RBF 神经网络模型以及未经优化的 RBF 神经网络模型对北京市人口增长进行了预测, 选出最优的预测模型; 建立影响北京市人口数量增长的数量化数据因素对象集, 利用包括相关分析、格兰杰因果检验和预测分析等在内的多种数据挖掘技术, 对这些因素进行分层筛选、组合优化, 挖掘出影响北京市人口增长的最主要的数量化数据因素, 为进一步提高预测模型的准确度打下基础; 建立影响北京市人口数量增长的文本类因素对象集, 运用文本挖掘技术, 通过关联规则算法, 挖掘出对北京市人口增长有重大影响的文本类因素, 从而建立北京市人口膨胀影响因素的复合知识库; 在此基础上, 将神经网络模型与决策树技术进行有机结合, 利用数据挖掘和文本挖掘获得的数据, 建立基于复合知识挖掘的北京市人口膨胀趋势预测模型; 利用基于复合知识挖掘的北京市人口膨胀趋势预测模型, 通过动态情景分析法, 对北京市中长期人口增长趋势进行动态预测; 此外, 将 REP - Tree 技术与人口预测领域最经典的 Logistic 模型有机结合, 建立 REPTree - Logistic 人口预测模型, 并利用这种模型对北京市中长期人口增长情况进行动态预测; 在此基础上, 将基于复合知识挖掘的神经网络人口预测模型与 REPTree - Logistic 人口预测模型的预测结果进行对比分析, 实现对北京市未来人口增长情况的准确把握; 最后, 基于预测结果, 对北京市人口管理与调控提出合理的建议, 以期为决策管理机构合理控制人口、有效利用资源、科学有序地进行城市规划及环境保护等提供有价值的决策依据。

(1) 研究创建北京市人口膨胀影响因素的数量化因素和文本类因素对象集。综合运用数据挖掘技术和知识挖掘技术, 对促进北京市人口膨胀的数量化因素和非数量化的知识性因素进行全面、深入挖掘, 构建一个包含数据库和知识库在内的北京市人口膨胀影响因素对象集。

(2) 在构建北京市人口膨胀影响因素对象集过程中, 研究创造各类数量化影响因素对象集及推理规则库和经验知识库。影响北京市人口膨胀的知识性因素主要包括城市化进程、全国人口规模、经济增长速度、人口与户籍政策等, 可据此建立相应的文本库, 例如城市化进程文本库包括“城市化进程较快、城市化进程较慢、城市化进程停止”等描述性知识, 为知识发现、识别和提取做好准备。同时运用人工智能和关联规则建立推理规则库, 例如“城市化进程较快必然导致农村人口快速涌入城镇, 必然导致北京市人口快速增加”等。此外, 通过对人口的认识的采集、归纳比较及知识挖掘, 建立预测

经验知识库,为预测的预处理分析和后干预检验提供依据。

(3) 研究基于复合知识挖掘的建模预处理技术。在对数据型知识进行数据挖掘和对定性类知识进行知识挖掘的基础上,将两类影响因素分别进行关联度分析,并按重要程度进行排序,分别获取前 n 个重要影响因素。利用 REPTree 决策树对知识性因素进行 If-Then 规则分析,用分析结果动态调整输入神经网络预测模型中的数量化因素的预测误差,得到含有复合知识的、与北京市人口增长具有高度相似复合知识特征的建模所需的一系列历史数据,以此为建模基础,结合预测模型方法上的创新,可突破性提高预测精度。

(4) 研究神经网络预测模型的优化技术。传统的统计类预测模型无法处理与预测目标之间存在复杂非线性关系的数量化因素。而神经网络预测模型具有无须识别变化形状和规律、可模拟任意非线性复杂映射、可处理历史数据和定量影响因素作为输入变量、具有经学习训练得到最终模型的一定智能性的优点,用其作为北京市人口增长预测模型是比较合适的。但是普通神经网络预测模型也有一定的缺陷,例如易陷入局部最优、导致预测精度不高、学习与训练速度降低等。为了克服神经网络的上述不足,尝试用 PSO、DE、GA 等生物进化算法对神经网络进行优化,提高神经网络的全局寻优能力和学习训练速度,并将各种算法加以比较、筛选和改进,选择和构建预测误差最小的优化神经网络预测模型。

(5) 研究构建基于复合知识挖掘的智能优化神经网络人口预测模型。神经网络预测模型没有充分考虑知识性文本类影响因素对预测建模的作用,不适合大量知识性文本类影响因素的预处理和建模。针对神经网络上述缺陷,进一步研究在优化神经网络预测模型的基础上,嵌入智能化 REPTree 决策树分类器,解决知识性文本类影响因素的输入问题,解决知识性因素对数量化因素的预测误差进行动态调整的技术问题。运用模型进行仿真实验,不断修正模型,提高预测精度。

(6) 优化 Logistic 人口预测模型,建立 REPTree-Logistic 人口预测模型。针对人口预测领域最经典的 Logistic 模型存在的只能描述长期人口总体增长变化趋势、对具体某个时间的人口数量预测精度不高,以及不同因素对人口增长的影响无法体现的问题,将 REPTree 决策树文本挖掘技术与 Logistic 模型结合起来,建立了同时包含数据因素和文本因素的 REPTree-Logistic 人口预测模型,从而进一步优化了传统的 Logistic 人口预测模型。

(7) 在上述研究的基础上,运用基于复合知识挖掘的智能优化神经网络人口预测模型和 REPTree - Logistic 模型对北京市人口增长趋势做中长期预测,对预测结果和影响因素进行分析并提出政策建议,为北京市有效调控人口规模、合理调动和规划资源供给、合理进行城市规划等提供参考和借鉴。

1.4 研究方法

本书综合应用人口预测理论、神经网络和群智算法理论与技术、数据挖掘和文本挖掘技术与方法,以及现代数学方法与计算机科学技术,对人口预测理论及方法进行了全面系统的研究。

(1) 查阅国内外相关研究资料和成果,学习、分析和改进前沿的智能信息处理技术,如数据挖掘、知识挖掘、神经网络、模糊理论、决策树技术等智能技术,研究人口预测方法与智能技术结合的途径与技术。

(2) 调研促进北京市人口膨胀的历史数据和影响因素,进行数据采集,建立人口增长影响因素数据库;研究基于文本关联规则分析的文本知识的分类与挖掘,建立便于机器学习的各类知识的文本分库。

(3) 研究利用经验推理、决策树技术、关联规则等对知识性文本因素进行分析挖掘,建立推理规则库;通过调研、类比、归纳,建立经验知识库。

(4) 对数据库中的历史数据和知识库中的知识因素进行关联分析和排序,分别挖掘出与预测目标具有高度相似特征的前 n 个数量化影响因素和知识性影响因素。

(5) 将知识性因素进行 If - Then 规则分析,用 If - Then 规则分析结果动态调整数量性因素在预测中产生的误差,创造出包含复合知识的新的历史数据,形成基于复合知识挖掘的新的预处理技术。

(6) 运用各种优化算法对神经网络进行优化,从中选择出预测效果最好的优化神经网络。

(7) 将 REP - Tree 决策树嵌入优化神经网络预测模型,解决知识性文本类影响因素的输入及对数量化影响因素的动态调整问题,实现基于复合知识挖掘的最终目标。

(8) 建立 Logistic 模型,并利用 REP - Tree 决策树对传统的 Logistic 模型

进行优化,建立既包含数量化因素、又包含文本类因素的基于复合知识挖掘的 REPTree - Logistic 人口预测模型。

(9) 依据上述理论方法,进行仿真实验和案例验证,不断优化和改进我们提出的“基于复合知识挖掘的智能优化神经网络人口预测模型”,不断修正模型,扩大其适用性,使预测精度突破性提高。

(10) 运用基于复合知识挖掘的智能优化神经网络预测模型、REPTree - Logistic 人口预测模型和动态情景分析法对北京市人口增长趋势做中长期预测,在对预测结果和影响因素进行分析的基础上,对北京市人口增长管理与调控提供政策建议。

1.5 创新点

本书首次提出“基于复合知识挖掘的人口预测”的概念,综合运用多种算法和软件工具,将数据挖掘和文本挖掘技术同时应用到人口预测领域中,极大地提高了人口预测精度,为人口预测领域的研究提供新的思路和方法。

(1) 创建北京市人口膨胀影响因素“复合知识库”。针对北京市人口增长做预测,如果忽略那些对人口膨胀具有重要影响的知识性因素,无论你使用多么先进的预测模型,都不会有好的预测效果。据此,我们提出“复合知识库”的概念,突破传统的人口预测中只依赖数据库建模的思维,将知识性因素在北京市人口预测中重点加以考虑,具有创新性。

(2) 提出一种基于复合知识挖掘的人口预测建模预处理新方法。在数据挖掘和知识挖掘的基础上,运用 REP - Tree 智能决策树技术对知识性因素进行 If - Then 规则分析,不断用 If - Then 规则分析结果动态调整数量化因素预测中产生的误差,创造出一种包含复合知识的历史数据,从而形成一种新的预测建模预处理方法。

(3) 构建基于复合知识挖掘的智能优化神经网络人口预测模型。通过运用生物进化等优化算法对神经网络进行优化和改进,创建出优化神经网络人口预测模型;同时研究用智能化决策树技术来解决知识性文本因素的输入和对数量因素的动态调整的技术问题,从而创造性地构建了一个能同时处理定量和定性影响因素的智能优化神经网络人口增长预测模型,期望能突破性提高预测精

度，为北京市人口增长趋势做出准确预测。

(4) 构建基于复合知识挖掘的 REPTree - Logistic 人口预测模型。对人口预测领域最经典的 Logistic 模型进行改进，建立既包含数量化因素又包含文本类因素的基于复合知识挖掘的 REPTree - Logistic 人口预测模型，从而解决了传统的 Logistic 模型存在的只能描述长期的人口总体增长变化趋势、对具体某个时间的人口数量预测精度不高以及不同因素对人口增长的影响无法体现的问题。

(5) 将情景分析法运用到人口预测中，通过设定不同的情景，实现对北京市未来人口增长情况的动态预测。利用情景分析法对北京市未来人口增长趋势进行预测，并不会得到一个“唯一的、准确的”结果，而是通过探究北京市人口未来发展趋势的多种可能途径，检查可能的选择，为未来决策提供依据和框架。相比传统预测方法对人口单一的预测结果，基于情景分析法的人口预测是一种开放式的预测，可以提供多种可能的结果及其路径，可以让决策者更好地掌控风险，在环境发生变化时可以更加迅速而灵活地做出反应。

第2章 北京市人口增长的特点及影响因素分析

人口问题是一切发展的核心问题，许多矛盾和问题都与人口问题密不可分。随着经济社会的发展和人们生活水平的提高，北京市常住人口出现了新的变化趋势。本章首先从分析北京市常住人口的现状入手，对北京市人口增长的特点进行总结，在此基础上，从多个角度对影响北京市人口增长的原因进行分析。

2.1 北京市常住人口现状分析

常住人口指实际经常居住在某地区一定时间（半年以上，含半年）的人口。北京市常住人口的增长主要由户籍常住人口的增长和非户籍常住人口的增长两部分构成。其中，非户籍常住人口是指在京居住半年以上的外来人口，也就是通常说的流动人口。在无特别说明的情况下，本文所说的北京市人口均指北京市常住人口。

2.1.1 常住人口的增长趋势

改革开放以后，北京市常住人口的增长经历了5个发展阶段：1979—1994年为缓慢增长的阶段；1995—2000年为稳步增长的阶段；从2000年起，常住人口大幅度增长；而2008年北京奥运会过后，常住人口又呈现飞速增长的状态；2010年以后，常住人口增速开始放缓（参见图2-1）。

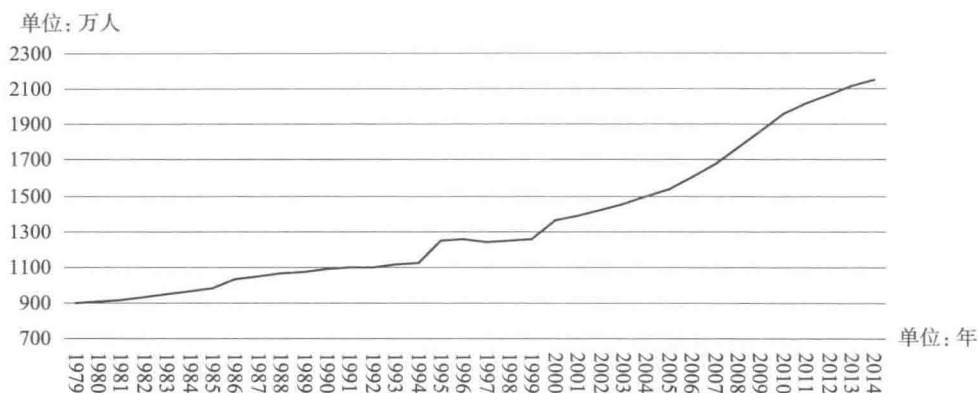


图 2-1 北京市常住人口增长趋势 (1979—2014 年)

据北京市统计局、国家统计局北京调查总队联合发布的数据显示^①：2014 年年末北京常住人口 2151.6 万人，比上年年末增加 36.8 万人，增速为 1.7%，但与上年相比，增量减少 8.7 万人，增速下降 0.5 个百分点。其中，北京常住外来人口为 818.7 万人，比上年年末增加 16 万人，增长 2%，但增量比上年减少 12.9 万人，增速比上年回落 1.7 个百分点。

由于 2008 年奥运会以后北京市人口迅速增长，人口问题明显加剧，所以从 2010 年起，北京市人口政策开始趋向严格。因此，2011 年以来，北京市常住人口快速增长趋势得到有效遏制，增量和增速逐步放缓，常住人口增量从 2011 年的 56.7 万人降至 2014 年的 36.8 万人，增速从 2011 年的 2.9% 降至 2014 年的 1.7%，均达到“十二五”以来最低水平。

2.1.2 常住人口的地区分布状况

北京市常住人口地区分布可以从环路、行政区划、人口密度和从业人员 4 个方面加以分析^[58-60]。

(1) 从环路上看。

环路人口分布呈圈层向外拓展，即由二、三环内向四环外聚集。2014 年人口抽样调查结果显示，三环至六环间聚集了 1228.4 万人的常住人口，占北

① 关于本文引用数据的说明：1. 在无特殊说明的情况下，本文所使用数据均来源于北京市统计年鉴及全国统计年鉴；2. 考虑到数据的有效性及可获得性，本文使用各项数据默认起止年份为 1979—2013 年，部分数据更新至 2014 年。

北京市总人口的 57.1%；四环至六环间聚集了 941 万人，占北京市总人口的 43.8%；五环以外有 1098 万人，占北京市总人口的 51.1%。

常住外来人口与常住人口在环路分布情况基本一致，且向外拓展聚集的特点更加突出。三环至六环间聚集了 637.6 万的常住外来人口，占北京市常住外来人口的 77.9%；四环至六环间聚集了 532.1 万人，占北京市常住外来人口的 65%；五环以外有 422.5 万人，占北京市常住外来人口的 51.6%。

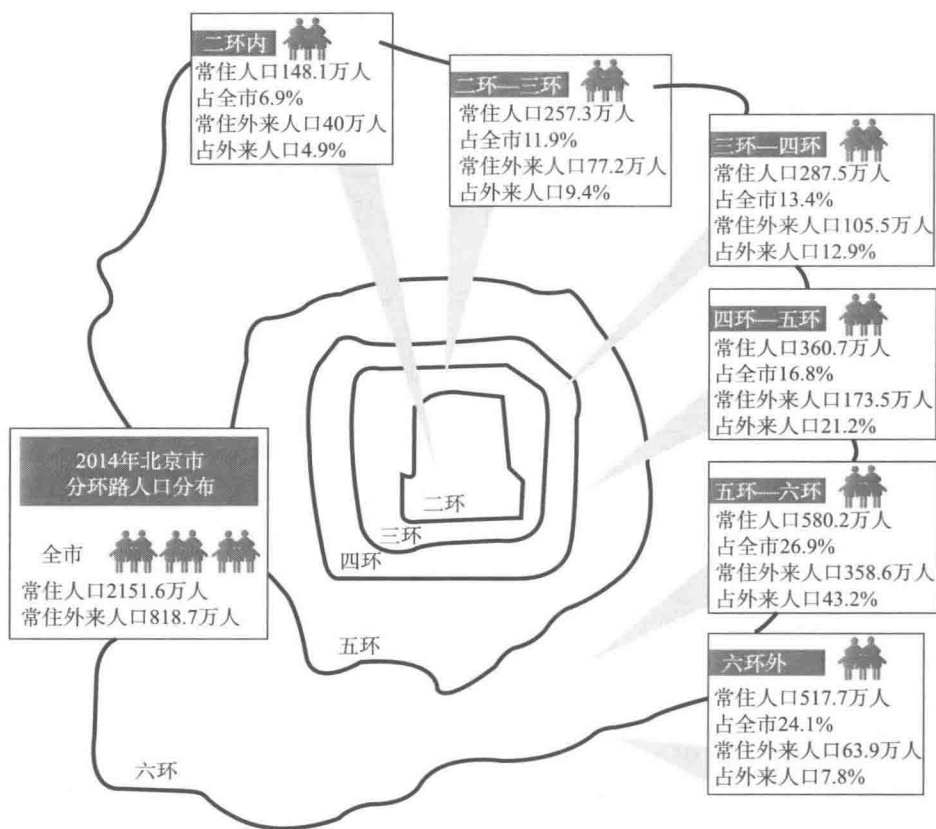


图 2-2 北京市常住人口环路分布图

(2) 从行政区划上看。

根据北京市土地利用总体规划，北京城市功能区划分为首都功能核心区、城市功能拓展区、城市发展新区和生态发展涵养区 4 个功能区。其中，首都功能核心区包括东城区、西城区；城市功能拓展区包括朝阳区、丰台区、石景山区、海淀区；城市发展新区包括房山区、通州区、顺义区、昌平区、大兴区；

生态发展涵养区包括门头沟区、怀柔区、平谷区、密云区、延庆区。

首先,从4个功能区看,截至2014年年末,城市功能拓展区常住人口最多,达到1055万人,占49%;其次是城市发展新区,常住人口为684.9万人,占31.8%;首都功能核心区和生态涵养发展区常住人口相对较少,分别为221.3万人和190.4万人,所占比重分别为10.3%和8.9%。城市功能拓展区常住外来人口最多,达到436.4万人,占北京市常住外来人口的53.3%;其次是城市发展新区,常住外来人口为296.9万人,占36.3%;而首都功能核心区和生态涵养发展区常住外来人口相对较少,分别为54万人和31.4万人,所占比重分别为6.6%和3.8%。

表2-1 按区域分北京市人口分布情况(2013—2014年)

地区	2014年常住人口 (万人)	2013年常住人口 (万人)	同比增长 (%)	2014年常住 外来人口 (万人)	2013年常住 外来人口 (万人)	同比增长 (%)
全市	2151.6	2114.8	1.7	818.7	802.7	2
首都功能核心区	221.3	221.2	0	54	55.4	-2.5
东城区	91.1	90.9	0.2	21.2	21	1
西城区	130.2	130.3	-0.1	32.8	34.4	-4.7
城市功能拓展区	1055	1032.2	2.2	436.4	426	2.4
朝阳区	392.2	384.1	2.1	179.8	176.1	2.1
丰台区	230	226.1	1.7	85.1	85	0.1
石景山区	65	64.4	0.9	21.2	21.4	-0.9
海淀区	367.8	357.6	2.9	150.3	143.5	4.7
城市发展新区	684.9	671.5	2	296.9	289.6	2.5
房山区	103.6	101	2.6	26.7	24.6	8.5
通州区	135.6	132.6	2.3	55.5	53.6	3.5
顺义区	100.4	98.3	2.1	38.9	37.3	4.3
昌平区	190.8	188.9	1	100.2	100.6	-0.4
大兴区	154.5	150.7	2.5	75.6	73.5	2.9
生态涵养发展区	190.4	189.9	0.3	31.4	31.7	-0.9
门头沟区	30.6	30.3	1	4.9	5	-2
怀柔区	38.1	38.2	-0.3	10.4	10.6	-1.9
平谷区	42.3	42.2	0.2	5.3	5.3	0
密云县	47.8	47.6	0.4	7.2	7.2	0
延庆县	31.6	31.6	0	3.6	3.6	0

其次,从区县看,常住人口数量位居前3位的区县是朝阳区、海淀区和丰台区,3个区人口占北京市人口总量的46%。常住外来人口位居前3位的是朝阳区、海淀区和昌平区,3个区外来人口占北京市外来人口总量的52.6%。门头沟区、平谷区、密云县和延庆县4个区县的常住外来人口均不足10万,合占比重仅为2.6%。

(3) 从人口密度上看。

随着人口数量的不断增长,人口密度也逐步上升。2014年,北京市常住人口密度为1311人/平方公里,比2011年增加81人/平方公里(2011年为1230人/平方公里)。西城区常住人口密度最高,为25767人/平方公里,而延庆县最低,只有158人/平方公里。常住人口密度从首都功能核心区开始向外围逐渐降低。核心区人口密度为23953人/平方公里,是城市功能拓展区的2.9倍,是城市发展新区的22倍,是生态涵养发展区的109.9倍。

(4) 从从业人员上看。

首先,从产业情况看,截至2013年年末,北京市第二产业和第三产业法人单位共有从业人员1111.3万人,比2008年年末增加294.5万人,增长36.1%。其中,第二产业从业人员216.6万人,增加15.4万人,增长7.6%;第三产业从业人员894.7万人,增加279.1万人,增长45.3%。在法人单位从业人员中,位居前5位的是:批发和零售业147.8万人,占13.3%;制造业138.5万人,占12.5%;租赁和商务服务业141.7万人,占12.8%;信息传输、软件和信息技术服务业93.0万人,占8.4%;科学研究和技术服务业95.2万人,占8.6%。

其次,第二、第三产业法人单位的从业人员主要集中在城市功能拓展区和城市发展新区。城市功能拓展区595.6万人,占53.6%;城市发展新区246.9万人,占22.2%;首都功能核心区202.3万人,占18.2%;生态涵养发展区66.2万人,占6.0%。

2.1.3 常住人口的年龄构成状况

2014年北京市常住人口中,0~14岁的少儿人口为213.0万人,占总人口数量的9.9%;15~59岁的劳动年龄人口为1617万人,占75.2%;60岁及以上老年人口为321.6万人,占14.9%,其中,65岁及以上人口为212.3万人,占9.9%。从以上结果可以看出,北京市已经开始面临严重的人口老龄化问题,未来这一问题将更加严峻。