



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

基于多元主体协同的算法治理研究——以智慧派单算法治理方案为例

上海交通大学

SHANGHAI JIAO TONG UNIVERSITY

学士学位论文

BACHELOR'S THESIS



论文题目：基于多元主体协同的算法治理研究

——以智慧派单算法治理方案为例

学生姓名：李瑞琪

学生学号：515130910024

专 业：行政管理

指导教师：樊博

学院(系)：国际与公共事务学院

上海交通大学

本科生毕业设计（论文）任务书

课题名称： 基于多元主体协同的算法治理研究
——以智慧派单算法治理方案为例

执行时间： 2018 年 12 月 至 2019 年 6 月

教师姓名： 樊博 职称： 教授

学生姓名： 李瑞琪 学号： 515130910024

专业名称： 行政管理

学院(系)： 国际与公共事务学院

毕业设计（论文）基本内容和要求：

- 1、 数据质量分析——通过对 6903 条线上爬取的微博、贴吧、新闻网等评论进行文本挖掘，测量情感强度、分析评论议程；
- 2、 智慧派单算法设计与分析；
- 3、 社会治理对策

毕业设计（论文）进度安排：

序号	毕业设计（论文）各阶段内容	时间安排	备注
1	数据整理	2019.1-2019.2	
2	算法分析	2019.2-2019.3	
3	实证分析	2019.3-2019.4	
4	论文撰写	2019.4-2019.5	

课题信息：

课题性质：设计 论文

课题来源*：国家级 省部级 校级 横向 预研

项目编号 14ZDB152

其他_____

指导教师签名：樊博

2018年12月21日

学院（系）意见：

院长（系主任）签名：郭俊华

2018年12月22日

学生签名：李瑞琪

2018年12月21日

上海交通大学

毕业设计（论文）学术诚信声明

本人郑重声明：所呈交的毕业设计（论文），是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

作者签名：

日期： 年 月 日

上海交通大学

毕业设计（论文）版权使用授权书

本毕业设计（论文）作者同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权上海交通大学可以将本毕业设计（论文）的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本毕业设计（论文）。

保密 ，在 ___ 年解密后适用本授权书。

本论文属于

不保密 。

（请在以上方框内打“√”）

作者签名：

指导教师签名：

日期： 年 月 日

日期： 年 月 日

基于多元主体协同的算法治理研究

——以智慧派单算法治理方案为例

摘要

随着第三次人工智能数字化浪潮的来临,算法已日益成为影响社会运行与政府决策的基础性法则,算法的效能提升和应用能力普及无疑极大地造福了民众,但与此同时,新的治理挑战也随之而来。从治理目标角度来说,政府治理以社会福利最大化为目标;企业治理以所有利润最大化为导向;民众则以个体效益最大化为目的,因此在多元主体视角下,算法治理如何更好地权衡三方关系优化治理成为了亟待解决的治理难题。在目前的大数据背景下,由于算法的设计偏向以及未公开的“算法黑箱”等问题,城市生活中的人们在受益于算法服务的同时也默默地被算法影响与控制。本文根据多元协同治理理论,对当前智慧派单算法原理进行分析,主要的研究内容包括:第一,通过爬取 6903 条与“派单算法”相关的线上微博、贴吧评论数据,进行整理分析;第二,以知网 HowNet、台湾大学 NTUSD、清华褒贬义三个词典为基础,结合三者构建了一个全面的情感分类词典,对情感强度进行分析并根据词频提取主题词,以此来挖掘公民的态度或情感变化以及最关注的话题等;第三,运用评论数据中的各利益方的诉求特征总结出不同规律,如不同地域的民众对车辆选择有自己的偏好;第四,将各利益方关注的变量引入优化智慧派单算法,并提出基于算法的社会治理模式与传统治理模式的区别;最后,本文根据实证案例分析结果,对政府、企业、公众三方互相协作治理新生事物进行了探讨。

关键词: 算法治理, 多元主体, 派单算法, 情感强度分析

ALGORITHM GOVERNANCE RESEARCH BASED ON MULTI-AGENT COLLABORATION: A CASE STUDY ABOUT ORDER DISPATCH ALGORITHM

ABSTRACT

With the advent of the third wave of artificial intelligence digitalization, algorithm has increasingly become the basic law in the society, which affects social operation and government decision-making. The improvement of algorithm efficiency and the popularization of application ability have undoubtedly greatly benefited the public. However, at the same time, new challenges have also been created. From the perspective of governance objectives, government aims to maximize social welfare; enterprises oriented at gaining the maximum profit; citizens, nevertheless, aims at gaining their high efficiency. Therefore, from the perspective of multiple agents, how to balance the triple relationship and optimize the overall governance of algorithm has aroused a heated discussion. According to the current social background, due to the design bias of algorithm and the confidential algorithm black box, citizens have benefited from the algorithm services but also affected and controlled by algorithm silently. Based on the theory of multi-agent collaboration, this paper analyzes the principle of the current order dispatch algorithm. The main research contents include: first, crawling 6903 online weibo and other comment data related to “order dispatch”. Second, based on HowNet, NTUSD from Taiwan University and Tsing Dictionary, a comprehensive emotion classification dictionary is constructed, which helps analyze the words extracted from the comments, thus exploring the change of citizens’ attitudes or emotions and their most concerned topics. Third, the individual heterogeneity in the review data is used to summarize various rules. Fourth, the heterogeneous variables are then introduced into the optimized order dispatch algorithm and the differences between the social governance model based on the algorithm and the traditional governance model are proposed. Finally, based on the result of the empirical case analysis, this paper discusses the new issues of collaborative governance among the government, enterprises and the citizens.

Key words: algorithm governance, multiple agents, order dispatch algorithm, emotional intensity analysis

目 录

第一章 绪论.....	1
1.1 研究背景	1
1.2 研究意义	2
1.2.1 理论意义	2
1.2.2 实践意义	2
1.3 研究方法和思路	3
1.3.1 研究方法	3
1.3.2 思路和章节布局	3
第二章 文献综述.....	5
2.1 算法治理理论研究	5
2.2 多元协同治理理论研究	6
2.2.1 国内外多元协同治理理论	6
2.2.2 多元协同治理存在的问题	8
2.3 文献述评	8
第三章 理论构建概述.....	10
3.1 传统的多元协同治理模式	10
3.2 算法治理中多元协同模式建构	11
第四章 多元主体视角下的智慧派单算法应用情景构建.....	15
4.1 情感强度分析研究	15
4.2 当前智慧派单算法设计与分析	17
4.3 基于文本挖掘的智慧派单评论分析	20
4.3.1 评论数据采集与处理	20
4.3.2 情感强度分析	23
4.3.3 评论主题分析	25
第五章 结合各利益相关方的智慧派单算法治理方案.....	27
5.1 个体选择行为影响因素分析	27
5.1.1 乘客角度	27
5.1.2 司机角度	29
5.2 政府治理优化方案	30
5.3 基于多元协同的算法治理新模式	31
第六章 结论与展望.....	33
6.1 研究结论	33
6.2 不足与展望	33
参考文献.....	35
谢辞	39
附录一 评论数据爬取代码.....	40
附录二 情感强度分析代码.....	42
附录三 主题词词频分析代码.....	44
附录四 TF-IDF 算法运行部分结果	46
附录五 优化后的智慧派单算法伪代码.....	47

第一章 绪论

1.1 研究背景

我们每天都生活在一个“算法时代”中，计算机科学正蓬勃发展并影响、塑造和指导我们的行为和社会治理。伴随着数字化程度的显著提升，我们需要充分意识到算法治理的必要性；另一方面，基于对算法应用模式及基础原理的认知，我们应该对新业态和新科技带来的治理挑战形成客观、合理的评估^[1]，进而在此基础背景上有针对性地提出结合公共管理与算法治理的创新思路。如今，关于算法治理的演变、应用及影响等逐渐成为了学者关注并讨论的热点话题。

回溯算法治理历史，自韦伯时代以来，社会学家强调了国家的法律官僚组织治理模式将会与工业工厂设计相同的现代化趋势^[2]，首次提出了社会治理将会是一个类似“机器”的治理系统：任务将会被细分，角色会变得专属，以便尽可能有效地执行治理事务，内容包括收集有关该系统适用的社会和公民的数据等^[3]，从 20 世纪中期计算机时代开始，以程序代码为表现形式的“算法”开始辅助人们的日常生活和工作，人们已尝试使部分或全部过程自动化。例如，控制论中主张在社会治理中使用计算机化的数据收集、处理和决策系统^[4]。

近年来在国内，算法逐渐被广泛应用，较为常见的形式是结合大数据，通过打分、评级、排序等方式在电子政务或商务环境、客户及推荐对象间建立链接，为其自动化推荐，例如电子商务中的“猜你喜欢”，社交平台上的“你可能认识的好友”，搜索引擎中的内容或关键词推荐等。

然而，2017 年 9 月，《人民日报》发文三评算法推荐，直接指出了以“今日头条”为代表的单纯依靠算法推荐的平台的严重危机，它们普遍存在“价值观缺失”、“制造信息茧房”、“竞争手段屡破底线”的三大问题，并称技术的迭代与创新在便民惠民的同时，也容易踏入规则的“无人区”，甚至有可能“为所欲为”走向创新的反面^[5]。实际上，除了国内的“算法推荐”外，国外的一些算法技术也被大部分民众认为会包含偏见，早在 2015 年，美国芝加哥法院预防犯罪的风险评估算法 COMPAS 就被证明对黑人有系统的种族歧视：黑人被普遍认为更具高犯罪风险，从而往往会被法官判处更长的刑期^[6]。因此，算法的目的、设计与数据使用等问题都涉及相关利益与主观偏见，使得算法操控、算法过滤等情况频繁发生，导致整个算法的决策或预测水平与能力大幅下降。

在此背景下，解决算法带来的阻碍和麻烦已是不容小觑的话题，“由谁来治理”与“如何治理”成为了至关重要的问题，实际上，“治理”是当代“新公共管理”学科的称谓，

^[1] 贾开. 人工智能与算法治理研究. 中国行政管理, 2019(10):42-47.

^[2] Kanter, R. M. The Future of Bureaucracy and Hierarchy in Organizational Theory: A Report from the Field In: Bourdieu, P./Coleman, JS (Eds.): Social Theory for a Changing Society. 1991.

^[3] I.Hacking, The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction, and Statistical Inference, 2nd ed., Cambridge University Press, New York, 2006.

^[4] Medina, E. Cybernetic revolutionaries: Technology and politics in Allende's Chile. Mit Press. 2011.

^[5] 人民日报. 三评算法推荐：不能让算法决定内容、别被算法困在“信息茧房”、警惕算法走向创新的反面[P/OL]. <http://opinion.people.com.cn/n1/2017/0918/c1003-29540709.html>

^[6] 许可. 人工智能的算法黑箱与数据正义[N]. 社会科学报, 2018-03-29(006).

它更加注重协调、互动和合作，由于各方有不同的利益诉求，为公平地治理算法问题，参与治理的主体往往是多元的，包括政府、公民和企业，这些主体在参与治理的过程中相互合作、达成共识，使不同的诉求、利益和冲突得以调和^[7]。因此，基于多元主体协同的算法治理能够有效地使各种社会利益方整合、优化，最终推动算法的善治。

因此，本文根据多元协同治理理论，对当前智慧派单算法原理进行分析，研究的主要内容包：第一，通过爬取 6903 条与“派单算法”相关的线上微博、贴吧评论数据，进行整理分析；第二，以知网 HowNet、台湾大学 NTUSD、清华褒贬义三个词典为基础，结合三者构建了一个全面的情感分类词典，对情感强度进行分析并根据词频提取主题词，以此来挖掘公民的态度或情感变化以及最关注的话题等；第三，运用评论数据中多元利益主体的目标诉求，并提出基于算法的治理模式，并阐述其与传统时也会治理模式的区别；最后，本文根据实证案例结果，对政府、企业、公众三方互相协作治理新生事物进行了探讨。

1.2 研究意义

1.2.1 理论意义

“算法治理”不仅是一种科技技术，更是一种新的权力形态，虽然目前许多算法的应用是跨学科的，涉及心理学、计算机科学等跨自然与人文的学科，但也无法实现完全的“价值中立”，需要引入一种相对公平可行的治理理论模式，本文将多元主体协同治理理论运用到算法治理中，以此来更加的明晰政府、企业、公众三方的权责分配与利益诉求，使得算法整体更公平更有逻辑，摒弃算法偏见等问题，让算法治理更加科学。目前，尚未有将“算法治理”与“多元主体协同理论”结合的文章或讨论，所以该理论的结合具有新颖性，对未来优化的治理模式具有很大的意义。

基于“马尔可夫决策过程”（Markov Decision Process, MDP）的派单算法研究是滴滴 AI 实验室在大数据挖掘方向的顶级会议 KDD2018 上发表论文“Large Scale Order Dispatch in On-Demand Ride-Hailing Platforms: A Learning and Planning Approach（网约车平台的按需大规模派单调度：一种学习和规划方法）”中提及的算法，该算法是目前较新优化的智慧派单算法，与网约车平台相关的其它研究较少，有很多应用方面的探讨空间，使得网约车整体算法提升并对跨部门实现治理优化。

通过爬取的网络评论数据，本研究将文本挖掘技术应用于用户对“派单算法”在线评论的情感分析和主题分析中，通过对评论情感词与特征词进行量化并可视化呈现，来找出目前公众、政府、企业三方各存在的问题，再通过地理位置、收入水平等特征找出个体异质性，加入变量并优化目前的算法，以此来深化各部门之间的协同关系的认识，提升多元主体协同的治理体系。

1.2.2 实践意义

当前算法治理日益受到重视，算法的目的、设计与数据使用等问题都涉及相关利益与主观偏见，使得算法操控、算法过滤等情况频繁发生，导致整个算法的决策或预测水平与能力大幅下降。为了对目前现有的治理情况进行改变，相关的优化方案与算法优化是亟待解决的问题。

此外，本研究将“智慧派单算法”作为实证研究案例，因为目前已有的和网约车相关的讨论大多侧重在定性的研究，如周向红（2018）等人探讨网约车市场监管^[8]，王晓蕾等人

^[7] 范文, 魏婷, 魏娜. 现代城市社区的多元主体协同治理实践——以治理理论为分析视角[J]. 改革与开放, 2016(08):62-64.

^[8] 王俐, & 周向红. 平台型企业参与公共服务治理的有效机制研究——以网约车为例. 东北大学学报(社会科学版), 2018, 20(6):601-607.

(2016)建立网约车平台的定价策略^[9]，朱嘉豪(2019)则探讨公众对出租车和网约车的选择等^[10]，众多研究中较少会从网约车平台决策过程的内部算法与居民出行行为的个体差异性相结合的角度探讨。

本研究中的实证案例将集中分析如何更高效地优化订单分配过程，为乘客匹配合适的司机或出行方式，“马尔可夫决策过程”能有效提高平台的长期有效性，从整体上实现效用最大化，但忽略了个体异质性和偏好，因此本研究在此算法基础上运用文本挖掘的方式来探索用户群体在出行目的、出行时段、环保理念等不同主、客观因素影响下，对网约车出行模式进行权衡时的表现，因此该研究具有实践意义，能够为各个部门提供可操作的政策建议来完善目前的治理模式。

1.3 研究方法和思路

1.3.1 研究方法

本文运用文献分析法、特征观点词对模式匹配、TF-IDF 权重测量等方法对线上微博和贴吧等网页爬取的 6903 条与“派单算法”相关的评论数据进行情感强度与主题词提取等分析。具体方法介绍包括：

(1) 文献分析法：通过对数据库检索文献以及阅读相关教材书籍，对“算法治理”、“多元协同治理”、“情感强度分析”等相关概念的研究成果进行整理和分析，接着形成文献综述和文献述评，在整理过程中发现了“派单算法”目前研究领域的空白和发展空间，从而确立了本文的研究方向和实证案例。

(2) 特征观点词对模式匹配：通过在情感强度分析中，将清华褒贬义情感词典、知网情感词典 HowNet、台湾大学 NTUSD 三个词典为基础，结合三者构建了一个全面的情感分类词典，对情感强度进行分析，以此来挖掘公民的态度或情感变化以及最关注的话题等，该方法有助于提高特征词与观点词提取的准确性。

(3) TF-IDF 权重测量：通过计算评论特征词的频率与影响算法的某变量的特征项的关联度，寻找特定用户群体对于算法治理中某一个关注话题的内容，以此来分析用户的异质性，并在后文引入变量优化派单算法的伪代码。同时，由于派单算法的伪代码是公开的，公民的评论是线上公开的，所以相较于问卷和访谈，通过线上评论文本挖掘的方式分析，能够较真实、直接地挖掘出公民、企业和政府实时关注的热点话题与情感态度，排除了主观因素及不可控客观因素的干扰。

1.3.2 思路和章节布局

第一章为绪论。简要地介绍了选题的研究背景，同时阐明了研究内容及其理论意义和实践意义，介绍研究方法和章节布局，并绘制研究框架图。

第二章为文献综述。对研究中涉及的核心概念和理论进行说明。文献内容包括“算法治理理论研究”、“情感强度研究”、“多元协同治理理论研究”等，通过介绍国内外历年的相关研究来总结阐述算法治理的研究积累与应用情况，并分析现有研究的不足，从而提出本研究的立足点与创新点。

第三章为理论构建概述。内容为对目前的算法提供治理模式建构与假设检验，指出可能适合算法治理的理论模式，并在文章结论处进行总结和假设验证。

第四章为多元主体视角下的智慧派单算法应用情景构建。内容包括对多元治理中各主体间的关系进行说明并对当前的派单算法进行分析与说；接着爬取 6903 条与“派单算法”相

^[9] 刘名敏, 王晓蕾. 网约车对城市居民出行方式影响的建模与分析[J]. 物流科技, 2016, 39(10):79-82.

^[10] 朱嘉豪. 网约车经营服务监管法律问题研究[J]. 现代商贸工业, 2019, 40(10):140-142.

相关的线上微博、贴吧评论数据后对文本预处理，分析其情感强度并根据词频等信息提取主题词，以此来挖掘公民的态度或情感变化以及最关注的话题等；然后运用评论数据中的个体异质性特征总结出不同规律，如不同地点或收入的民众对车辆选择有自己的偏好，将异质性变量引入优化智慧派单算法，并提出基于算法的社会治理模式与传统治理模式的区别；最后，本文根据实证案例分析结果，对政府、企业、公众三方互相协作治理新生事物进行了探讨。

第五章是结合个体异质性的智慧派单算法治理方案，该章节将把文本挖掘得出的变量引入伪代码中，分析个体选择行为影响因素，并提出政府治理以及企业治理等具体优化方案。

第六章为结论和展望。针对研究结果进行讨论，验证前文的理论模式建构及假设检验，分析研究存在的局限，并对研究进行展望。

研究框架图如图 1 所示。

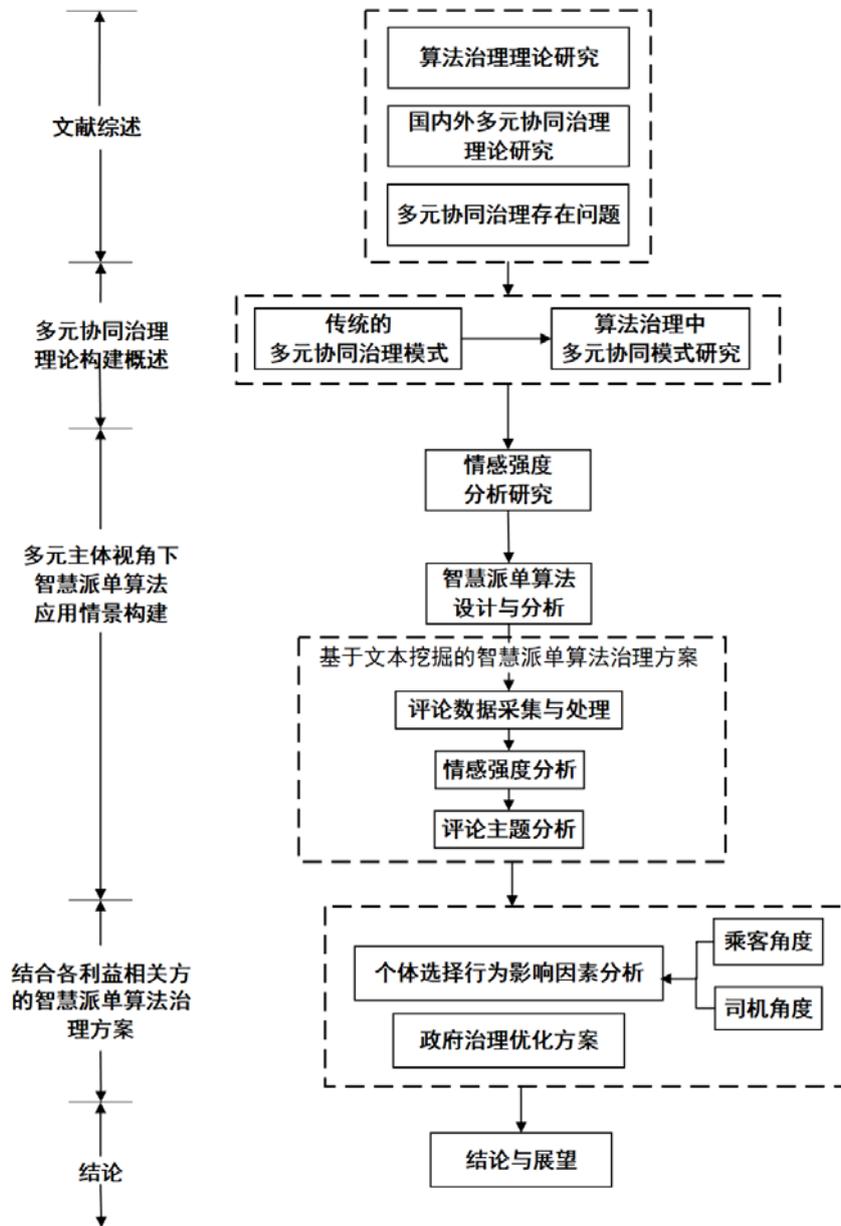


图 1 研究框架图

第二章 文献综述

2.1 算法治理理论研究

“算法”这个名词虽然在现代社会生活中较为普遍，但其定义却未在学术界形成高度共识。在相关定义研究中，Gillespie（2014）等人提出，从狭义上来说，算法被认为是一种特殊的决策技术，但就广义而言，算法往往被认为是建构社会秩序的“理性模型”，大部分与其相关的研究主要聚焦于合理的分类部分，而且也特别关注“和公共利益有联系的算法”^[11]。

关于“算法治理理论”的文献，近年来在国内属于新兴的话题，以国外的文献居多，本文主要将从三个方面进行分析：（1）算法治理的形式或模式；（2）算法治理存在的问题；（3）算法治理的研究方法等，并从这三个维度指出它们如何揭示算法治理的现象，以及研究如何尝试在它们的基础上进行构建。

虽然算法的形式在不断增多，Cormen 等人（2013）提出存在分解为子类型的搜索和排序算法，例如二进制搜索，插入排序，选择排序，合并排序和快速排序等算法^[12]。在评估算法治理系统的社会和规范属性时，理解这些不同类型非常重要。例如，Kitchin 与 McArdle（2016）强调了近年来算法设计中最重要的高级转变之一是“自上而下”算法转变为“自下而上”的机器学习算法^[13]。这种转变很重要，因为当涉及到算法治理系统的不透明度及黑箱问题时，自上而下的算法使用会产生某些问题，特别当这些算法被纳入已经不透明的治理结构时，后果尤为恶劣。

算法治理有许多政策建议方面的好处，包括 Domingos（2016）强调的速度（Speed），效率（Efficiency），综合性（Comprehensiveness）和公平性（Fairness）^[14]。但是，据 Gillespie 和 Seaver（2016）分析，还有越来越多的关键算法研究领域，旨在找出可能产生的社会，伦理，政治和法律问题。例如，Polonetsky 和 Tene（2013）指出关于与算法治理有关的监视系统相关的隐私和数据保护问题^[15]。Crawford（2013）指出有几项研究强调了数据收集和利用中的潜在偏差^[16]。还有其他研究表达了对新出现的“大数据鸿沟”的担忧，这种鸿沟确保只有大型机构才能实现数据革命的好处。有许多学者谈论算法治理中可能存在的不透明性和缺乏透明度，Burrell 与 Danaher（2016）提出尤其是在治理系统由机器学习驱动时，保密法网络如何才能高效地保护这些算法^[17]。Kitchin（2017）则具体列出了三项目前算法治理的

^[11] Gillespie T. The Relevance of Algorithms[C]. 2014.

^[12] Cormen T. Algorithms Unlocked[M]. 2013.

^[13] Kitchin, R., & McArdle, G. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1), 2016.

^[14] Dhar V, Domingos P. Pedro Domingos on The Master Algorithm: A Conversation with Vasant Dhar.[J]. *Big Data*, 2016, 4(1):10-13.

^[15] Gillespie T and Seaver N. Critical algorithm studies. 2016: A reading list. Available at: <https://socialmediacollective.org/reading-lists/critical-algorithm-studies/>

^[16] Crawford K. The hidden biases of Big Data. *Harvard Business Review*. Epub ahead of print 1 April 2013. Available at: <https://hbr.org/2013/04/the-hidden-biases-in-big-data>

^[17] Burrell J. How the machine thinks: Understanding opacity in machine learning algorithms[J]. *Social Science Electronic Publishing*, 2015, 3(1).

研究挑战,包括“黑箱问题”、“异质性和文本嵌入”以及“个体发展和表现”。其中“黑箱问题”指的是算法通常由公司和政府拥有和控制,它们的精细机制常常被隐藏起来,导致难以研究;“异质性和文本嵌入”说明许多关键算法通常由大型团队创建,由现有的代码包组装而成,并嵌入到其他算法的复杂网络中,难以具体剖析;“个体发展和表现”说明算法是静态的、不变的。它们经常根据用户交互进行修改和调整并以无法控制和不可预测的方式发展和变化。此外,问题还包括对这些系统的不准确性,低效率和意外后果的担忧。所有这些问题都有可能破坏算法治理的有效性和合法性。

关于算法治理的研究方法及研究焦点, Kitchin (2017) 提出了较为系统的研究框架,他特别指出,当政策目标需要转换为计算机代码时,要注意“翻译转换”问题,此外,他提出了算法治理过程中有六种潜在方法,包括检查代码、反射性生成代码、逆向工程、程序员访谈、拆解社会技术组合、研究真实世界效应^[18],其中,“检查代码”是指通过筛选文档解构代码,绘制算法谱系图,检查相同的任务如何被翻译转换成不同的编码语言,并在不同的平台上运行;“反射性生成代码”即如何将问题转换为规则集和相关代码;“逆向工程”是指选择虚拟数据,查看在不同场景下输出的内容(如测试谷歌的网页排名^[19]);“程序员访谈”则意味着在编码团队成员构建算法时,仔细观察和访谈他们;“剖析社会技术组合”即对公司文件、行业材料、采购标书、法律标准和框架进行散乱分析;“研究真实世界效应”则是进行用户实验、用户访谈或民族研究,探索算法的社会效应。

2.2 多元协同治理理论研究

2.2.1 国内外多元协同治理理论

当前,国内的很多关于公共服务的研究大部分借鉴了国外的相关概念,并在此基础上再结合我国国情,对其进行修正和提升。目前国内外对于多元主体协同治公共事务的专业词汇都比较复杂和多样,包括合作治理(Cooperative Governance)、网络治理(Network Governance)、公私伙伴关系(Public-Private-Partnership, PPP)、协同治理(Synergy Governance)、无缝隙政府(Seamless Governance)等^[20],结合本文的主要研究内容“基于多元主体协同的算法治理研究”,笔者将对公私伙伴关系、网络治理、协同治理进行定义和详细说明。

公私伙伴关系(PPP)是指两个及以上的公共和私营部门之间的长期性合作^[21],各国政府在历史上常常使用这种合作模式为公民提供服务^[22]。多年来,各国的公私伙伴关系都呈大幅增加的趋势,部分原因是地方和州政府严重依赖于越来越多的非营利组织来提供它们无法提供的许多公共服务^[23]。如果不谨慎监管和教育,进入公私合作伙伴关系既可以是有益的,也可以是破坏性的^[24]。伙伴关系需要双方平衡以及持续维护,但同时也要避免过分

^[18] Kitchin, Rob. Thinking critically about and researching algorithms[J]. Information, Communication & Society, 2016:1-16.

^[19] 邵晶晶. 基于 PageRank 排序算法改进的若干研究[D]. 武汉: 华中师范大学, 2009.

^[20] 李利文. 公共服务供给碎片化研究进展: 类型, 成因与破解模型[J]. 国外理论动态, 2019 (1): 97-107.

^[21] Hodge G A, Greve C. Public-private partnerships: an international performance review[J]. Public administration review, 2007, 67(3): 545-558.

^[22] Roehrich J K, Lewis M A, George G. Are public-private partnerships a healthy option? A systematic literature review[J]. Social science & medicine, 2014, 113: 110-119.

^[23] Smith S R. The challenge of strengthening nonprofits and civil society[J]. Public Administration Review, 2008, 68: S132-S145.

^[24] Babiak K, Thibault L. Challenges in multiple cross-sector partnerships[J]. Nonprofit and voluntary sector quarterly, 2009, 38(1): 117-143.

依赖导致的破坏性结果。根据中国财政部的数据，中国现有的 PPP 项目超过 14000 个，总价值达到 2.7 亿元^[25]。

网络治理 (Network Governance) 涉及了大量的相互依存的行为者，它们相互作用以产生结果^[26]，其决策基于谈判理性而非管理国家规则的控制和监管^[27]，它解释了在高度动荡的环境中存在的组织提高效率和减少代理问题。一方面，通过分布式知识获取和分散问题解决提高效率；另一方面，通过在不同的自我管制活动部门出现集体解决方案来提高效率^[28]。在公共部门，公共行政学者对于网络治理的发展各执一词：一些人怀疑其作为民主治理结构发挥作用的能力，而另一些人则将其视为促进公共产品和服务的有效和有效提供的现象^[29]。

协同治理 (Synergy Governance) 中，“协同”即“相互协调、共同作用”^[30]。这个概念的形成和发展可以说是大科学时代以来系统科学的重要成果之一。在国际上，首次以科学的名义提出“协同”概念的是德国斯图理论物理学教授赫尔曼哈垦 (H.Haken)。在 20 世纪 70 年代，他首次创立了一门新兴的系统学科——“协同学 (Synergetics)”^[31]。“协同学”即“协同合作之学”，它来源于古希腊语，意味着开放系统中，存在着许多子系统相互作用的、集体的或协调的效应。这个理念后来被很好地运用于社会公共事务的治理，根据李汉卿 (2014) 的定义，协同治理就是不断地去寻找有效的治理结构，在这一过程中虽然也强调各种公司部门或者社会组织之间的竞争，但是更多的是强调各个组织行为体之间的协作，以这样的方式来实现整体大于部分的效果^[32]。此外，按照协作程度由浅至深，可以将主体间的具体协同类型划分为九种，分别是纳入、对话、共同规划、共同工作、共同冒险、小型整合、战略联盟、联合、合并。

在我国的传统治理模式中，以政府主导的一元主体下的行政垄断可能会导致许多权力行政的碎片化，使许多公私组织与个人的自身发展与能力受到限制和约束，从而阻碍了公共利益的进一步增长^[33]。随着创新城市治理的发展，城市社区治理的主体不再单单是政府，也有来自不同区域和部门的公私组织以及公民个人。根据俞可平 (2002) 在《全球治理引论》中所述，因为这些组织和行为体具有不同的价值判断和社会资源，所以在社会系统中，它们保持着竞争和合作两种关系，因为在现代社会没有任何一个组织或者行为体具有能够单独实现一个社会理想的目标的知识和资源^[34]。

因此，通过多元的治理可以更有效地满足多方的利益诉求，并以此达到更有效、更有针对性的社会治理。但总体来说，国内的大部分文献都仍局限于表面上的泛泛而谈，很少

^[25] Andoh R. The Politics of Railway Transportation and Development in Ghana, A Case Study of Nsawam and Akim Achiase Junction Railway Stations[D]. University Of Ghana, 2018.

^[26] Nesti G, Graziano P R. The democratic anchorage of governance networks in smart cities: an empirical assessment[J]. Public Management Review, 2019: 1-20.

^[27] Scharpf F W. Games real actors play: Actor-centered institutionalism in policy research[M]. Routledge, 2018.

^[28] Deduurwaerdere T. The Contribution of Network Governance to Sustainability Development[J]. 2007.

^[29] Ferlie E, Fitzgerald L, McGivern G, et al. Making Wicked Problems Governable?: the case of managed networks in health care[M]. OUP Oxford, 2013.

^[30] 何水. 协同治理及其在中国的实现——基于社会资本理论的分析[J]. 西南大学学报(社会科学版), 2008, 34(3):102-106.

^[31] Haken H, Plath P J, Ebeling W, et al. Beiträge zur Geschichte der Synergetik[M]. 2016.

^[32] 李汉卿. 协同治理理论探析[J]. 理论月刊, 2014(1):138-142.

^[33] 陈振明. 公共管理学:一种不同于传统行政学的研究途径[M]. 中国人民大学出版社, 2003.

^[34] 俞可平. 全球治理引论[J]. 马克思主义与现实, 2002(1):20-32.

有切入要害的一些针对性的治理方案。本文将在此理论视角的基础之上，提出一些具体可实施的举措，从政府、企业的角度引入合适的治理自变量，以此来优化整体的算法治理。

2.2.2 多元协同治理存在的问题

多元协同治理模式吸引了社会各方的广泛使用，得以解决许多复杂公共问题，但在实际应用方面仍存在四个方面的问题：

一是公共服务供给中责任边界不清晰；实现多元服务供给主体的良性协作，关键是要明确公民的需求、服务供给和评估监督中各主体的定位与作用，形成责任边界明晰、相互分工协作的关系模式^[35]，但到具体操作时，往往比较困难，协同治理的重要前提是各方要拥有一致的政策目标，但是在许多公共事务上，共同的目标很难同时达成，各方初始目标的偏离或异化就在所难免，这在政府主导下建立的协同治理网络中表现得尤为明显^[36]。例如跨区域环境治理，由于涉及不同地方政府和市场主体，其治理污染的决心受经济形势和舆论氛围影响，当经济下行压力加大时，治理力度往往会减弱，以回应民众对经济发展的诉求。尽管政府可以引导和激励各方将个体利益与公共利益兼容，但各方利益固有的竞争性以及价值理念的差异化使得政策目标难以达成一致。

二是绩效测量比较困难。在多元主体协同治理的条件下，各方如何才能协调整合出合理、合法的绩效测评方案，以此来保证每个季度或者每年的审核评估达标也是目前存在的问题^[37]。

三是管理方面的机会成本和协作安排上的时间成本。协同治理研究中的一个理论难题就是与参与者的协商行为有关，许多研究者用制度主义方法来描述协同治理的结构和过程，但却鲜少有研究会真正关注个体或集体层面的协商行为^[38]。此外，也尚缺少对参与者的协商行为如何影响集体决策质量进行的研究。

四是政府干预经常会发生失灵的情况。协同治理是多元主体的共同行动，政府通常在其中发挥主导作用。为实现协同治理目标并确保治理过程有序，政府需要对其他参与主体进行适当干预，其手段包括监督管理、经济激励、教育引导等。然而经验告诉我们，政府干预行为经常会失灵，几乎所有政府至今未能寻找到协同治理中适度监管的标准和节点^[39]。例如，为充分调动参与各方的积极性，政府会通过放松监管和结成良好的公私伙伴关系来营造宽松的合作环境，其结果可能会造成高额财政支出和低质量公共服务，这一情况在水电气等城市公用设施领域屡见不鲜。

2.3 文献述评

关于“算法治理理论”的文献，近年来在国内属于新兴的话题，以国外的文献居多，但与社会治理领域相关的部分文献探索还是停留在表面、宏观的角度来进行讨论，提出“算法治理”的影响和意义，但具体的细节的算法优化方法却很少有学者进行讨论，因此，本论文具有新颖性，将在研究后引入治理变量提升并优化目前的算法。

^[35] 王德福. 需求对接与协同供给: 社区服务体系运作机制的完善路径[J]. 地方治理研究, 2019(01): 20-32+78-79.

^[36] 刘伟忠. 协同治理的价值及其挑战[J]. 江苏行政学院学报, 2012, 5: 113-117.

^[37] 王莹. 协同治理理论内涵及在城市公共安全治理中的应用[J]. 改革与开放, 2018(19): 112-116.

^[38] 杨金石, 朱有明, 刘金程. 多元主体的社会组织协同治理模式研究——基于“免费午餐项目”的实证研究[J]. 行政科学论坛, 2016(02): 34-39.

^[39] 王焱. 政府与社会组织协同治理公共危机的困境及对策研究[D]. 西北大学, 2018.

“多元协同治理”虽然引起了众多学者的关注，但在实际研究中并没有真正把握其对社会组织提升自身能力的作用机制，主要还是停留在宏观的意义和作用介绍上，举例来说，比如强调协同治理对政府的机制作用、对基层治理水平的提高、对城市社区市民的关注，同时对于社会部门和组织协同治理的困境、影响因素及导致原因作了理论分析。此外，部分学者从主体机制、信任关系等层面提出了协同治理模式实现的基本要素，为处理各种社会组织协同治理困境表达了需要因素，并同时提出了多层合作型等协同治理模式的初步设想。总体来说，国内的研究由于基本大部分从宏观的理论视角或意识形态的立场出发展开讨论，但是却忽视了一些具体案例的实证研究，但同时，也很难借鉴国外的案例研究，因为基本国情也有很大差异。

学界目前尚未有将“算法治理理论”与“多元协同治理”结合的文献，而在本研究中的实证案例里，将集中分析如何更高效地优化订单分配过程，为乘客匹配合适的司机或出行方式，“马尔可夫决策过程”能有效提高平台的长期有效性，从整体上实现经济效益最大化，但忽略了各方的利益诉求和偏好，因此本研究在此算法基础上融入了“多元协同治理”模式，运用文本挖掘、情感强度分析的方式来探索用户群体在出行目的、出行时段、环保理念等不同主、客观因素影响下，对网约车出行模式进行权衡时的表现，在企业视角的基础上，加入了公民和政府的治理影响因素，因此该研究具有实践意义，能够为各个部门提供可操作的政策建议来完善目前的算法治理模式。

第三章 理论构建概述

3.1 传统的多元协同治理模式

为了对传统的多元协同治理模式进行梳理，本节主要阐述了制度分析与发展 IAD (Institutional Analysis and Development) 分析框架、政策网络理论对协同治理进行研究，并以此构建协同治理的理论研究框架。

制度分析与发展 IAD 分析框架将所追求的社会目标、个人关于是否合作的决定、集体行为的质量以及所达到的结果几个元素之间建立起系统性的联系。Ostrom 和 McGinnis (2012) 认为个人是否决定纳入合作，一部分的原因在于潜在合作者的实际行为以及该行为给个人带来的感，对于多方参与的协同行为来说，适当的组织架构和机制将有利于保障协同的高效率^[40]。但是 IAD 分析框架也存在研究的缺陷，即其缺乏对协同治理和整体政治经济环境的相互作用的探讨，因此，IAD 分析框架忽略对许多低收入国家至关重要的政治经济复杂性等客观因素。

政策网络理论包含于某一领域有关联的政府和非政府组织，因为他们的利益相互以来，但并不完全互补，所以政策网络中的行为人彼此结盟，并于其他联盟竞争，以期能影响与自己切身利益相关的决策过程 (Peterson 2006)^[41]。政策网络理论认为政策的制定不仅依赖于政府，还依赖于以松散或正式网络结合起来的个人和机构。网络的不同构架会影响信息的沟通、共享以及协同的驱动力。

表 1 传统的多元协同治理理论构建框架

	动因层面		思想层面		能力层面		
	纳入	授权	共同目标与规划	信任	权力	领导力	成员结构
IAD 分析框架	个人是否决定纳入合作，一部分的原因在于潜在合作者的实际行为以及该行为给个人带来的感受。	受协同行为影响的个人都被授权可以参加规则的制定和修改工作。	相关利益者在空间和管辖边界上达成共识是有效协同管理的先决条件。	发展协同各方之间的互信互惠关系是协同行为长期存在的先决条件。	最小权力认可：参与者制定或参与制定规则的权力受到政府认可	对于多方参与的协同行为来说，适当的组织架构和机制将有利于保障协同的高效率。	
政策网络理论	政策网络包含与某政策领域有关联的政府和非政府组织，因为他们的利益相互依赖，但也不完全互补，所以政策网络中的行为人彼此联盟，并与其他联盟竞争，以期能影响与自己切身利益相关的决策过程。		政策网络的参与者之间具有长期的联系，这使得他们能及时、有效地交换与政策提案相关的信息和资源。	网络参与者会根据共同的核心信仰和互惠关系结成小规模联盟，以倡导关注和解决特定的政策困境。	政策网络理论认为政策的制定和执行不仅依赖于政府，还依赖于以松散或正式网络结合起来的个人和机构。	不同的政策领域中，有效协同网络领导力具有不同特征。	网络的不同构架会影响信息的沟通、共享以及协同的驱动力。

^[40] Understanding Institutional Diversity. Princeton: Princeton University Press. The Euro Crisis Page 30 Pisani-Ferry, J. 2012.

^[41] Peterson J. is the wolf at the door this time? transatlantic relations after Iraq[J]. European Political Science, 2006, 5(1):52-61.

基于上述两个理论，主题维度被分为动因、思想和能力三个层面，动因层面包括纳入、授权，思想层面包括共同目标与规划、信任，能力层面包括权力、领导力与成员结构几个维度，以此把各个理论对多元协同治力不同主题的关注点及结论放入表 1 理论框架的相应位置。

3.2 算法治理中多元协同模式建构

随着互联网的发展，网约车成为了公众出行的重要选择，它在极大程度上方便公众出行，促进交通系统的高效发展，在快速发展的同时也引起了极高的舆论关注度，根据《中国网约车（2019）舆情分析报告》，从 2018 年 4 月 2019 年 3 月期间，五大网约车公司滴滴出行、易到用车、曹操出行、神州专车以及首汽约车均有不同程度的社会关注度，其中滴滴出行的新闻关注度远高于其他网约车公司，五大网约车公司新闻关注对比图可见图 2。

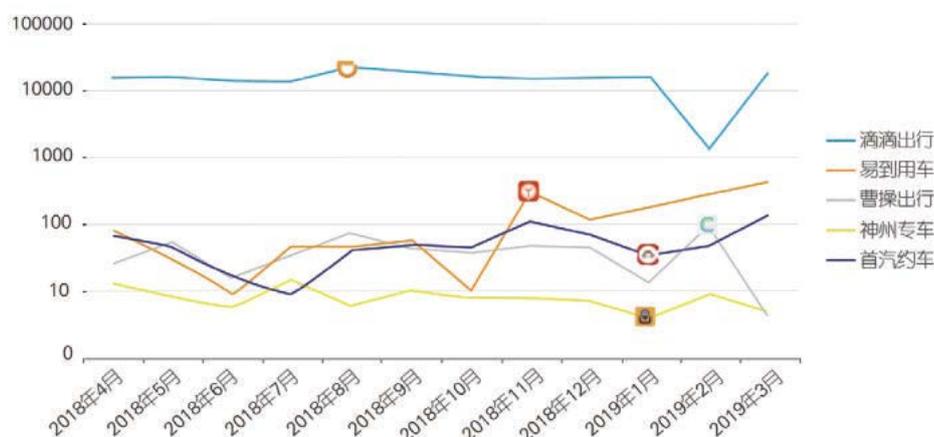


图 2 网约车公司新闻关注度对比 (2018.4-2019.3)^[42]

(来源:《中国网约车(2019)舆情分析报告》)

虽然滴滴的新闻关注度远高于其他同行业公司，但随着数年的野蛮增长，网约车业态暴露出了安全事件频发、竞争无序等问题^[43]。2018 年 5 月，郑州空姐乘滴滴顺风车被司机杀害^[44]；同年 8 月，浙江一名女孩乘顺风车遭司机杀害^[45]；2019 年 3 月，滴滴司机陈某被 19 岁乘客连捅数刀后身亡……根据南方周末统计，在近四年里，滴滴司机性侵、性骚扰等事件至少有 50 个案例，3 起故意杀人案和 19 起强奸案等。面对安全事件和负面新闻，滴滴从去年 8 月下线顺风车业务进行无限期整改，至今未恢复^[46]。

在社会转型的阶段，面对多元利益主体，传统的管理模式逐渐难以应对复杂、多变的情况，包括司乘、平台与政府的关系、矛盾冲突的历史以及长期博弈条件下两者的信任关系等，构建以政府、企业、公民等主体参加的协同治理模式逐渐成为了解决许多冲突和困

^[42] 注：由于图 4 中纵坐标数据差异较大，该数据以对数形式进行表示。

^[43] 李明月. 网络预约出租车的发展问题探索——基于“滴滴”事件[J]. 经济研究导刊, 2019(09):160.

^[44] 文辰西. 网络媒体热点题材报道的话语分析——以澎湃新闻报道“空姐乘滴滴遇害事件”为例[J]. 新闻研究导刊, 2019, 10(03):85+106.

^[45] 朱子芸. 以情境式危机传播理论浅析企业的舆情应对——以乐清滴滴事件为例[J]. 新闻研究导刊, 2019, 10(03):93.

^[46] 史泽帅. 网约车平台的用户权益保护机制——基于滴滴顺风车事件的视角[J]. 科技经济导刊, 2018, 26(36):178-179.

难的有效途径^[47]。寻找一个抽象化、概念化的视角来分析算法治理中不同种类的协同治理是非常大的挑战，结合 IAD 分析框架中相关利益者的纳入合作与达成共识、政策网络理论中的彼此联盟与资源信息共享，引入信息学、算法治理理论等元素构建了新的治理模型，由起始条件（starting conditions）、催化领导（facilitative leadership）、制度设计（institutional design）和协同过程（collaborative process）四个部分组成^[48]。算法治理中多元协同模型如图 3 所示。

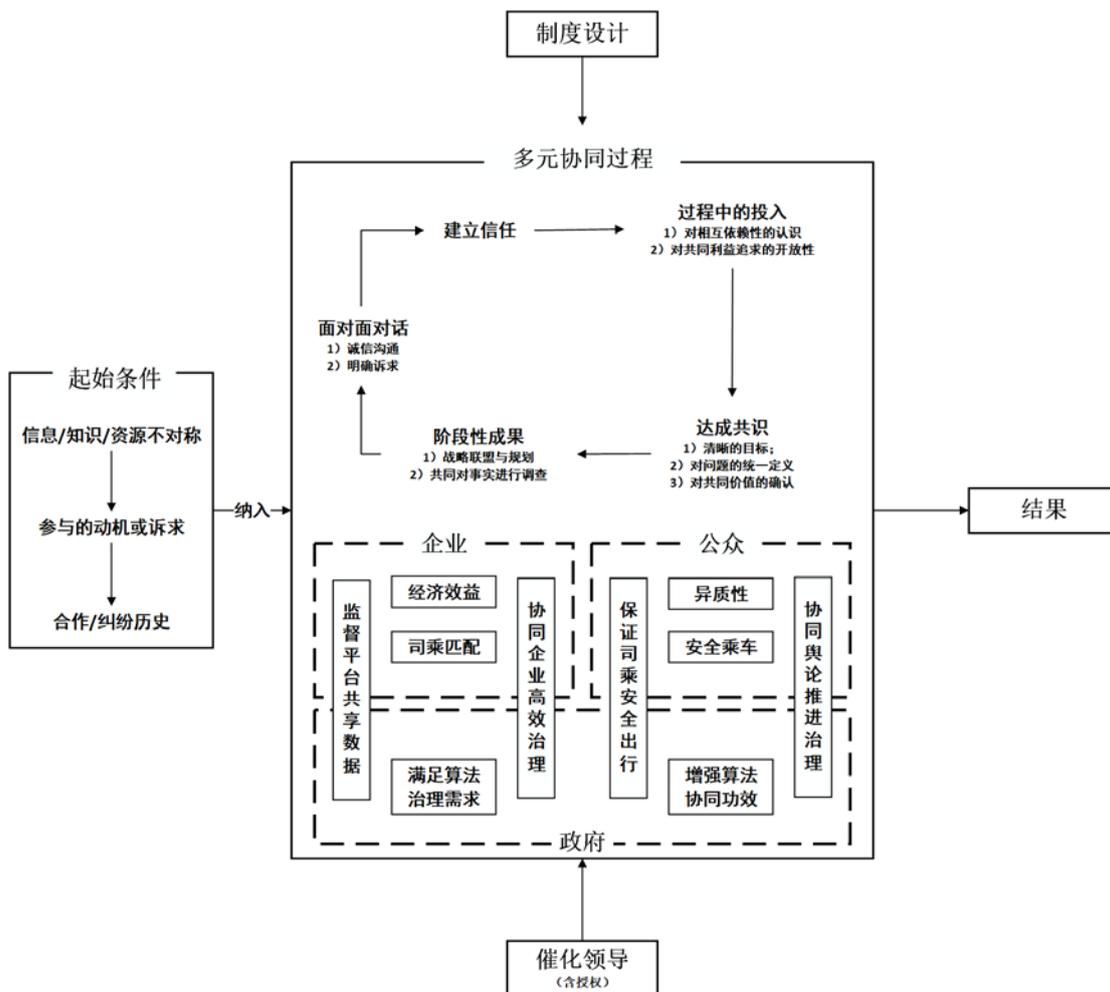


图 3 算法治理中多元协同治理模型

算法治理中的多元协同治理模型由五个独立但又相互存在影响关系的单元组成，各个单元中又包括多个影响因子，影响因子之间呈线性关系，箭头意味着影响方向^[49]。第一单元是起始条件，由于信息、知识、资源的不对成性，影响了参与的动机或诉求，再影响了合作和纠纷的历史，从而决定起始时双方的信任程度，因此纳入了各方进行多元的协同治理；第二单元为催化领导，包括向各个主体的授权；第三单元为制度设计，结合外部的政治与是社会环境等客观因素，制定基础规则，并保持其清晰性与高透明度；第四单元为多元协同过程，各种影响因子形成了互相影响的闭合式循环，包含着各方的利益诉求，首先协同各方开展面对面对话，明确诉求并建立信任，接下来开始各项投入，包括形成相互依

^[47] 李玉明, 陈建平. 关于绿色新政中多元主体协同治理模式建构[J]. 长春理工大学学报(社会科学版), 2013(12).

^[48] Ansell C, Gash A. Collaborative Governance in Theory and Practice[J]. Journal of Public Administration Research & Theory, 2008, 18(4):543-571(29).

^[49] 赵锦. 中国网约车服务业的协同治理研究[D]. 2016.

赖的共识、形成清晰的目标，达成阶段性的成果，例如形成一定的战略规划等，但同时可能会出现一些新的环节和问题，因此需要携同各方反复地完成上述环节，直到达成整体的目标，结束循环并输出协同过程结果。

结合本文的实证案例，网约车服务业的治理问题，不仅仅是管理好一类交通运营车辆，涉及到的影响因子复杂，包括起始条件下的网约车司机与政府的关系、矛盾冲突的历史以及在长期博弈条件下两者的信任关系；也包括在制度设计过程中，网约车司机参与的程度，对制度的认知程度、认可程度以及制度在执行过程当中透明程度等；还包括各个群体之间的权责分配，政府所给予的授权对象以及各主体之间的权责关系是否对等等。以上这些因素都会影响到网约车司机与政府合作的意愿、效率，能否达成共同的愿景，从而促进网约车服务业向着更好的方向发展。因此，对网约车服务业的规制，并不是政府部门闭门造车或简单地出台相关的规定就能做好的，它需要借助一定的理论模型，分析清楚各影响因子内在的逻辑关系，从而有效地促进各类社会因素协同到网约车服务业的治理过程中，达到共同目标。

该模型可以较好地与网约车治理结合是因为从认识到协同的过程不是一个一劳永逸的过程，而是需要协同各方在达成共识、共同投入资源、形成清晰目标并达到阶段性成果之后，进行进一步的交流和转型，应对新的问题从而达成新的目标。网约车服务业治理同样需要政府和网约车司机搁置分歧、达成共识、共同寻找共同价值、并积累阶段性目标的循环往复的过程。所以，该模型可以有效地模拟网约车治理当中各方影响因子的相互关系及其作用过程，具有科学性和可行性。

政府在社会治理中总是扮演着“调控者”与“组织者”的角色，它的政策水平和执行水平直接关系到协同治理的成效结果^[50]。因此，算法治理需要政府对平台数据的监督和管理，在紧急情况下保证司乘安全出行，以此来协同企业进行高效治理，并通过听取舆论建议来推进治理，最终目的是为了满足不同算法治理的需求、增强算法治理协同的功效。政府不仅仅是一名监督员和引导员，它还要保证措施和方案进行落地与执行，所以它还起着“协调员”的角色^[51]，在信息、组织和管理方面对企业和公众进行协调，对于当前算法进行针对性的治理^[52]。

企业是经济发展中十分重要的市场主体，它在多元协同和算法治理中发挥着不可替代的作用^[53]，企业拥有的核心算法主要是为其经济效益服务，实现企业价值最大化，并在算法治理汇总保证司乘可以有效匹配。在日新月异的发展中迅速适应变化，摒弃失败和旧观念，向成功和创新精神发展。但在企业发展的过程中，也需要从公共管理的角度约束其肆意发展，节约资源、保护环境、安全保证并可持续发展应该作为企业发展的首要理念^[54]，因此，企业需要在核心算法的基础上考虑公民的偏向性和政策建议，以此来完善当前算法，并接受政府监督，共同保障司乘安全^[55]。

公众在算法治理的多元协同模式中特指司机和乘客，他们出行乘车的选择上具有个体

^[50] 肖瑶，黄江松. 整体型政府:西方公共管理改革的新目标[J]. 北京行政学院学报, 2009(5):33-36.

^[51] 论协同型政府[D]. 吉林大学, 2010.

^[52] 刘光容. 政府协同治理:机制、实施与效率分析[D]. 华中师范大学, 2008.

^[53] 韩炜. 企业网络组织治理机制与绩效:基于协同视角的研究[J]. 软科学, 2011, 25(6):97-102.

^[54] 曾爱军. 企业多元利益主体共同治理机制研究——基于可持续发展视角[J]. 财经理论与实践, 2009, 30(3):62-65.

^[55] 算法政治:风险、发生逻辑与治理[J]. 厦门大学学报(哲学社会科学版), 2018, 250(76):32-43.

差异性，可能会关注各种不同的话题和主题，同时也非常注重安全出行的重要性。公众的舆论常常被政府和企业主体考虑和听取，以此来推进整个治理体系的运行^[56]。所以，公众也行使着监督权，积极扮演对政府和企业执行方案和政策时的监督角色。

本文新构建的多元协同治理模型考虑到了资源和信息的不对称性，分析的维度会更加丰富全面，例如考虑了资源信息的不平等、安全事件的频发、民众对于环境等因素的考量等。此外，多元协同过程是一个循环的流程闭环，取得一定的阶段性成果后须再进行面对面对话、建立信任等讨论，这样保证最终的算法治理是考量丰富且有效的，传统的理论模型未有这样的循环验证，此外，传统的理论模型，例如资源依赖理论或协同理论还是较为抽象和概念化，没有具体的流程或措施，难以付诸实践，本文构建的多元协同理论直接地阐述了具体流程和主要环节。

本文将以文本挖掘线上评论的方式发现智慧派单算法中各利益方的诉求并寻找共识达成的契合点，以情感强度分析来观察公众方与政府方对于智慧派单的态度。因此，在一章节中，笔者将会以 IAD 分析框架与政策网络理论相结合的理论模型来分析并优化当前的智慧派单算法，用情感强度分析来测量公众对其的态度与建议。

^[56] 基于政府与公民社会多元协同的城市社区治理机制研究[D]. 青岛大学, 2014.

第四章 多元主体视角下的智慧派单算法应用情景构建

本章将主要介绍多元主体视角下的智慧派单算法应用构建，分为情感强度分析研究、当前的智慧派单算法设计与分析以及基于文本挖掘的智慧派单算法评论分析。情感强度分析将结合新构建的治理理论模型用以挖掘民众对智慧派单算法的态度、建议与情感倾向。

4.1 情感强度分析研究

情感强度分析（Sentiment Analysis）又称为评论挖掘（Comments Mining），指的是通过文本挖掘技术，对某现象或事件的在线评论文本内容进行语义挖掘和分析，旨在识别公众对于该事件或某一具体内容^[57]，例如算法的褒贬态度和意见，关于用户评论的情感分析大致可以分为两部分，其一是通过评论语句判断用户对产品的整体情感倾向性，一般分为正面、中性与负面评论；其二是判断用户对于产品具体特征的情感倾向及其强度。

情感分析的流程图如图 4 所示。

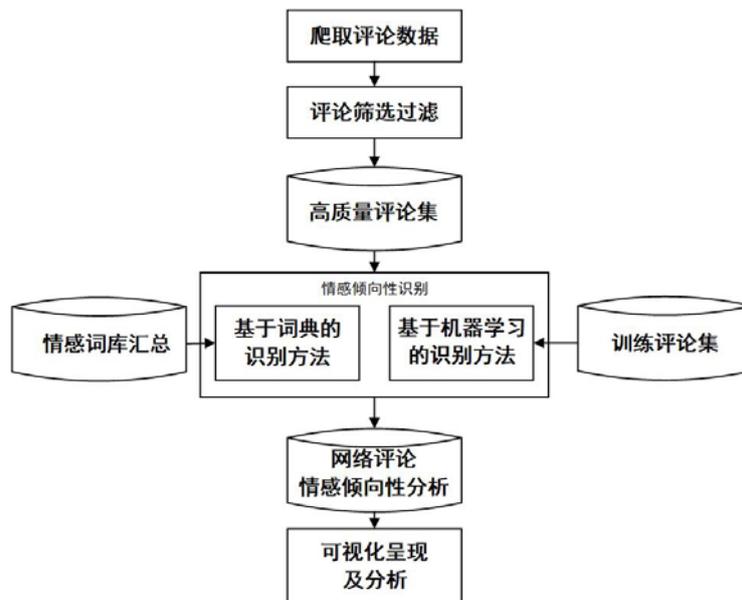


图 4 情感强度分析图

一般来说，按照分析的目的不同，情感分析被分为“粗粒度情感分析”和“细粒度情感分析”^[58]，粗粒度情感分析是剖析整篇长文本，判断文本整体的情感极性或识别作者对某件事物的正面、负面或中性情感倾向；细粒度情感分析是以短句、句子或文本片段为单位，从客观和主观两个维度对某事件或事物的具体特征及其对应观点词进行情感极性分析，通常客观情感维度包含对某事物的具体评价（如派单速度慢、服务态度好、派单距离远等），主观情感维度是指用户并没有对某事物的具体特征进行客观评论，而是表达了自己的主管情感（如非常满意，很讨厌等）。

^[57] 文本情感分析综述[J]. 计算机应用, 2013, 33(6):1574-1607.

^[58] 钟凯, 张传庆. 消费者感知价值对网络购买意愿影响研究——以在线口碑为调节变量[J]. 社会科学辑刊, 2013(3):125-131.

此外,按照任务类型的不同,情感分析可分为情感检索、情感分类、情感信息抽取等子问题^[59]。情感分类是识别一段带有主观描述的文本所表达的正向、负向和中性的情感倾向,以此反映用户对于某事件的褒贬态度;情感检索是指用户输入一个带有情感的查询词或问题,要求系统返回的记过不仅与用户输入的查询词相关,而且还要满足一致的情感倾向,因此情感检索比情感分类更加复杂;情感信息抽取则需要根据需求识别文本的情感表达者、主题词、观点词、评价对象等信息价值。

情感分析中,评论特征词的提取主要分为基于文本的提取方法和基于非本体的提取方法,观点词的提取可以分为基于词典的方法和基于机器学习的方法,具体的提取方法整理详见表 2。

表 2 评论特征词与观点词的提取方法整理

提取词性	研究方法	作者	主要特点
特征词	基于本体的特征词提取	B. Liu	根据线上评论自动化构建基于某事件特征和情感词的模糊领域情感本体树,以便于特征词的提取 ^[60] 。
		唐晓波	利用特征情感词对抽取某事件的评论特征,并根据情感词典计算情感词语特征分类结果的搭配权重,识别产品的隐式特征 ^[61] 。
	基于非本体的特征词提取	肖璐	使用词频词性、关联规则文本模式等自动抽取方法进行显示特征抽取 ^[62] 。
		Z.J. Yan	使用同义词词典扩展的网页排序算法进行评论中的显示特征抽取 ^[63] 。
		邱云飞	使用 Hownet 和聚类算法从评论文本中进行评论对象属性的抽取 ^[64] 。
观点词	基于词典的观点词抽取	M. Q.Hu	通过 WordNet 中种子词生成情感词典 ^[65] ,使用反义词和同义词进行极性分类,生成包含积极和消极情感词的词典。
		F. L. Cruz	使用 WordNet 和 PMI 对情感词进行分类,运用随机森林算法对情感词典进行扩展 ^[66] 。
	基于机器学习	J. B. Zhu	提出多特征的 Bootstrapping 算法用于从未标

^[59] 马晓玲, 金碧漪, 范并思. 中文文本情感倾向分析研究[J]. 情报资料工作, 2013(1).

^[60] Hu M, Liu B. Mining and summarizing reviews[C]// Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004. ACM, 2004.

^[61] 唐晓波, 兰玉婷. 基于特征本体的微博评论情感分析[J]. 图书情报工作, 2016(16):121-127.

^[62] 肖璐, 陈果, 刘继云. 基于情感分析的企业产品级竞争对手识别研究——以用户评论为数据源[J]. 图书情报工作, 2016(1):83-90.

^[63] Yan Z, Xing M, Zhang D, et al. EXPRS: An extended pagerank method for product feature extraction from online consumer reviews[J]. Information & Management, 2015, 52(7):850-858.

^[64] 邱云飞, 倪学峰, 邵良杉. 商品隐式评价对象提取的方法研究[J]. 计算机工程与应用唯一官方网站, 2015, 51(19):114-118.

^[65] Hu M, Liu B. Mining and summarizing reviews[C]// Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004. ACM, 2004.

^[66] Cruz F L, Troyano J A, Enríquez F, et al. 'Long autonomy or long delay?'The importance of domain in opinion mining[J]. Expert Systems with Applications, 2013, 40(8): 3174-3184.

	的观点词抽取		注的数据中学习和特征相关的项目，从评论中抽取种子词进行情感抽取 ^[67] 。
		C.H.Lin	通过对文本中的正向及负向话题使用共同情感话题模型进行情感词抽取 ^[68] 。

基于以上理论基础，结合本文的研究目的，本文属于细粒度情感分析，对用户在线评论的具体特征词的情感分析，以此来捕捉治理问题，在算法治理中，各个参与治理的主体都有各自的诉求，算法治理需要政府对平台数据的监督和管理，在紧急情况下保证司乘安全出行，以此来协同企业进行高效治理，并通过听取舆论建议来推进治理，最终目的是为了满 足算法治理的需求、增强算法治理协同的功效。企业需要在核心算法的基础上考虑公民的偏向性和政策建议，以此来完善当前算法，并接受政府监督，共同保障司乘安全公众的舆论常常被政府和企业主体考虑和听取，以此来推进整个治理体系的运行。公众也需行使着监督权，积极扮演对政府和企业执行方案和政策时的监督角色。

4.2 当前智慧派单算法设计与分析

网约车指的是通过线上网络进行预约的出租汽车。李民（2017）曾指出用户通过在移动通讯设备在网约车平台提供并发布自己的出行需求，平台通过内部算法匹配给合适的平台注册司机，接收到订单的司机主动与乘客联系并前往乘客始发地点，行程结束后，平台自动扣费或由用户主动缴费完成订单，最后将支付回执传送给乘客^[69]，服务流程图如图 5 所示：

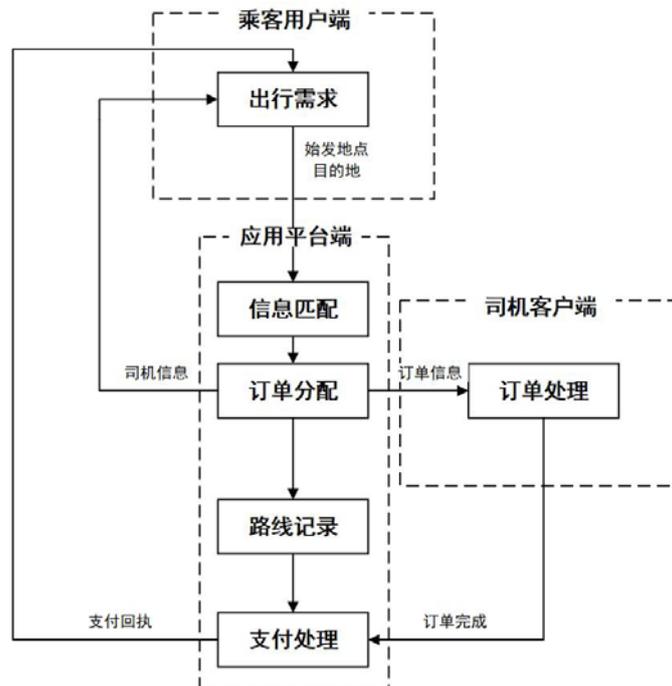


图 5 网约车乘客用户、应用平台、司机客户端服务流程图

实际上，日常出行场景中的司乘匹配是很复杂难测的，车辆在位置上不断发生移动，状态也在不停改变，乘客的需求也各不相同，每一次的派单都会影响未来车辆的分布和交通的拥挤情况。因此，这些背景情况为算法提出了更高的要求，要对司乘进行实时匹配，还要

^[67] Zhu J, Lahiri S N. Bootstrapping the empirical distribution function of a spatial process[J]. Statistical inference for stochastic processes, 2007, 10(2): 107.

^[68] Wu W H, Wu Y C J, Chen C Y, et al. Review of trends from mobile learning studies: A meta-analysis[J]. Computers & Education, 2012, 59(2): 817-827.

^[69] 李民. 基于智能感知技术的约车车辆管理系统设计与实现. (Doctoral dissertation). 2016

计算并保证整体的长期收益。

在本文中，本课题研究将集中分析如何更高效地优化订单分配过程，为乘客匹配合适的司机或出行方式。目前的智慧派单算法使用马尔可夫决策过程（MDP）进行建模，以此来提高平台的长期有效性，从整体上实现效用最大化。马尔可夫决策过程（MDP）是滴滴 AI 实验室于 2018 年 8 月在大数据挖掘方向的顶级会议 KDD2018 上发表论文“Large Scale Order Dispatch in On-Demand Ride-Hailing Platforms: A Learning and Planning Approach（网约车平台的按需大规模派单调度：一种学习和规划方法）”中提及的算法^[70]，该算法也于 2018 年 9 月在滴滴平台运营，截至目前，滴滴使用的也是该算法。该文指出马尔可夫决策过程（MDP）通常用于序列决策问题建模^[71]。在该模型中，每一个司机都被模拟成为一个“代理商”（Agent），其目标是使获得利润 G_t 最大化，其中 G_t 包含了从时间 t 开始的预期未来累计奖励，公式如下所示：

$$G_t = \sum_{i=t}^T R_{t+1} \quad (1)$$

在奖励函数 R_{t+1} 中，文章引入了折扣因子 γ 控制着马尔可夫决策过程中的价值函数，因为从长期视角来看，奖励会随着时间的增长呈边际效益递减，所以，对于具有价格 R 和折扣因子 γ 的持续时间为 T 的订单，奖励函数为

$$R_\gamma = \sum_{t=0}^{T-1} \gamma^t \frac{R}{T} \quad (2)$$

该模型建立了价值函数 $V(s)$ ，基于网约车平台的历史数据，运用动态性的规划（Dynamic Programming）的方法解出了每个时空状态下司机的预期收益。在此算法中，“状态（State）”指司机所处的时间和空间状态，将时空量化成为十分钟的时间段和固定大小的区域，那么一天就有 144 个时间片段组成。图 6 显示了司机分别在“空车（idle）”和“载客（serving）”行为状态下的示意图与价值函数更新。

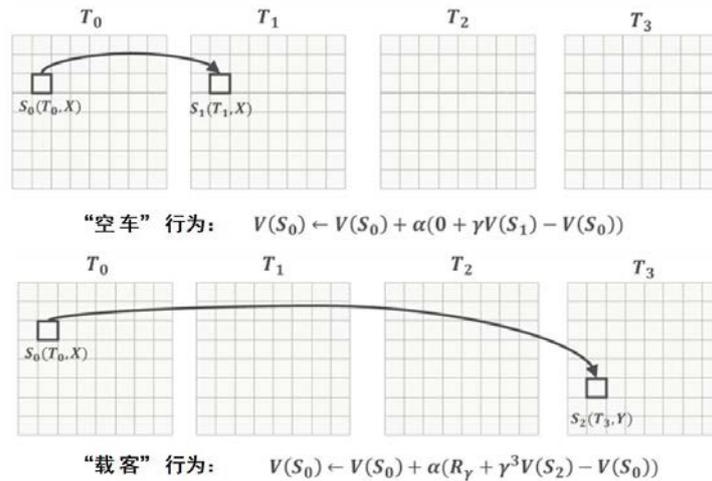


图 6 “空车”与“载客”行为状态的价值函数更新变化

在“空车”行为状态下，假设 $s = (t, g)$ 是现在司机的状态， $s' = (t+1, g)$ 是司机下一时刻

^[70] Xu, Z., et al. Large-Scale Order Dispatch in On-Demand Ride-Hailing Platforms: A Learning and Planning Approach. 2018.

^[71] 白尘. 马尔可夫决策问题的关键状态优先学习算法. 中国管理信息化, 19(7), 198-202. 2016.

的状态，而在此过程中司机没有得到即时的奖励，算法更新规则如下：

$$V(s) \leftarrow V(s) + \alpha[0 + \gamma V(s') - V(s)] \quad (3)$$

在“载客”行为状态下，假设 $s = (t, g)$ 是现在司机的状态， $s'' = (t + \Delta t, g_{\text{dest}})$ 是司机将乘客送达目的地时刻的状态，此处的 Δt 包括了司机赶来接客、等待与送客的所有时间，算法更新规则如下：

$$V(s) \leftarrow V(s) + \alpha[R_\gamma + \gamma^{\Delta t} V(s'') - V(s)] \quad (4)$$

在最后完成订单后，系统会让司机自动发生时空状态的转移，同时会带来奖励，奖励函数会受到折扣系数的影响。

图 4 是在“空车”与“载客”两种不同行为状态下的价值函数 $V(s)$ 的变化，仅在“载客”行为状态下，司机才会获得收益，系统才会开始记录司机的时间和空间位移变化 $T_0 \rightarrow T_3$, $X \rightarrow Y$ 。

在“空车”行为状态中，系统默认司机仍未发生空间上的位置移动，仅发生了时间上的变化。这样的价值函数同时捕获了需求侧和供给侧的时空模式，在以追求平台利益最大化的前提下保证司乘之间的匹配。由此可知，当前的马尔可夫决策过程算法流程如下：

<p>Algorithm Pseudocode (Markov Decision Process)</p> <p><i>Policy evaluation (dynamic programming) for the local-view MDP</i></p> <p>Input: Collect historical state transitions $D = \{(s_i, a_i, r_i, s'_i)\}$; each state is composed of a time and space index $s_i = (t_i, g_i)$.</p> <p>1: Initialize $V(s)$, $N(s)$ as zeros for all possible states.</p> <p>2: for $t = T - 1$ to 0 do</p> <p>3: Find a subset $D^{(t)}$ where $t_i = t$ in s_i.</p> <p>4: for each sample $\{(s_i, a_i, r_i, s'_i)\}$ in $D^{(t)}$ do</p> <p>5: $N(s_i) \leftarrow N(s_i) + 1$,</p> <p>6: $V(s_i) \leftarrow V(s_i) + \frac{1}{N(s_i)} [\gamma^{\Delta t(a_i)} V(s'_i) + R_\gamma(a_i) - V(s_i)]$.</p> <p>7: end for</p> <p>8: end for</p> <p>Return: Value function $V(s)$ for all states</p>

该智慧派单算法从经济利益最大化的角度保证了平台的整体效益，旨在优化平台的长期有效性，同时满足客户的即时需求，通过价值函数大小的比较，将司乘更高效地进行匹配。目前，该方法已应用于滴滴出行的真实调度系统中，在总收入和订单完成率的效果都有显著提高^[72]。

然而，从公共管理的视角来说，该算法未考虑政府、企业、公众三方的信任程度、政府的治理因素、乘客的偏好、能源的消耗等问题，通过政府、企业、公众三方的协同与合作可以有助算法的整体优化，接下来的部分中，本文爬取了线上 6903 条与“派单算法”相关的网络评论数据，通过对评论情感词与特征词进行量化并可视化呈现，来找出目前公众、政府、企业三方各存在的问题，加入变量并优化目前的算法，以此来深化各部门之间的协

^[72] Xu, Z., et al. Large-Scale Order Dispatch in On-Demand Ride-Hailing Platforms: A Learning and Planning Approach. 2018.

发布用户信息	发布者链接	String
评论数据 内容信息	关键词	String
	评论内容	String
	发布网址	String
	发布来源	String
	发布时间	DateTime
	转发数	Double
	评论数	Double
	点赞数	Double
评论数据来源	新浪微博	
	新浪新闻	
	百度贴吧	
	百度搜索	
	百度新闻	
	人民网新闻	
评论数据采集时间	2018年9月1日0时至2019年4月20日24时	

本文对于 6903 条清理后的数据使用 K-平均聚类法及 TF-IDF 算法进行文本聚类并予以可视化分析，目的是为了寻找该评论数据是否可以进行分类，寻找各个族的中心点，并为后续的评论主题筛选打下基础。

代码操作时使用的环境为 Spyder (python3.7)，运用 sklearn 机器学习包、jieba 中文分词以及 matplotlib 的可视化对 6903 条评论数据进行聚类分析，具体流程为导入相关包、加载语料、中文分词、计算 TF-IDF 权重、导出权重、对向量聚类并可视化地输出图片。需要注意的是，将文本向量化后，各个文档的维度较高，进行可视化之前需要对文档降低维度，最后对降维后的数据进行可视化的图片分析。

K-平均聚类算法是一种比较典型的基于距离远近的算法，可以查看主题的相似性，如果两个主题相似性高，则它们的空间立体距离近^[74]。具体算法操作是通过指定某个簇的个数 k ，算法开始随机选择 k 个记录点作为中心点，然后遍布整个评论数据集的每一条，将每条评论数据放至离它最近的中心点所在的簇中，接着更新每个簇的中心点，不断迭代最终直至收敛，算法伪代码如下，其中 k 是中心点的个数， N 是评论数据集的大小，而 T 表示的是迭代次数。

<p>Algorithm Pseudocode (K-MEANS)</p> <p>$K\text{-MEANS} (\{\bar{x}_1, \dots, \bar{x}_N\}, K)$</p> <p>1 $(\bar{s}_1, \bar{s}_2, \bar{s}_3, \dots, \bar{s}_k) \leftarrow \text{SELECT RANDOM SEEDS} (\{\bar{x}_1, \dots, \bar{x}_N\}, K)$</p> <p>2 for $k \leftarrow 1$ to K</p> <p>3 do $\bar{\mu}_k \leftarrow \bar{s}_k$</p> <p>4 while stopping criterion has not been met</p> <p>5 do for $k \leftarrow 1$ to K</p> <p>6 do $w_k \leftarrow \{ \}$</p> <p>7 for $n \leftarrow 1$ to N</p> <p>8 do $j \leftarrow \text{argmin}_{j'} \bar{\mu}_{j'} - \bar{x}_n$</p>
--

^[74] 李鹏浩, 朱立敬, 石秀君. 基于 K-means 算法微博热点话题预测分析[J]. 数字通信世界, 2019(03):84+122.

```

9    $w_j \leftarrow w_j \cup \{\bar{x}_n\}$  (reassignment of vectors)
10  for  $k \leftarrow 1$  to  $K$ 
11  do  $\bar{\mu}_k \leftarrow \frac{1}{|w_k|} \sum_{\bar{x} \in w_k} \bar{x}$  (recomputation of centroids)
12  return  $\{\bar{\mu}_1, \bar{\mu}_2, \bar{\mu}_3, \dots, \bar{\mu}_K\}$ 

```

在将文档主题的相似度转换为数字向量矩阵问题的时候，可以通过 VSM 向量空间模型存储每个文档的词频和权重，特征抽取完成后，因为词语对中心度的贡献程度存在差异，所以需要测量词语赋予的不同权重，本文使用的权重方法是 TF-IDF，它表示 TF（Term Frequency，词频）和 IDF（Inverse Document Frequency，逆文本频率指数），词频指的是某某个词语出现在某文件中所有词语的频率：

$$tf_{ij} = \frac{n_{ij}}{\sum_k n_{kj}} \quad (5)$$

而逆向文件频率则意味着包含某个词条 t 的文档越少，即 n 越小，则 IDF 越大^[75]。它代表一个词语的普遍重要性，计算方式为用总的文件数目除以存在该词语的文件数目，再将得到的商取以 10 为底的对数即可得到

$$idf_i = \log \frac{|D|}{|\{j: t_i \in d_j\}|} \quad (6)$$

因此 TF 与 IDF 的乘积即可测量权重的不同，某一特定文件内的高词语频率，以及该词语在所有的整个文件集合中的低文件频率，可以产生出较高权重的 TF-IDF^[76]

$$TF - IDF_{ij} = tf_{ij} \times idf_i \quad (7)$$

在本研究中，K-平均聚类算法中的 k 值需要预先给定，而确定 k 值是较为困难的，本文通过代码（见附件二）中 `kmeans.inertia_值` 作为评估标准进行参照，确定了 $k=6$ ，最终生成的可视化结果如下图所示：

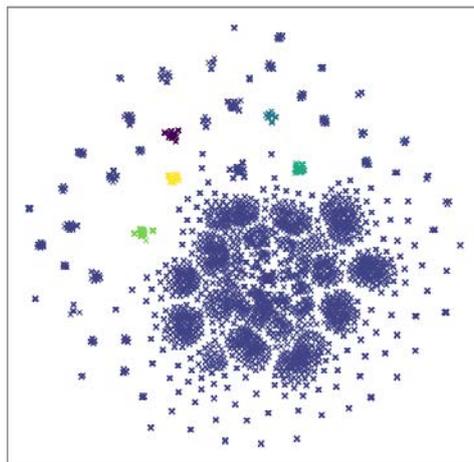


图 8 K-means 评论数据分类结果

^[75] 公冶小燕, 林培光, 任威隆, 张晨, 张春云. 基于改进的 TF-IDF 算法及共现词的主题词抽取算法[J]. 南京大学学报(自然科学), 2017, 53(06):1072-1080.

^[76] 刘家成, 王艺憬, 孙燕红. 基于 TF-IDF 算法和 K-means 聚类的商品评论与价格波动相关性研究——以 ThinkPad 电脑为例[J]. 科技创业月刊, 2018, 31(07):45-49.

可以看到聚类结果并没有把六组结果完全分开，图中 5 组分类相对较为聚集，但数量较少，对这种情况的一个解释是，民众在评论中习惯以不同的角度发表观点，也就是说评论内容看很可能存在交叉性；此外，在语义中经常存在一词多义的情况，因此，在不考虑语义和词义的前提下，该算法无法对词进行准确归类，为克服这一技术问题，本文接下来将对情感词进行提取和分析，并对主题词在 TF-IDF 算法的基础上进行人工辅助分类分析，找到更多影响因变量价值函数的自变量。

4.3.2 情感强度分析

情感倾向判断即指对情感进行分类，对主观性文本内容进行情感类别的判断，通过情感词表可以准确地找到情感词进行匹配，然后根据情感词的大小进一步算出每句话的情感极性值，从而对民众的态度与情感倾向进行客观地衡量^[77]。本文收集了现有的三个情感词典清华大学褒贬词典、知网 HowNet、台湾 NTUSD 情感词典的正面和负面词汇，将情感词进行合并与去重，以此整合为情感倾向的判断词表，情感词表生成流程如图 9 所示。

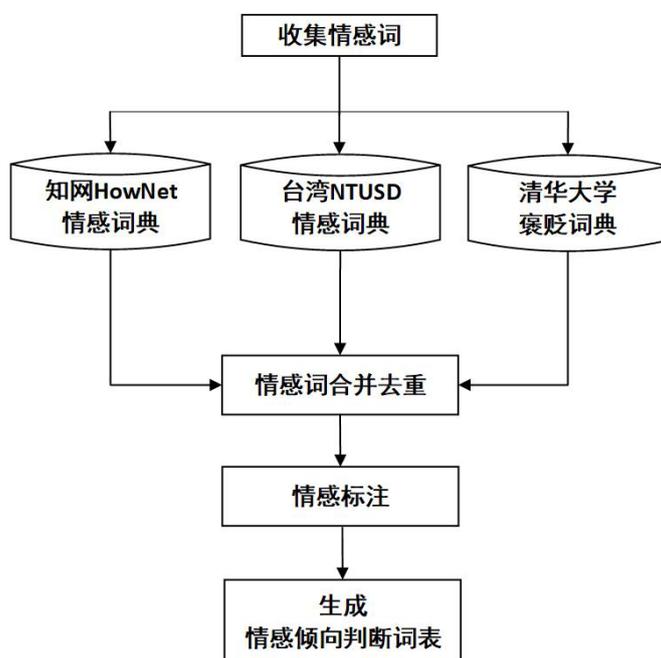


图 9 情感词表生成流程图

知网 HowNet 情感词典、台湾 NTUSD 情感词典、清华大学褒贬词典为本文使用的三个情感词典。其中知网 HowNet 情感词典包括 836 个正向情感词汇与 1254 个负向情感词汇，3730 个正向评价词和 3116 个负向评价词，合并所有情感词后，删除 38 个正向的词汇重复项和 50 个负向的词汇重复项，共得到正向情感词汇 4528 个，负向情感词汇 4320 个，情感词汇共计 8848 个^[78]。

台湾 NTUSD 情感词典包括简体和繁体两个版本，各有 3810 个正向词汇，8274 个负向词汇，通过对 NTUSD 情感词典进行修正，最终得到正向情感词汇 1511 个，负向情感词汇 4597 个，情感词汇共计 6108 个。

清华大学中文褒贬词典共包含 5568 个褒义词，4470 个贬义词，共计 10038 个词汇^[79]。情感词来源汇总整理如表 4 所示。

^[77] 张玉婷. 基于商品评论的情感分析方法研究[D]. 哈尔滨工程大学, 2017.

^[78] 李丽. 基于 HowNet 语义的中文信息抽取及摘要生成方法[D]. 哈尔滨工程大学, 2015.

^[79] 三个情感词典（知网 HowNet，台湾大学 NTUSD，清华大学李军中文褒贬义词典）. 来源：https://download.csdn.net/download/weixin_42018090/10358043. 2018.

表 4 情感词来源汇总

情感词汇来源	正向情感词汇	负向情感词汇	共计
知网 HowNet 情感词典	4528	4320	8848
台湾大学 NTUSD 情感词典	1511	4597	6108
清华大学中文褒贬义词典	5568	4470	10038
汇总情感词典	11607	13387	24994

本文使用的情感强度分析方法是运用 jieba 分词，去除停用词，提取出关键词中的正向情感词与负向情感词，与上述汇总的情感词典中的词语匹配，每个词语对应不同的情感分数，输出每一评论数据的正向情感程度与负向情感程度，并将正向情感程度减去负向情感程度，得到最终评论数据的情感得分，以此来进行文本情感量化分析。部分评论数据的情感得分结果如下表所示：

表 5 部分评论数据情感分析结果

评论	正向情感程度	负向情感程度	最终得分	提取关键词
1. 滴滴的算法我算看明白了，打车的人多的地段越近路程派单越远，今早上班 3 公里，派单 10 公里，你还不能取消，取消了给你派更远，繁忙路段闲一会不怕没生意的路段也硬要给你配上，资源配置一点没浪费，就是不重视客户需求，真是糟糕！	1.3	3.2	-1.9	没浪费 不重视 糟糕 硬要
2. 当规模不再是门槛，滴滴需要向投资方证明自己模式的变现能力。市场有声音认为，滴滴通过信息不对称利用派单及溢价机制来从乘客和司机身上赚取价格差，从而获得利润。多名知乎网友反馈称溢价系统不透明，司机并不知道订单被加价。矛头指向的是滴滴派单背后的算法体系。	2.1	3.2	-1.1	能力 利用 赚取 不透明 溢价
3. 坐滴滴快车听司机吐槽派单的算法和规则。规则是系统派单必须接，否则扣钱无补贴。可在东五环一带活动的司机，接到的超过一半的派单都是来自西四环的叫车，司机叫苦不迭，我说你苦，乘客也郁闷啊，明明叫的快车，还要等你从那么远跑来接，能让平台多边都不爽的规则和算法，也是够尴尬的。	0.1	4.2	-4.1	吐槽 扣钱 叫苦 郁闷 不爽 尴尬
4. 滴滴算法有问题，应该服务客户（打车人）至上，采用就近原则，而不是根据司机积分派单。等滴滴，10 多分钟，还不如坐公交了。	0	-2.2	-2.2	有问题 不如
5. 【滴滴出行】滴滴出行宣布已经与北上广深杭等十余个城市的近 50 家出租车企业达成战略合作，这些出租车企业均将	3.5	0	3.5	合作 探索 融合

引入滴滴互联网技术、计算机算法、平台大数据和运营管理思路,滴滴与其将在智能派单、流量融合等方面展开探索。				
--	--	--	--	--

本文对 6903 条评论数据进行分析,最终提取出正向情感词程度为 6922.2,负向情感程度为 8328.7,最终得到的情感分数为-1406.5。因此,大部分网上对于智慧派单算法的评论还是以负面的评价为主,主要提供了许多改进意见并暴露了目前的算法问题所在。所以,平台运行算法的背后也潜藏着资本的玄机,例如溢价系统不透明、无法满足客户需求等问题浮现出来。在明晰了民众对于当前算法的态度后,可以更好地从他们的建议和评价入手,以政府、企业、公众的多元协同治理体系模式为模型,不断完善、规范当前的算法治理。

结合第三章提出的多元协同的算法治理模型,评论数据反映了公民、政府的利益诉求和偏好,在挖掘各方诉求的基础上进行面对面对话,进行进一步的交流和转型,应对新的问题从而达成新的目标。网约车服务业治理同样需要政府、公民和企业搁置分歧、达成共识、共同寻找共同价值并积累阶段性目标的循环往复的过程。

4.3.3 评论主题分析

本节首先对评论数据中的主题词频进行了统计,通过高频的词频来观测公众的集中兴趣点和讨论的主题,除了搜索的关键词例如“滴滴”,“派单”之外,其他词例如“黑车”、“客服”、“被扣”、“罚款”“运营证”、“环境”等也较为高频,可以看出司机和乘客对于派单算法有不同的关注侧重点,排名前 30 的主题词如表所示。

表 6 排名前 30 的主题词词频统计

排名	主题词	词频	排名	主题词	词频	排名	主题词	词频
1	司机	429	11	公司	123	21	部门	64
2	黑车	272	12	整改	235	22	检查	64
3	车	188	13	担	103	23	交通运输	64
4	客服	186	14	发现	297	24	进驻	62
5	被扣	182	15	郑州	97	25	深夜	54
6	罚款	160	16	环境	90	26	整改	35
7	运营证	130	17	电动车	81	27	漏洞	20
8	无证	140	18	资格证	78	28	合规	16
9	平台	132	19	调查	73	29	费用	16
10	记者	125	20	发布	70	30	不派	15

在对于主题词的词频进行统计后,利用 python3.7 (Spyder) 中的 Wordcloud 包生成了词云图,将高频的主题词可视化呈现,主题词的大小反映了词语出现的频次,字越大则表示出现的频率越高。通过直观的词云图可以发现乘客或司机主要关心的一些主题和问题,词云图可见图 10。但单纯的词频统计只能够说明乘客比较频繁提及的主题分布,但并不能够反映其重要性,通过 TF-IDF 进行权重的测量可以更好地反映词语的重要性。TF 与 IDF 的乘积可测量权重的不同,某一特定文件内的高词语频率,以及该词语在整个文件集中的低文件频率,可以产生出较高权重的 TF-IDF^[80]。

^[80] 刘家成,王艺憬,孙燕红.基于 TF-IDF 算法和 K-means 聚类的商品评论与价格波动相关性研究——以 ThinkPad 电脑为例[J]. 科技创业月刊, 2018, 31(07):45-49.

第五章 结合各利益相关方的智慧派单算法治理方案

根据上一章的 TF-IDF 权重结果，本章将从两个角度挖掘各方的利益诉求，分别是乘客及司机的偏好和需求以及政府的监管。在目前智慧派单算法的基础上，本章从多元协同的角度，根据上文算法治理中多元协同治理模型，通过挖掘各方诉求，建立信任并达成共识，优化目前的派单模式，从公共管理和城市治理的角度，优化当前的派单算法，寻找各方影响因变量价值函数的自变量。

5.1 个体选择行为影响因素分析

5.1.1 乘客角度

从乘客的角度来说，当前爬取的评论从不同角度体现了存在的不同需求，包括乘客的出行方式的偏好和属性，包括乘车目的、出行时间、旅程距离等，此外，也包括出行者的个人情况属性，包括地域、收入、选择司机的性别倾向等。如下表所示，与反映乘客个体异质性相关的部分代表性评论有：

表 7 评论自变量提取（乘客角度）

评论原文	来源	时间	自变量
离上班的地方只有1000米的距离 每天早上却要等从2000到3000米以外接单的滴滴司机 只有一个结果：迟到扣钱。	新浪微博	2019.3.25	乘车目的（通勤）； 出行时间（早上）； 旅程距离（1km）； 派单距离（2-3km）
滴滴司机把车停在离始发地 300 米以外的地方，故意拖延时间，让客人另外加钱才走，让他取消订单他也不取消。我一个月也没赚多少钱，这些司机恶意收费、糊弄乘客！这样的人太损坏南阳的声誉了！希望有关部门能注意！！	新浪微博	2019.3.7	乘客收入（较少）； 乘客地域（南阳）
如果滴滴在派单的时候有个挑选司机性别的选项会不会加大中年女性的就业机会？还是我想多了？至少我每次出行如果遇到女司机整个用户体验非常的好而且有安全感，当然坏人不分性别...	百度贴吧-滴滴打车吧	2018.9.14	选择司机性别倾向 (女乘客偏向女司机)
因为每天打两次滴滴，所以经常自动升级成专车，今早接单的师傅说又见面啦，这个师傅确实和我见过很多次，师傅说我一看就知道是你，周末上班辛苦啊，我朝他笑还是打车方便啊，不用考虑停哪，不用关注路况，算了不买车了！	百度贴吧-滴滴吧	2019.1.15	打车频率 (每天两次)

通过相关的评论研究可知，根据乘客角度的需求可以建立一个新的价值函数 $V_p(n)$ ，该价值函数衡量的是乘客的满意度，可表示为：

$$V_p(n) = ASC_n + \beta_F F_n + \sum_{n=1}^N \beta_X X_n + \sum_{n=1}^N \beta_D D_n + \sum_{n=1}^N \beta_{SC} SC_n \quad (8)$$

主要包括的元素为：第一，乘客 n 出行的固有常量 ASC_n (Alternative Specific Constants)，即乘客的出行出发点与目的地的地址与距离等信息；第二，乘客乘车的频率 F_n ，以此来反映每天乘客乘车的次数和对网约车的依赖性；第三，乘客对于出行的相关偏好及属性以 X_n 来表示，其中包括出行目的 TP (Travel Purpose)、出行时间 DT (Departure Time)、出行距离 TD (Travel Distance)、车型选择 PV (Preferred Vehicle) 等；第四，乘客对于司机的相关偏好以 D_n 来表示，其中包括司机性别 DG (Driver's Gender)、乘客评价的司机星级 DR (Driver's Rating) 等；第五，表示个人的社会属性 SA_n (Social Attribute)，其内容包括乘客性别 PG (Passenger's Gender)、乘客所在城市 PC (Passenger's City) 等。

$\beta_F, \beta_X, \beta_D, \beta_{SC}$ 分别表示公式 (8) 价值函数 $V_p(n)$ 对应的参数属性，该价值函数相关的属性变量与水平详见表 8。

表 8 属性自变量及水平（乘客）

属性变量		属性水平
乘客乘车频率	频率 F_n	一天 3 次及以上/一天 2 到 3 次/一天 1-2 次/ 几乎不使用/完全不使用
X_n 出行相关偏好 及属性	出行目的 TP	通勤（上班/下班）/非通勤（购物/娱乐）
	出行时间 DT	高峰（7:00-9:30, 17:00-19:30, 21:00-次日 6:00）/ 平峰（除高峰外其它时段）
	车型选择 PV	燃油汽车/电动汽车
	出行距离 TD	5km/10km/15km
D_n 乘客对司机的 相关偏好	司机性别 DG	男/女
	司机星级 PR	一星/二星/三星/四星/五星 ^[81]
SA_n 个人社会属性	乘客性别 PG	男/女
	乘客城市 PC	一线城市/新一线城市/二线城市/三线城市及其它城市 ^[82]

在爬取的评论数据中，关于乘客个体异质性相关的信息只有乘客发布评论时所在城市，根据该信息可以得出不同地区的民众对于当前派单算法的关注点不同，但由于仅有 272 评论中公开了评论所在位置，所以只能针对上述评论进行地区差异化分析。其中地理位置为广州的频次最高，共 32 次，其次为嘉兴、南康，深圳、亳州、济南等地，详见表 9。

表 9 排名前 10 的评论发布地点

序号	字词	城市分级	频次	序号	字词	城市分级	频次
1	广州	一线	32	6	济南	二线	12
2	嘉兴	二线	22	7	青岛	新一线	12
3	南康	三线	18	8	天津	新一线	9
4	深圳	一线	16	9	哈尔滨	二线	9
5	亳州	四线	13	10	上海	一线	8

可以从表中得出，除了广州和深圳之外，许多二三线及以下城市的滴滴派单情况也会被常常讨论，并非假设二中提及的一线城市的派单情况最会广受热议，评论内容主要包括乘客

^[81] 刘贝特. 网约车新规背景下消费者满意度影响因素研究[J]. 产业创新研究, 2018(09):62-65.

^[82] 2018 中国城市新分级名单. [2018.4.26]. <https://baike.baidu.com/item/中国城市新分级名单/12702007?fr=aladdin#2>

打车现状、司机接单情况描述、乘客投诉内容等。其中，一线城市如北京、上海、广州、深圳的网友常常会更倾向于说明派车现状并提出可以改进的意见，而二三线城市，如嘉兴、南康的市民则常常会投诉司机的服务和滴滴整体派单的运营情况，其中也不乏一些激烈的言辞。因此，根据评论数据发布地点分析可得全国各地都存在居民对派单算法表达过自己的建议和意见，并不仅仅局限于四个一线城市。

从乘客角度出发，考虑个体异质性的优化算法伪代码详见附录六。在新构建的算法伪代码中，公司的利益以及乘客的需求都被纳入考虑范围内，该算法相较之前的模式会更加公平合理，有较为清晰的目标和共同价值确认，与新构建的理论模型契合，图中 $V_C(s_i)$ 是目前从公司（corporate）平台角度出发的价值函数，其目的是为了保证平台的整体效益，旨在优化平台的长期有效性，同时满足客户的即时需求，通过价值函数大小的比较，将司乘更高效地进行匹配，而 $V_p(n)$ 则是从乘客（passenger）的角度出发，引入乘客乘车频率、出行相关偏好及属性、对司机选择的相关偏好、个人属性等自变量，以乘客的满意度作为价值函数的另一考量因素，最终将两者按比例累加得出一个综合满足乘客需求和平台利益的结果，以此来作为优化后的价值函数 V_{All} ，在公司平台与乘客的共同价值确认，取得一定的阶段性成果后，下一部分将继续挖掘公司平台、乘客、司机的利益诉求。

5.1.2 司机角度

在爬取的评论数据中，除了乘客的观点表达之外也不乏许多司机的声音，他们对特别注意安全事件、派单距离、获得利润等因素。滴滴平台已出台相应功能来保护乘客的出行安全，但仍未有完善的方案保护司机的安全，2019年3月24日，常德市发生司机遇害的命案，一名19岁少年杨某在车内朝网约车司机陈某连捅数刀，致其死亡^[83]。因此，在派单之前如何预防危险、在行驶过程中如何保护司乘安全成为了网约车平台亟待解决的问题。此外，也有司机在线上评论中反映平台的派单距离过远，行程距离小于派单距离等问题，这些都属于司机的出行偏好性。

如下表所示，与反映乘客个体异质性相关的部分代表性评论有：

表 10 评论自变量提取（司机角度）

评论原文	来源	时间	自变量
接客 3 公里，跑单 2 公里， <u>到手的收入连油钱都不够。远距离派单，接乘客远、送乘客近</u> ，即使旁边有车也不会指派，这种派单模式也太坑了吧！	新浪新闻	2019.3.17	接客距离（3 公里） 获得利润（收入）
滴滴恢复夜间运营司机安全服务超一千单才可深夜接单	人民网新闻	2018.9.15	接单量（一千单）
平台只是关注客户的需求有没有关注司机的需求？司机也有 <u>选择和评价乘客</u> 的权利，例如考虑乘客等级评分等因素，但同时也应注重保护乘客隐	百度新闻	2018.9.18	司机对乘客的 相关偏好 (性别、乘客评分等)

^[83] 韦也那. 关于社会安全事件的新媒体舆论导向分析——以“滴滴事件”为例[J]. 视听, 2019(03):154-155.

私。

通过相关的评论研究可知，根据司机角度的需求可以建立一个新的价值函数 $V_d(m)$ ，该价值函数衡量的是司机的满意度，可表示为：

$$V_d(m) = \beta'_O O_m + \sum_{m=1}^M \beta'_X X_m + \sum_{m=1}^M \beta'_P P_m + \sum_{m=1}^M \beta'_{SC} SC_m \quad (9)$$

主要包括的元素为：第一，司机 m 的接单量 O_m ，以此来反映司机的接单和驾龄经验；第二，司机对于接客的相关偏好及属性以 X_m 来表示，其中包括接客距离 PD (Passenger-driver Distance)等；第四，司机对于乘客的相关偏好以 P_m 来表示，其中包括乘客性别 PG (Passenger's Gender)、司机评价的乘客星级 PR (Passenger's Rating)等；第五，表示个人的社会属性 SC_n (Social Attribute)，其内容包括司机性别 DG (Driver's Gender)、司机所在城市 DC (Driver's City) 等。

$\beta'_O, \beta'_X, \beta'_P, \beta'_{SC}$ 分别表示公式 (9) 价值函数 $V_d(m)$ 对应的参数属性，该价值函数相关的属性变量与水平详见表 11。

表 11 属性自变量及水平 (司机)

属性变量		属性水平
司机接单量	接单量 O_m	每月 0-300 单/300-500 单/500-700 单/ 700-1000 单/1000 单及以上
X_m 出行相关偏好 及属性	接客距离 PD	0-1 公里/1-2 公里/2-3 公里/3 公里及以上
D_m 司机对乘客的 相关偏好	乘客性别 PG	男/女
	乘客星级 PR	一星/二星/三星/四星/五星
SA_m 个人社会属性	司机性别 DG	男/女
	司机城市 DC	一线城市/新一线城市/二线城市/三线城市及其它城市

从司机、乘客与平台的角度出发，考虑司机利益方的优化算法伪代码详见附六。在新构建的算法伪代码中，公司的利益以及乘客、司机的需求都被纳入考虑范围内，该算法相较之前的模式会更加公平合理，有较为清晰的目标和共同价值确认，与新构建的理论模型契合，图中 $V_d(m)$ 是目前从司机 (driver) 角度出发的价值函数，目的是为了保证司机的安全与利益，引入司机接单量、出行相关偏好及属性、对乘客选择的相关偏好、个人属性等自变量，以此来作为价值函数的另一考量因素，最终将三者按比例累加得出一个综合满足司乘需求和平台利益的结果，以此来作为优化后的价值函数 V_{All} 。

5.2 政府治理优化方案

2018 年 8 月，滴滴被指在全国范围内拒绝将数据接入政府部门监管，由于没有具体的运营车辆数据及驾驶人员信息，政府部门无法进行针对性的执法，因此导致很多案件无法及时处理^[84]。接着也就在 8 月份，由于接连在郑州、温州发生顺风车乘坐人遇害案件，滴滴出行创始人与总裁发表“郑重道歉”，并宣布将以“安全”而不是“规模和增长”作为公司发展的核心考核标准，整体的组织向安全和客服体系倾斜^[85]。2018 年 12 月，滴滴提出 27 项措施宣布全面进行落实整改，其中包括完善政府和企业的合作机制，建立警方调出证据的绿色通道，安全专线 24 小时待命。因此，政府可时刻监督异常情况，一旦发现任何路线严

^[84] 张津溶. 共享经济时代的网约车监管的路径新选择[J]. 法制博览, 2019(06):264.

^[85] 滴滴出行发文：郑重道歉！（2018.8.28）. http://www.sohu.com/a/250617485_99965852

重偏移、异常停留过久或频繁取消的现象就直接进入干预,即对算法中的时间与空间的状态 $s_i = (t_i, g_i)$ 进行实时监控。

除此之外,在爬取的评论数据中,关键词“环境”与“电动车”也有极高的出现频次,随着涌入城市的网约车数量越来越多,汽车尾气的污染也愈加严重,政府已在全国许多城市发布了鼓励使用电动车作为网约车的条例,例如在 2018 年 8 月份开始,深圳新增的网约车必须全部使用纯电动车辆;佛山从 11 月开始新增网约车须全部使用电动车辆^[86]。因此,政府可通过行程距离和网约车的排量来监督平台整体的碳排放量,即对乘客出行及相关属性 X_n 进行实时监控。

5.3 基于多元协同的算法治理新模式

从上述分析中,我们可以发现派单模式与新构建结合 IAD 框架与政策网络理论的多元协同治理模型相吻合,从起始条件开始,由于资源和信息的不对称,乘客的派单需求与司机的条件往往不相吻合,导致了许多安全事件和纠纷的产生。因此,线上的公开评论成为了很好的挖掘各方利益诉求的切入点,通过寻找问题、发现问题、取得阶段性成果,再继续寻找新的问题的循环协同模式,优化当前的算法,此外,政府需要加强与协同各方的沟通,通过召开听证会、研讨会,广泛听取协同各方代表的声音,积极地谋求与打车软件公司的合作,依法依规地使用权力,尊重打车软件公司作为一个企业的正当利益。

从催化领导方面,平台数据库应接入服务所在地道路运输管理机构出租汽车监管平台,实现信息的共享,如司机的基本信息,乘客的评价等,在本文的评论爬取过程中就发现,网约车平台不会将负面的司乘评价都公布于自己的平台,因此只能通过贴吧与微博爬取评论数据,所以,政府应加强监管,平台也应该主动开放所有公民的评论,保持信息对称,使司乘双方都可以做公平、理性的选择。

在制度设计方面,为了确保网约车发展的可持续性,需要出台相关法律,一是要从法律上给予网约车服务业存在的合法性,拜托行业的“灰色地带”,二是要对目前出现的问题和未来可预计出现的问题给出详细的管理规范,加入相应的法律法规、保险条例,促使网约车服务向更安全、更便利的方向发展。

在多元协同过程中,政府需要明细各利益方的目标,这样各方面的资源才能够围绕目标进行配置,避免资源和人力的浪费。此外,政府需要加强对各项目标的验收,当达到了阶段性的成果以后,政府需要比照成果与原定目标之间的差异,对不符合现实条件的网约车服务业治理目标做出调整,从而完善当前的多元协调体系,为协调过程指明方向。

因此纳入了司乘、企业平台与政府三方,需要面对面对话明确各方的诉求,本文通过文本挖掘的方式分析线上评论,以此来明确各方的诉求,彼此建立信任后通过投入达成共识,形成战略联盟与规划,在算法伪代码中直接的体现就是把影响三方的自变量纳入进整体的价值函数中讨论,以此来满足三方的利益诉求,保证司乘安全出行,协同企业高效治理。

除了派单算法之外,该多元协同治理模式也可适用于其他算法,例如商品搜索的个性化推荐算法,探寻用户的消费习惯并予以推荐,常常会反复推荐类似的商品,如果以用户体验作为产品设计基础,商品搜索排序算法应多元地考虑商品的质量维度、商家的诚信唯独以及用户的偏好来提升该算法,需要通过平台、商家、用户三方进行多元协同,通过面对面对话明确诉求、达成共识、取得阶段性成果,因为网络上的评论和算法是公开透明的,这就决定了利主体的协调和派单算法是一致的,所以,多元协同治理模式能够比较好地适用于其它算法。

因此,本文新构建的多元协同治理模型考虑到了资源和信息的不对称性,分析的维度会

^[86] 佛山网络预约出租汽车经营服务管理实施办法(暂行)[N]. 佛山日报,2018-11-03(004).

更加丰富全面，例如考虑了资源信息的不平等、安全事件的频发、民众对于环境等因素的考量等。此外，多元协同过程是一个循环的流程闭环，取得一定的阶段性成果后须再进行面对面对话、建立信任等讨论，这样保证最终的算法治理是考量丰富且有效的，传统的理论模型未有这样的循环验证，此外，传统的理论模型，例如资源依赖理论或协同理论还是较为抽象和概念化，没有具体的流程或措施，难以付诸实践，本文构建的多元协同理论直接地阐述了具体流程和主要环节，在大数据背景下，帮助人工智能与算法治理实现善治。

第六章 结论与展望

6.1 研究结论

通过上述内容,本文尝试从多元协同的角度寻找智慧派单算法的优化模式,当前平台以利益最大化为导向,但未考虑乘客、司机的偏向和政府监管治理的变量,总体来说,本文在当前智慧派单算法的基础上构建了新的理论模式,结合文本挖掘的方式对政府、企业、公民的评论进行分析,对当前的算法原理进行分析及优化,本篇文章主要的研究结果包括:

(1) 通过爬取 6903 条与“派单算法”相关的线上微博、贴吧等评论数据,进行清洗、整理并使用 K-平均聚类法及 TF-IDF 算法进行文本聚类并予以可视化分析,目的是为了寻找该评论数据是否可以进行分类,寻找各个簇的中心点,并为后续的评论主题筛选打下基础。聚类结果并没有把六组结果完全分开,图中五组分类相对较为聚集,但数量较少,因此,在不考虑语义和词义的前提下,该算法无法对词进行准确归类,为克服这一技术问题,本文接下来将对情感词进行提取和分析,并对主题词在 TF-IDF 算法的基础上进行人工辅助分类分析,找到更多影响因变量价值函数的自变量。

(2) 情感强度分析方法是以知网 HowNet、台湾大学 NTUSD、清华褒贬义三个词典为基础,结合三者构建了一个全面的情感分类词典,接着运用 jieba 分词,去除停用词,从评论数据中提取正向情感词与负向情感词,与情感词典中的词语匹配,每个词语对应不同的情感分数,输出每一评论数据的正向情感程度与负向情感程度,并将正向情感程度减去负向情感程度,得到最终评论数据的情感得分,以此来进行文本情感量化分析。最终得出大部分网上对于智慧派单算法的评论还是以负面的评价为主,主要提供了许多改进意见并暴露了目前的算法问题所在。

(3) 通过 TF-IDF 算法确定了评论中民众的主要关注点,使用 Python3.7(Spyder)中的 jieba 分词与 sklearn 对爬取的评论进行分词,并计算字词的 TF-IDF 值,通过输出结果可以看到重要性较高的词语为“黑车”、“陈某(滴滴遇害司机)”、“顺风”、“预约”、“需要”、“附近”、“环境”、“高德”等。分别对应的问题有合法运营问题、司机安全隐患、乘客安全隐患、乘客乘车需求、环境因素考量等问题,而这些并未被纳入当前的智慧派单算法的因变量考核范围内,因此从公共管理的角度来说,这些建议和评论为算法治理提供了较大的帮助。

6.2 不足与展望

由于时间和资源的受限,本文仍存在一些不足之处,希望在后续的研究中能够针对这些不足作出改进:

(1) 情感测量方式以提取情感词进行分数计算,虽然根据情感词的程度会有权重大小的不同,但语境会有很多变化,难以准确分析,因此测量结果难免存在误差。另外,在情感词计算规则方面,本文仅考虑了否定词、主观词、正面情绪词和负面情绪词,但未考虑表情符号等直接影响或反映用户情感的元素。

(2) 滴滴平台智慧派单的源代码无法获得,只能在伪代码基础上进行修正,难以进行深度的测评和整改。滴滴平台仅在 KDD2018 上发布了自己的算法伪代码,但由于商业竞争和内部机密等原因,未能获得详细的源代码。

(3) 本文提出的优化方案需要政府、企业、民众的三方协调,虽然引入了更多的相关

变量，但显示的情况和利益关系往往有无法预料的情况和冲突，因此该方案需要根据现实情况及时做出调整。

(4) 线上爬取的评论数据有些未公开发布的地点，因此对于个体异质性的不可避免地存在偏差。在测量中，大部分在二三线城市居住或工作的发布者会更习惯于发布自己的地理位置，因此可以更好地针对该类人群进行个体异质性测量，但一线城市的公民的偏好需求则难以准确测量。

因此，本文希望在今后的研究中有针对性地做出如下改进：

(1) 在今后的研究中，结合访谈等方法，对于未能在爬取的评论数据中获取的信息进行搜集和提炼，以此来丰富异质性以及政府监管的方案，提升算法的结果。

(2) 目前的结果是引入了更多相关的衡量因变量价值函数的自变量，但仍是在伪代码的基础上进行修改，未来的研究中，希望可以尝试获取关于智慧派单算法的源代码，以此来进行更深入的算法提升和改进。

(3) 算法治理问题还会与各方利益和许多部门内部的细节相关，这些无法在伪代码或爬取的评论中获得，因此需要更多的文本资料爬取，因为本研究时间和资源的有限，未能在规定时间内爬取并分析所有相关的文本。

所以，算法治理应避免将问题理想化，要以更贴近现实需求的角度来考虑各种可能发生的实际情况，更好地找到治理的一般性和普遍性，以此来提升社会各个领域的算法，为算法治理研究做出更好地贡献。

参考文献

- [1] Andoh R. The Politics of Railway Transportation and Development in Ghana, A Case Study of Nsawam and Akim Achiase Junction Railway Stations[D]. University Of Ghana, 2018.
- [2] Ansell C , Gash A . Collaborative Governance in Theory and Practice[J]. Journal of Public Administration Research & Theory, 2008, 18(4):543-571(29).
- [3] Babiak K, Thibault L. Challenges in multiple cross-sector partnerships[J]. Nonprofit and voluntary sector quarterly, 2009, 38(1): 117-143.
- [4] Burrell J . How the machine thinks: Understanding opacity in machine learning algorithms[J]. Social Science Electronic Publishing, 2015, 3(1).
- [5] Cormen T. Algorithms Unlocked[M]. 2013.
- [6] Crawford K. The hidden biases of Big Data. Harvard Business Review. Epub ahead of print 1 April 2013. Available at: <https://hbr.org/2013/04/the-hidden-biases-in-big-data>
- [7] Cruz F L, Troyano J A, Enríquez F, et al. ‘Long autonomy or long delay?’The importance of domain in opinion mining[J]. Expert Systems with Applications, 2013, 40(8): 3174-3184.
- [8] Deduurwaerdere T. The Contribution of Network Governance to Sustainability Development[J]. 2007.
- [9] Dhar V, Domingos P. Pedro Domingos on The Master Algorithm: A Conversation with Vasant Dhar.[J]. Big Data, 2016, 4(1):10-13.
- [10] Ferlie E, Fitzgerald L, McGivern G, et al. Making Wicked Problems Governable?: the case of managed networks in health care[M]. OUP Oxford, 2013.
- [11] Gillespie T and Seaver N. Critical algorithm studies. 2016: A reading list. Available at: <https://socialmediacollective.org/reading-lists/critical-algorithm-studies/>
- [12] Gillespie T. The Relevance of Algorithms[C]. 2014.
- [13] Hacking, The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction, and Statistical Inference, 2nd ed., Cambridge University Press, New York, 2006.
- [14] Haken H, Plath P J, Ebeling W, et al. Beiträge zur Geschichte der Synergetik[M]. 2016.
- [15] Hall P , Taylor R . Political Science and the Three New Institutionalisms*[J]. Political Studies, 1996, 44(5):936-957.
- [16] Hodge G A, Greve C. Public-private partnerships: an international performance review[J]. Public administration review, 2007, 67(3): 545-558.
- [17] Hu M , Liu B . Mining and summarizing reviews[C]// Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004. ACM, 2004.
- [18] Hu M , Liu B . Mining and summarizing reviews[C]// Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004. ACM, 2004.
- [19] Huxham C , Beech N . Points of Power in interorganizational forms: Learning from a Learning Network [J]. Academy of Management Annual Meeting Proceedings, 2002: B1-B6.
- [20] Kanter, R. M. The Future of Bureaucracy and Hierarchy in Organizational Theory: A Report from the Field In: Bourdieu, P./Coleman, JS (Eds.): Social Theory for a Changing Society. 1991.
- [21] Kitchin, R., & McArdle, G. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society, 3(1), 2016.
- [22] Kitchin, Rob. Thinking critically about and researching algorithms[J]. Information, Communication & Society, 2016:1-16.
- [23] Medina, E. Cybernetic revolutionaries: Technology and politics in Allende's Chile. Mit Press. 2011.
- [24] Nesti G, Graziano P R. The democratic anchorage of governance networks in smart cities: an

- empirical assessment[J]. *Public Management Review*, 2019: 1-20.
- [25] Peterson J. is the wolf at the door this time? transatlantic relations after Iraq[J]. *European Political Science*, 2006, 5(1):52-61.
- [26] Roehrich J K, Lewis M A, George G. Are public-private partnerships a healthy option? A systematic literature review[J]. *Social science & medicine*, 2014, 113: 110-119.
- [27] Scharpf F W. *Games real actors play: Actor-centered institutionalism in policy research*[M]. Routledge, 2018.
- [28] Smith S R. The challenge of strengthening nonprofits and civil society[J]. *Public Administration Review*, 2008, 68: S132-S145.
- [29] *Understanding Institutional Diversity*. Princeton: Princeton University Press. The Euro Crisis Page 30 Pisani-Ferry, J. 2012.
- [30] Wu W H, Wu Y C J, Chen C Y, et al. Review of trends from mobile learning studies: A meta-analysis[J]. *Computers & Education*, 2012, 59(2): 817-827.
- [31] Xu, Z., et al. Large-Scale Order Dispatch in On-Demand Ride-Hailing Platforms: A Learning and Planning Approach. 2018.
- [32] Xu, Z., et al. Large-Scale Order Dispatch in On-Demand Ride-Hailing Platforms: A Learning and Planning Approach. 2018.
- [33] Yan Z, Xing M, Zhang D, et al. EXPRS: An extended pagerank method for product feature extraction from online consumer reviews[J]. *Information & Management*, 2015, 52(7):850-858.
- [34] Zhu J, Lahiri S N. Bootstrapping the empirical distribution function of a spatial process[J]. *Statistical inference for stochastic processes*, 2007, 10(2): 107.
- [35] 2018 中国城市新分级名单. [2018.4.26]. <https://baike.baidu.com/item/中国城市新分级名单/12702007?fr=aladdin#2>
- [36] 白尘. 马尔可夫决策问题的关键状态优先学习算法. *中国管理信息化*, 19(7), 198-202. 2016
- [37] 曾爱军. 企业多元利益主体共同治理机制研究——基于可持续发展视角[J]. *财经理论与实践*, 2009, 30(3):62-65.
- [38] 陈振明. *公共管理学:一种不同于传统行政学的研究途径*[M]. 中国人民大学出版社, 2003.
- [39] 滴滴出行发文: 郑重道歉! (2018. 8. 28). http://www.sohu.com/a/250617485_99965852
- [40] 范文, 魏婷, 魏娜. 现代城市社区的多元主体协同治理实践——以治理理论为分析视角[J]. *改革与开放*, 2016(08):62-64.
- [41] 佛山网络预约出租汽车经营服务管理实施办法(暂行)[N]. *佛山日报*, 2018-11-03(004).
- [42] 公冶小燕, 林培光, 任威隆, 张晨, 张春云. 基于改进的 TF-IDF 算法及共现词的主题词抽取算法[J]. *南京大学学报(自然科学)*, 2017, 53(06):1072-1080.
- [43] 韩炜. 企业网络组织治理机制与绩效:基于协同视角的研究[J]. *软科学*, 2011, 25(6):97-102.
- [44] 何水. 协同治理及其在中国的实现——基于社会资本理论的分析[J]. *西南大学学报(社会科学版)*, 2008, 34(3):102-106.
- [45] 基于政府与公民社会多元协同的城市社区治理机制研究[D]. 青岛大学, 2014.
- [46] 贾开. 人工智能与算法治理研究. *中国行政管理*, 42-47. 2019.
- [47] 李汉卿. 协同治理理论探析[J]. *理论月刊*, 2014(1):138-142.
- [48] 李黎. 基于文本挖掘的网络舆情情感倾向及演化分析[D]. 湘潭大学, 2017.
- [49] 李丽. 基于 HowNet 语义的中文信息抽取及摘要生成方法[D]. 哈尔滨工程大学, 2015.
- [50] 李利文. 公共服务供给碎片化研究进展: 类型, 成因与破解模型[J]. *国外理论动态*, 2019(1):97-107.
- [51] 李民. 基于智能感知技术的约车车辆管理系统设计与实现. (Doctoral dissertation). 2016
- [52] 李明月. 网络预约出租车的发展问题探索——基于“滴滴”事件[J]. *经济研究导刊*, 2019(09):160.
- [53] 李鹏浩, 朱立敬, 石秀君. 基于 K-means 算法微博热点话题预测分析[J]. *数字通信世*

- 界, 2019(03):84+122.
- [54] 李玉明, 陈建平. 关于绿色新政中多元主体协同治理模式建构[J]. 长春理工大学学报(社会科学版), 2013(12).
- [55] 刘贝特. 网约车新规背景下消费者满意度影响因素研究[J]. 产业创新研究, 2018(09):62-65.
- [56] 刘光容. 政府协同治理: 机制、实施与效率分析[D]. 华中师范大学, 2008.
- [57] 刘家成, 王艺憬, 孙燕红. 基于TF-IDF算法和K-means聚类的商品评论与价格波动相关性研究——以ThinkPad电脑为例[J]. 科技创业月刊, 2018, 31(07):45-49.
- [58] 刘家成, 王艺憬, 孙燕红. 基于TF-IDF算法和K-means聚类的商品评论与价格波动相关性研究——以ThinkPad电脑为例[J]. 科技创业月刊, 2018, 31(07):45-49.
- [59] 刘名敏, 王晓蕾. 网约车对城市居民出行方式影响的建模与分析[J]. 物流科技, 2016, 39(10):79-82.
- [60] 刘伟忠. 协同治理的价值及其挑战[J]. 江苏行政学院学报, 2012, 5: 113-117.
- [61] 论协同型政府[D]. 吉林大学, 2010.
- [62] 马晓玲, 金碧漪, 范并思. 中文文本情感倾向分析研究[J]. 情报资料工作, 2013(1).
- [63] 邱云飞, 倪学峰, 邵良杉. 商品隐式评价对象提取的方法研究[J]. 计算机工程与应用唯一官方网站, 2015, 51(19):114-118.
- [64] 人民日报. 三评算法推荐: 不能让算法决定内容、别被算法困在“信息茧房”、警惕算法走向创新的反面[P/OL].
<http://opinion.people.com.cn/n1/2017/0918/c1003-29540709.html>
- [65] 三个情感词典(知网HowNet, 台湾大学NTUSD, 清华大学李军中文褒贬义词典). 来源:
https://download.csdn.net/download/weixin_42018090/10358043. 2018.
- [66] 邵晶晶. 基于PageRank排序算法改进的若干研究[D]. 武汉: 华中师范大学, 2009.
- [67] 史泽帅. 网约车平台的用户权益保护机制——基于滴滴顺风车事件的视角[J]. 科技经济导刊, 2018, 26(36):178-179.
- [68] 算法政治: 风险、发生逻辑与治理[J]. 厦门大学学报(哲学社会科学版), 2018, 250(76):32-43.
- [69] 唐晓波, 兰玉婷. 基于特征本体的微博评论情感分析[J]. 图书情报工作, 2016(16):121-127.
- [70] 王德福. 需求对接与协同供给: 社区服务体系运作机制的完善路径[J]. 地方治理研究, 2019(01):20-32+78-79.
- [71] 王俐, & 周向红. 平台型企业参与公共服务治理的有效机制研究——以网约车为例. 东北大学学报(社会科学版), 2018, 20(6):601-607.
- [72] 王垚. 政府与社会组织协同治理公共危机的困境及对策研究[D]. 西北大学, 2018.
- [73] 王莹. 协同治理理论内涵及在城市公共安全治理中的应用[J]. 改革与开放, 2018(19):112-116.
- [74] 韦也那. 关于社会安全事件的新媒体舆论导向分析——以“滴滴事件”为例[J]. 视听, 2019(03):154-155.
- [75] 文本情感分析综述[J]. 计算机应用, 2013, 33(6):1574-1607.
- [76] 文辰西. 网络媒体热点题材报道的话语分析——以澎湃新闻报道“空姐乘滴滴遇害事件”为例[J]. 新闻研究导刊, 2019, 10(03):85+106.
- [77] 肖璐, 陈果, 刘继云. 基于情感分析的企业产品级竞争对手识别研究——以用户评论为数据源[J]. 图书情报工作, 2016(1):83-90.
- [78] 肖瑶, 黄江松. 整体型政府: 西方公共管理改革的新目标[J]. 北京行政学院学报, 2009(5):33-36.
- [79] 许可. 人工智能的算法黑箱与数据正义[N]. 社会科学报, 2018-03-29(006).
- [80] 杨金石, 朱有明, 刘金程. 多元主体的社会组织协同治理模式研究——基于“免费午餐项目”的实证研究[J]. 行政科学论坛, 2016(02):34-39.
- [81] 俞可平. 全球治理引论[J]. 马克思主义与现实, 2002(1):20-32.
- [82] 张津溶. 共享经济时代的网约车监管的路径新选择[J]. 法制博览, 2019(06):264.

- [83] 张玉婷. 基于商品评论的情感分析方法研究[D]. 哈尔滨工程大学, 2017.
- [84] 赵锦. 中国网约车服务业的协同治理研究[D]. 2016.
- [85] 钟凯, 张传庆. 消费者感知价值对网络购买意愿影响研究——以在线口碑为调节变量[J]. 社会科学辑刊, 2013(3):125-131.
- [86] 朱嘉豪. 网约车经营服务监管法律问题研究[J]. 现代商贸工业, 2019, 40(10):140-142.
- [87] 朱子芸. 以情境式危机传播理论浅析企业的舆情应对——以乐清滴滴事件为例[J]. 新闻研究导刊, 2019, 10(03):93.

谢辞

时光荏苒，初入交大校园时的场景还历历在目，而此时已经接近尾声，即将要用毕业设计来为我的本科生涯画上句号。在毕业设计整个过程里，从确认导师、选题再到现在的致谢收尾，毕业设计的完成给了我大学前三年从未有过的体验，能够更加专注地去潜心研究一个领域的知识核问题。与文献和代码度过的每一个漫长的白天和黑夜让我清晰地感受到了在国务学院的四年学习给我带来的改变，我想在这里感谢在这四年里帮助过我的人，感谢国务学院给予我的这一段珍贵的回忆。

师从樊老师是我本科生涯极大的幸运，深刻的学术研究和精准的问题剖析让我激发了我对学术研究的热忱。作为我学术道路上的引路人，樊博教授对于交叉学科以及数据分析的见地是对我在毕业论文研究中最大的激励。正是在他的影响和鼓励下，我才有勇气尝试之前未曾接触过的研究方法，有勇气自己编写代码处理数据，有勇气将自己的研究内容不断深挖。自从加入樊老师的师门，老师每周都会组织组会，关心师兄师姐以及我的学位论文进度，及时地为我们指点迷津。樊老师对我们的严格要求鞭策着我一次次地克服自己的畏难心理，让我能够将自己的想法映射成现实。樊老师尊重我的想法，愿意与我进行观点的交流和碰撞，更能发现我们每个人身上也许是微不足道的闪光点来给予我们信心，他是我们的毕业论文的最大驱动力。此外，师门的各位师兄师姐也都对我的论文写作给予了关心和支持。在这里要感谢秦川申老师、于峰老师、若玄姐、嫣婷姐、洲蓬师兄给我提供的许多毕设指导建议，也要感谢姚晨师兄、楚哲同学给我代码方面的指点和教学，他们的观点和意见带给我很多灵感，也帮助我将论文打磨得更加完善。

在我撰写论文的过程中，励洲蓬学长去年的优秀毕业论文是我写作过程中借鉴的蓝本，他规范性的研究帮助我更好地认识了自己所研究的问题。他对我的研究也非常关心，将他的经验和见解倾囊相授，也为我提供了非常有益的材料。在这里也要感谢若玄学姐对我提供的帮助和建议，使我能够更好地处理评论数据。

在研究过程中，除了阅读文献、学习前人的研究之外，与同学们的交流是我最主要的灵感来源。在和吕立远同学就我本人的论文交换意见、相互切磋的过程中，我的研究方法和研究内容被打磨得更加完善，遇到困难或是有所成果时我们的分享和鼓励也使得论文的写作过程变得不那么艰难。感谢李嘉慧和吴仪同学，我们因为大三的学期交换原因，大四有幸在一起上课，为彼此的毕业设计互相鼓励、充分交流，在借鉴对方的优点和长处中不断提高。也要感谢我的亲爱的室友王佳依、巴勒恒、杨心怡，是她们包容了我的无数个挑灯夜战，以及遇到挫折时在寝室中的各种崩溃，更重要的是他们是我的灵感和想法的第一倾诉对象，非常有幸能够收获他们对我的肯定。

最后要感谢一直无条件信任我、支持我，在我背后默默奉献的妈妈。在她每一天的询问和关怀中，我总是会提及毕业论文的压力和困难，是她一直安慰我、劝解我。也要感谢我的先生，在我每一次拥有一点点小进步的时候为我加油鼓励，让我看到前方的光明和胜利，好幸运身边有你们。

流泪撒种的必欢呼收割。筚路蓝缕，风雨无阻。值此结业之际，回望一路走来的每一步，感恩在我前进道路上每一个关心我、帮助我的人！谢谢大家！

附录一 评论数据爬取代码

```
#导入时间与需求包
import time
import requests

#导入数据展示的包
import numpy as np
import pandas as pd

#确认百度贴吧“滴滴打车吧”可访问
r=requests.get(url='http://tieba.baidu.com/f/search/res?isnew=1&kw=%B5%CE%B5%CE%B4%F2%B3%B5&qw=%B5%A5&rn=10&un=&only_thread=0&sm=1&sd=&ed=&pn=1')
print(r)
print(type(r))

#安装 BeautifulSoup
from bs4 import BeautifulSoup

soup = BeautifulSoup(r.text,'lxml')
type(soup)

#寻找目标评论区域题目标签与分类
test = soup.find('div',class_='s_post_list').find_all('a',class_='s_title')
print(test)

#寻找目标评论区域题目标签与分类
test = soup.find('div',class_='s_post_list').find_all('a',class_='bluelink')
print(test)

#get 请求模板
template_url =
'http://tieba.baidu.com/f/search/res?isnew=1&kw=%B5%CE%B5%CE%B4%F2%B3%B5&qw=
%B5%A5&rn=10&un=&only_thread=0&sm=1&sd=&ed=&pn=1'

#从一页中提取帖子
def extra_from_one_page(page_lst):
    """从一页中提取 帖子"""
    # 临时列表保存字典数据，每一个帖子都是一个字典数据
    tmp = []

    for i in page_lst:
        dic = {}
        # 帖子名称
```

```
dic['name'] = i.find(class_='p_title').text
# 帖子地址
dic['address'] = 'https://tieba.baidu.com' + i.find(class_='bluelink').a['href']

tmp.append(dic)

return tmp

# 爬取 n 页的数据
def search_n_pages(n):
    """爬取 n 页数据"""
    target = []

    # 发起 n 次的 get 请求
    for i in range(76):
        # 跟踪进度
        print('page:', i)

        # 按照浏览贴吧的自然行为, 每一页 50 条
        target_url = template_url.format(50*i)
        res = requests.get(target_url)

        # 转为 bs 对象
        soup = BeautifulSoup(res.text, 'html.parser')

        # 获取该页帖子列表
        page_lst = soup.find_all(class_='s_post')

        # 该页信息保存到 target
        target.extend(extra_from_one_page(page_lst))

        # 休息 0.2 秒再访问, 友好型爬虫
        time.sleep(0.2)

    return target

# 爬取贴吧 76 页数据
d = search_n_pages(76)

# 转化为 pandas.DataFrame 对象
data = pd.DataFrame(d)
data

# 导出到 excel 表格
data.to_excel('百度贴吧_滴滴打车吧_单_comments.xlsx')
```

附录二 情感强度分析代码

```
#1.导入 jieba
import jieba

#导入负面情感字典 allnegative.txt
negative_dic = open(r"C:\\Users\\lrq20\\Desktop\\allnegative.txt", "r", encoding="utf-8")
negative_arr= negative_dic.readlines()
negative_word = []
for arr in negative_arr:
    arr = arr.replace("\n", "")
    negative_word.append(arr)
# print(negative_word)

#导入积极情感词典 allnegative.txt
positive_dic = open(r"C:\\Users\\lrq20\\Desktop\\allpositive.txt", "r", encoding="utf-8")
positive_arr = positive_dic.readlines()
positive_word = []
for arr in positive_arr:
    arr = arr.replace("\n", "")
    positive_word.append(arr)
# print(positive_word)

#将文本进行分词
def segmentation(txt):
    seg_list = jieba.lcut(txt)
    seg_result = []
    for w in seg_list:
        seg_result.append(w)
    return seg_result

#去除停用词
def del_stopWord(seg_result):
    stop_dic = open(r"C:\\Users\\lrq20\\Desktop\\中文停用词.txt", "r", encoding="utf-8")
    stopwords = stop_dic.readlines()
    arr = []
    for stopword in stopwords:
        stopword = stopword.replace("\n", "")
        arr.append(stopword)
    new_sent = []
    for word in seg_result:
        if word not in arr:
            new_sent.append(word)
    return new_sent
```

```
def score(txt):
    # score = []
    seg_result = segmentation(txt)
    seg_result = del_stopWord(seg_result)
    print(seg_result)
    poscount = 0
    negcount = 0

    for word in segmentation(txt):
        if word in negative_word:
            negcount +=1
        elif word in positive_word:
            poscount +=1
    final_score = poscount - negcount

if __name__ == '__main__':
    with open (r"C:\\Users\\lrq20\\Desktop\\All_DIDI Comments.txt", "r", encoding="utf-8") as
f:
        txt=f.read()
        print(segmentation(txt))
        score (txt)

        print('poscount:',poscount)
        print('negcount:',negcount)
        print('final_score:',final_score)
```

附录三 主题词词频分析代码

```
#1.导入相关包
import jieba
import matplotlib.pyplot as plt
import re
import collections
from wordcloud import WordCloud, STOPWORDS
from zhon.hanzi import punctuation

#2.加载语料
# 结巴中文分词
segment_jieba = lambda text: " ".join(jieba.cut(text))

corpus = []
tokens = []
with open(r"C:\Users\lq20\Desktop\All_DIDI Comments.txt", "r", encoding="utf-8") as f:
    lines = f.readlines()
    i=0
    for line in lines:
        i+=1
        if i == 300:
            break
        #去除标点符号
        chi_nopuc = re.sub("[{}]+".format(punctuation), "", line)
        # 去掉标点符号
        chi_token = jieba.lcut(chi_nopuc)
        tokens.extend(chi_token)

#去除停用词
#首先我们需要引入停用词
f = open(r"C:\Users\lq20\Desktop\中文停用词.txt",'r',encoding = 'utf-8')
stopwords_n = f.readlines()
f.close()
#清洗停用词数据
stopwords = [sw.strip().replace('\n','') for sw in stopwords_n]

#导出分词结果
for chi in tokens:
    if chi not in stopwords:
        corpus.append(chi)
print(corpus)

# 对分词做词频统计
```

```
word_counts = collections.Counter(corpus)
# 获取前 100 最高频的词
word_counts_top100 = word_counts.most_common(100)
print(word_counts_top100)

#绘制词云
my_wordcloud = WordCloud(
    background_color='white', # 设置背景颜色
    # mask=img, # 背景图片
    max_words=400, # 设置最大显示的词数
    stopwords=STOPWORDS, # 设置停用词
    # 设置字体格式, 字体格式 .ttf 文件需自己网上下载, 最好将名字改为英文, 中文名路
    # 径加载可能会出现问題。
    font_path='simhei.ttf',
    max_font_size=250, # 设置字体最大值
    random_state=50, # 设置随机生成状态, 即多少种配色方案
    ##提高清晰度
    width=1000,height=600,
    min_font_size=20,
).generate_from_frequencies(word_counts)

# 显示生成的词云图片
plt.imshow(my_wordcloud)
plt.axis('off')
plt.show()
plt.savefig('my_wordcloud.png')
```

附录四 TF-IDF 算法运行部分结果

序号	关键词	TF-IDF	序号	关键词	TF-IDF
1	黑车	0.032916369	40	青岛	0.001589711
2	陈某	0.028614798	41	随便	0.001496199
3	顺风	0.023004053	42	随着	0.001496199
4	预约	0.021788392	43	难赚	0.001496199
5	需要	0.021694879	44	音乐	0.001496199
6	附近	0.012998225	45	领导	0.001496199
7	环境	0.006919918	46	马路上	0.001496199
8	难道	0.005984794	47	马鞍山	0.001496199
9	马上	0.005704257	48	阶段	0.001402686
10	顺路	0.005236695	49	骗人	0.001402686
11	非常	0.004582108	50	骚扰	0.001402686
12	首汽	0.004395083	51	高倍	0.001309174
13	高额	0.004395083	52	防晒	0.001215661
14	高德	0.004021034	53	随口	0.001215661
15	驾驶员	0.003927521	54	难以	0.001215661
16	限制	0.003740496	55	难过	0.001215661
17	阿姨	0.003553472	56	驾照	0.001215661
18	页面	0.003459959	57	齐全	0.001215661
19	非京牌	0.003272934	58	难受	0.001122149
20	随意	0.002898885	59	雕像	0.001122149
21	驾驶	0.002898885	60	青年	0.001122149
22	非法	0.002805372	61	面膜	0.001122149
23	除了	0.00271186	62	预计	0.001122149
24	面前	0.00271186	63	鳌头渚	0.001122149
25	高峰	0.002618348	64	陌生人	0.001028637
26	顺利	0.002431323	65	降维	0.001028637
27	高速	0.00233781	66	随后	0.001028637
28	麻烦	0.00233781	67	随时	0.001028637
29	陌生	0.002244298	68	预警	0.001028637
30	顾客	0.002150785	69	默默	0.001028637
31	黑色	0.002150785	70	陈露	0.000935124
32	需求	0.002057273	71	降低	0.000935124
33	面对	0.002057273	72	降到	0.000935124
34	风口浪尖	0.002057273	73	韩信	0.000935124
35	飞机	0.001963761	74	预订	0.000935124
36	饭店	0.001870248	75	风格	0.000935124
37	驾乘	0.001870248	76	食堂	0.000935124
38	顺便	0.001683223	77	首次	0.000935124

附录五 优化后的智慧派单算法伪代码

Optimized Algorithm Pseudocode 1

Input: Collect historical state transitions $D = \{(s_i, a_i, r_i, s'_i)\}$; each state is composed of a time and space index $s_i = (t_i, g_i)$.

1: Initialize $V_C(s_i)$, $N(s)$, $V_p(n)$ as zeros for all possible states.
2: **for** $t = T - 1$ to 0 do
3: Find a subset $D^{(t)}$ where $t_i = t$ in s_i .
4: **for** each sample $\{(s_i, a_i, r_i, s'_i)\}$ in $D^{(t)}$ **do**
5: $N(s_i) \leftarrow N(s_i) + 1$,
6: $V_C(s_i) \leftarrow V_C(s_i) + \frac{1}{N(s_i)} [Y^{\Delta t(a_i)} V_C(s'_i) + R_Y(a_i) - V_C(s_i)]$.
7: $V_p(n) \leftarrow V_p(n) + ASC_n + \beta_F F_n + \sum_{n=1}^N \beta_X X_n + \sum_{n=1}^N \beta_D D_n + \sum_{n=1}^N \beta_{SA} SA_n$
8: $V_{All} \leftarrow V_{All} + \beta_C V_C(s_i) + \beta_p V_p(n)$
9: **end for**
10: **end for**
Return: Value function V_{All} for all states and departments

图 12 优化算法一：加入“乘客角度自变量”后的智慧派单算法伪代码

Optimized Algorithm Pseudocode 2

Input: Collect historical state transitions $D = \{(s_i, a_i, r_i, s'_i)\}$; each state is composed of a time and space index $s_i = (t_i, g_i)$.

1: Initialize $V_C(s_i)$, $N(s)$, $V_p(n)$, $V_d(m)$ as zeros for all possible states.
2: **for** $t = T - 1$ to 0 do
3: Find a subset $D^{(t)}$ where $t_i = t$ in s_i .
4: **for** each sample $\{(s_i, a_i, r_i, s'_i)\}$ in $D^{(t)}$ **do**
5: $N(s_i) \leftarrow N(s_i) + 1$,
6: $V_C(s_i) \leftarrow V_C(s_i) + \frac{1}{N(s_i)} [Y^{\Delta t(a_i)} V_C(s'_i) + R_Y(a_i) - V_C(s_i)]$.
7: $V_p(n) \leftarrow V_p(n) + ASC_n + \beta_F F_n + \sum_{n=1}^N \beta_X X_n + \sum_{n=1}^N \beta_D D_n + \sum_{n=1}^N \beta_{SA} SA_n$
8: $V_d(m) \leftarrow V_d(m) + \beta'_O O_m + \sum_{m=1}^M \beta'_X X_m + \sum_{m=1}^M \beta'_P P_m + \sum_{m=1}^M \beta'_{SA} SA_m$
9: $V_{All} \leftarrow V_{All} + \beta_C V_C(s_i) + \beta_p V_p(n) + \beta_d V_d(m)$
10: **end for**
11: **end for**
Return: Value function V_{All} for all states and departments

图 13 优化算法二：加入“乘客和司机角度自变量”后的智慧派单算法伪代码

ALGORITHM GOVERNANCE RESEARCH BASED ON MULTI-AGENT COLLABORATION: A CASE STUDY ABOUT ORDER DISPATCH ALGORITHM

With the advent of the third wave of artificial intelligence digitalization, algorithm has increasingly become the basic law in the society, which affects social operation and government decision-making. The improvement of algorithm efficiency and the popularization of application ability have undoubtedly greatly benefited the public. However, at the same time, new challenges have also been created. From the perspective of governance objectives, government aims to maximize social welfare; enterprises oriented at gaining the maximum profit; citizens, nevertheless, aims at gaining their high efficiency. Therefore, from the perspective of multiple agents, how to balance the triple relationship and optimize the overall governance of algorithm has aroused a heated discussion. According to the current social background, due to the design bias of algorithm and the confidential algorithm black box, citizens have benefited from the algorithm services but also affected and controlled by algorithm silently. Based on the theory of multi-agent collaboration, this paper analyzes the principle of the current order dispatch algorithm. The main research contents include: first, crawling 6903 online Weibo and other comment data related to "order dispatch". Second, based on HowNet, NTUSD from Taiwan University and Tsing Dictionary, a comprehensive emotion classification dictionary is constructed, which helps analyze the words extracted from the comments, thus exploring the change of citizens' attitudes or emotions and their most concerned topics. Third, the individual heterogeneity in the review data is used to summarize various rules. Fourth, the heterogeneous variables are then introduced into the optimized order dispatch algorithm and the differences between the social governance model based on the algorithm and the traditional governance model are proposed. Finally, based on the result of the empirical case analysis, this paper discusses the new issues of collaborative governance among the government, enterprises and the citizens.

In this study, the empirical case focuses on analyzing how to more effectively optimize the order allocation process, matching the appropriate driver or transportation for passengers. Therefore, the Markov Decision Process can effectively improve the long-term effectiveness of the platform, maximizing the utility from the overall. However, it ignores the individual heterogeneity and preferences, so this study based on the algorithm, which uses the text mining methods to explore the user group in the travel purpose, travel time, environmental protection concept under the influence of various subjective and objective factors, such as to weigh on net about car travel mode, so the research has practical significance. More effective measures can be taken by various departments to improve the current governance model.

The method of literature review analysis, matching of characteristic viewpoint words to pattern, TD-IDF feature aggregation and other methods were used to analyze the emotional intensity and subject word extraction of 6903 comments related to "single assignment algorithm"

crawled from online Weibo, Baidu Tieba and other websites. Specific methods are as follows:

(1) Literature analysis method: based on the database literature retrieval and reading related books, the algorithm of "governance", "multiple cooperative governance", "emotional intensity analysis" research achievements of related concepts such as sorting and analysis, and then form a literature review and literature review, found in the process of finishing "send single algorithm" blank in the field of current research and development space, thus established in this paper, the research direction and the empirical case.

(2) Characteristic Viewpoints: To analyze the emotional intensity, the HowNet Dictionary, NTUSD Dictionary in Taiwan university and Tsing Dictionary are combined to build a comprehensive emotional classified dictionary. Then, we can excavate the attitude of the citizens or emotional change and most concerned topics, etc., this method is helpful to improve the accuracy of the paper.in and opinion word extraction method.

(3) TF-IDF feature aggregation method: By calculating the correlation between the frequency of comment keywords and the feature items of a variable that affects the algorithm, the content of a topic of concern for algorithm governance by a specific user group is searched to analyze the heterogeneity of users, and the pseudo-code of variable optimization single algorithm is introduced later. At the same time, compared with questionnaires and interviews, through the analysis of online comment text mining, the hot topics and emotional attitudes of citizens, enterprises and the government can be excavated in a more real and direct way, and the interference of subjective factors and uncontrollable objective factors can be eliminated.

In general, this paper analyzes and optimizes the principle of the current intelligent single algorithm based on the theory of multiple co-governance. The theoretical model is applicable to the governance of different algorithms, which is different from the traditional theoretical model in the following aspects:

(1) The traditional theoretical models, such as the IAD analysis framework, lacks collaborative governance and the interaction of the overall political and economic environment, and on the basis of SFIC multiple cooperative governance model took into account the resources and the asymmetry of information, analysis of the dimension will be more abundant. For example, consider the security incident frequent, public concerns about the environment factors, such as, etc.

(2) The newly constructed multiple collaborative governance model can be applied to a variety of algorithms, such as single-dispatch algorithm and personalized recommendation algorithm, etc. However, traditional theoretical models, such as resource dependence theory, are still abstract and conceptualized, without specific processes or measures, which are difficult to put into practice.

(3) The multiple collaborative process is a cyclic process closed loop, the face-to-face conversations and trust-building discussions shall be conducted after certain periodic results are achieved, so as to ensure that the final algorithm governance is rich and effective, and there is no such cyclic verification for traditional theoretical models.

(4) However, the newly constructed multiple co-governance model ignores the research on the consequences of co-governance. In the SFIC model, the cooperative consequences are regarded as the product of the cooperative process and only appear in the schematic diagram but are not discussed in detail. As a matter of fact, synergetic consequences can be expressed in various forms, and can be summarized, modified and upgraded in the following process.