

大数据驱动的地铁众包快递系统

张坤芳, 鲁鸣鸣, 郑林

(中南大学信息科学与工程学院, 湖南 长沙 410083)

摘 要: 近年来, 同城快递的需求增多, 涌现出一些提供 1 h 或 3 h 送达的快递公司, 如闪送等。然而, 其服务质量缺乏量化分析, 常常会出现全职/兼职快递员不接单或接单后没有按时投递的情况。为此, 提出了一种地铁众包快递系统, 能够基于地铁乘客出行历史记录分析快递服务质量, 并提供分时段的快递差异化服务。该系统既满足时延要求较高的快递需求, 又可作为同城范围快递的补充。同时, 提出快递中转方案。实验证明此中转方案是高效的。

关键词: 地铁; 众包; 快递; 大数据

中图分类号: TP39

文献标识码: A

Big data based metro crowd delivery system

ZHANG Kun-fang, LU Ming-ming, ZHENG Lin

(School of Information Science and Engineering, Central South University, Changsha 410083, China)

Abstract: In recent years, the demand for urban express delivery experiences fast growing. Several startup delivery corporations, such as iShansong, have provided the delivery services, which are promised to deliver a parcel within 1 or 3 hours in a city. However, since the quality of the promised services have not been fully quantitatively analyzed based on real data, it is quite often that couriers refuse to pick up delivery requests or the accepted parcels fail to be delivered on time. To address the above issue, a metro crowd-delivery system was proposed, which could utilize the historical records of metro passengers to analyze the quality of the delivery service and provide differential service within different period of time. The system not only meet the requirements of high delay requirements of the courier, but also as a city-wide express supplement. At the same time, a courier transit program was proposed. Experiments show that this transfer program is efficient.

Key words: subway, crowd sourced, express delivery, big data

1 引言

随着城市生活水平的日益提高以及电子商务和快递行业的快速发展, 在同一个城市内的快递需求日益增多。现在的快递公司一般雇佣快递员在城市公共道路上利用机动车进行同城范围的快递投递, 然而, 这种方式存在很多的问题。首先, 逐渐增长的快递需求使快递公司不得不雇佣更多的快递人员, 同时购买或租用更多的交通设备。这不仅

增加了快递公司的运营成本和社会的资源消耗, 也进一步恶化了本已比较拥堵的城市交通状况。其次, 在销售的旺季, 爆发式增长的快递需求远远地超出了快递公司的投递能力, 造成了较长时间的投递延迟。最后, 现有的同城快递投递时延的偏差往往比较大, 对于生鲜、快餐等投递时延要求比较高的同城快递, 投递时延的不确定性往往造成较差的用户体验。尽管国内已出现一些新型同城快递公司, 如闪送、人人快递、快速递等, 能提供 1 h

收稿日期: 2017-09-20

基金项目: 国家自然科学基金资助项目 (No.61232001, No.61173169, No.91646115, No.60903222); 湖南省自然科学基金资助项目 (No.2016JJ2149); 湖南省战略性新兴产业重大科技攻关项目基金资助项目 (No.2012GK4054)

Foundation Items: The National Natural Science Foundation of China (No.61232001, No.61173169, No.91646115, No.60903222), The Natural Science Foundation of Hunan Province (No.2016JJ2149), Major Science & Technology Research Program for Strategic Emerging Industry of Hunan (No.2012GK4054)

或 3 h 内送达的快递业务,但其承诺的服务缺乏基于真实数据量化分析,很容易造成无法按时投递的事故。

为了解决上述问题,本文提出了一个基于地铁网络的众包快递系统。该地铁众包快递系统可基于地铁乘客历史乘车数据统计每一段投递 OD 站点对(即起始(originator)和目标(destination)对)在不同时段的投递时延分布,从而估算出每个 OD 对在不同时段的投递服务质量。基于随时段变化的投递服务质量分析,地铁众包快递系统能够为不同用户提供差异化快递服务,既可以满足时延要求比较高的快递需求,又可以作为现有的同城范围快递的补充。

在该系统中,注册用户首先发起同城快递请求,快递公司收取包裹后对每一个包裹计算传递路径,然后向注册的地铁乘客发布快递投递请求,乘客在乘车前挑选符合自己出行路线的包裹,携带传输到自己的目的地,并通过转发或直接投递的方式最终将快递投送到目的车站,随后通过快递员进行最后一段的投递。完成投递的乘客可以得到相应的一笔费用。

相对于现有的同城范围的快递投递方案,本文设计的基于地铁网络的众包快递系统主要有以下的优点:相比于城市道路上的快递运输,利用地铁网络传输快递能够满足时延性要求较高的快递需求。同时,基于地铁人流规律的分析给出的投递线路减少了投递时间的波动性;巨大的地铁快递传输容量能够应对“双 11”等突发式的业务需求;通过建立合适的中转模型,本系统能预估投递方案的延迟和相应时间内成功投递的概率,量化投递指标。

本文的主要贡献有:提出了一个通过地铁乘客以众包的方式帮助投递快递的系统;基于上海地铁乘客乘车记录数据分析了地铁众包快递系统的服务质量以及可行性;为了提升地铁众包系统服务质量,提出了一种快递中转投递方案。

2 相关工作

共享经济是一种基于陌生人并且存在物品使用权暂时转移的一种新的经济模式。通过整合闲散社会资源,共享经济不仅给分享者带来收益,也能减少资源消耗。Ma 等^[1]早在 2013 年就详细阐述了城市出租车共享系统的实现和调度算法,Uber、滴滴等打车软件的广泛使用已经证明了该共享模式在当今时代的可行性。文献[2,3]特别研究了出租车拼车问题,并分别提出了高效的车辆

调度方案,能够减少总行驶路程,降低资源消耗。在城市交通设备的共享方面,越来越多的城市建立了有桩自行车共享系统,同时多家公司已经开始铺路共享单车市场。文献[4]研究了城市自行车停靠点的最佳位置。Yang 等^[5]对自行车共享系统的人流移动进行建模并提出了人流预测算法。Li 等^[6]通过对自行车站进行聚类,发掘不同类别之间和类别内部的车辆流动规律从而推断车辆租借情况,利用两者的研究工作能够提前进行车辆的资源调度,解决资源不平衡问题。

城市交通数据还被应用到其他研究领域。文献[7]通过分析人们的日常出行提出了地铁中的广告投放策略。Zhang 等^[8]结合了不同类型的城市交通设施,发掘出城市人流的出行规律和城市的热点区域。在快递运输方面,Zhang 等^[9]通过研究快递员的动态调度方式增加同城快递的效益。与此同时,国内已经出现了一些基于共享经济的同城快递服务,如闪送、人人快递、快速递等。用户可以通过这些软件在线提交投递请求,投递人员可以通过软件抢单成功后上门收取快递,然后自行选择交通方案进行投递。这些服务虽然利用了闲散的人力资源,但是由于缺乏对数据的详细分析和对服务质量的量化,容易造成无法按时投递的事故。

现有针对城市范围内的路由算法主要是基于概率模型或是最短时延的方法,如文献[10]研究了城市道路上基于最大概率的数据传输方法,文献[11]提出城市出租车可以构成一个 DTN (delay tolerant network)^[12],并研究了利用出租车轨迹数据完成城市范围内的消息传递的车载网络。由于在快递投递过程中还要考虑到成本因素,这些方法不能够直接适用于地铁快递系统。实际上由人流构成的地铁网络是一个特殊的 DTN,车站是网络中不同的节点,人流是网络中的边,由于人流量随时间变化的特性使得其具有时间演化图的性质^[13],现有的对时间演化图的研究主要集中在节点的可达性^[14]以及基于最短延迟研究图中的最短路径^[15],没有针对时延、概率和成本因素进行统一分析的合适方案。

3 系统设计

本文设计的基于地铁网络的众包快递系统整体结构如图 1 所示,主要包括 4 个部分:安装在乘客智能终端上的应用程序、建设在部分地铁车站中的中转系统、一个后台系统以及若干个快递员。

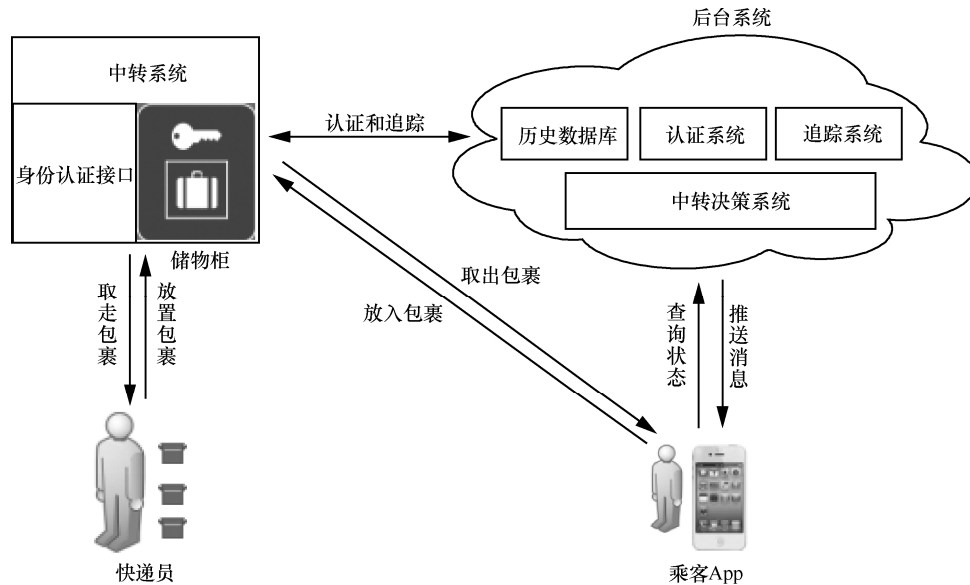


图 1 系统结构

3.1 快递员

因地铁站点无法覆盖整个城市，再加上成本等因素的考虑，中转站点在开始的时候只覆盖了部分热门地铁站点，因此，系统提供快递员上门收件的服务，并负责将包裹投递到最近的地铁中转站。此外，快递员还负责将快递从投递的目标站点取出并上门送给客户。对每个包裹而言，它的起始站点是发件位置的最近中转站，而目的站点是离收件位置最近的地铁站。

3.2 中转系统

中转系统主要用于和快递员以及参与投递的乘客进行交互，包括身份认证接口和储物柜 2 个部分。身份认证接口主要用于提取用户认证信息并和后台服务器验证，只有通过认证的用户才能够放置或取走包裹。储物柜是一个临时的包裹安置区域，包裹的收取和中转只能依托于储物柜进行，具有方便性和安全性。考虑到成本因素，中转系统一开始并不会覆盖所有的地铁站点，而只是覆盖一些热门的站点。具体中转系统的数量和部署方案参见 6.1 节的实验分析。

3.3 移动应用

安装在用户手机上的移动应用主要用于和地铁快递系统进行交互。下载应用并通过实名认证后，地铁乘客即可通过应用接收投递任务。接收投递任务的乘客在经过有中转系统的地铁站点时通过中转系统站即可取出包裹，此时，系统会将乘客和包裹进行绑定，并通过应用追踪包裹状态，以便寄送包裹的用户查询并防止包裹丢失。

3.4 后台系统

后台系统主要用于用户认证、包裹投递路径的选择以及包裹追踪，包含以下 4 个组成部分：历史出行数据库、认证系统、中转决策系统和追踪系统。

历史出行数据库：主要用于存储乘客的日常出行轨迹信息并挖掘出其中隐含的移动规律，为中转决策提供数据支持。

认证系统：该部分事先存储参与投递的乘客的实名认证信息，用于包裹在中转系统取出之前的身份确认。

中转决策系统：该部分主要是基于地铁乘客历史出行记录，在权衡快递时效性和经济效益的基础上选择最优包裹投递路径。关于中转决策系统的必要性分析和中转决策算法的设计分别参见 4.5 节和 5.2 节。

追踪系统：为了避免包裹的丢失，系统会为每个投递的包裹产生一个 ID，用来追踪包裹的状态。在投递过程中，包裹将会与参与投递的乘客进行绑定，并通过乘客的智能手机上的应用实时更新其状态。

3.5 投递过程

包裹投递过程如图 2 所示。

- 1) 用户发出寄件通知，快递公司取件填写信息，后台生成唯一 ID。
- 2) 快递员或用户直接投递包裹到中转系统的储物柜。
- 3) 地铁乘客收到 App 的推送包裹信息，选择合适包裹取出进行传递。
- 4) 乘客到达目的地后将包裹放入储物柜或交

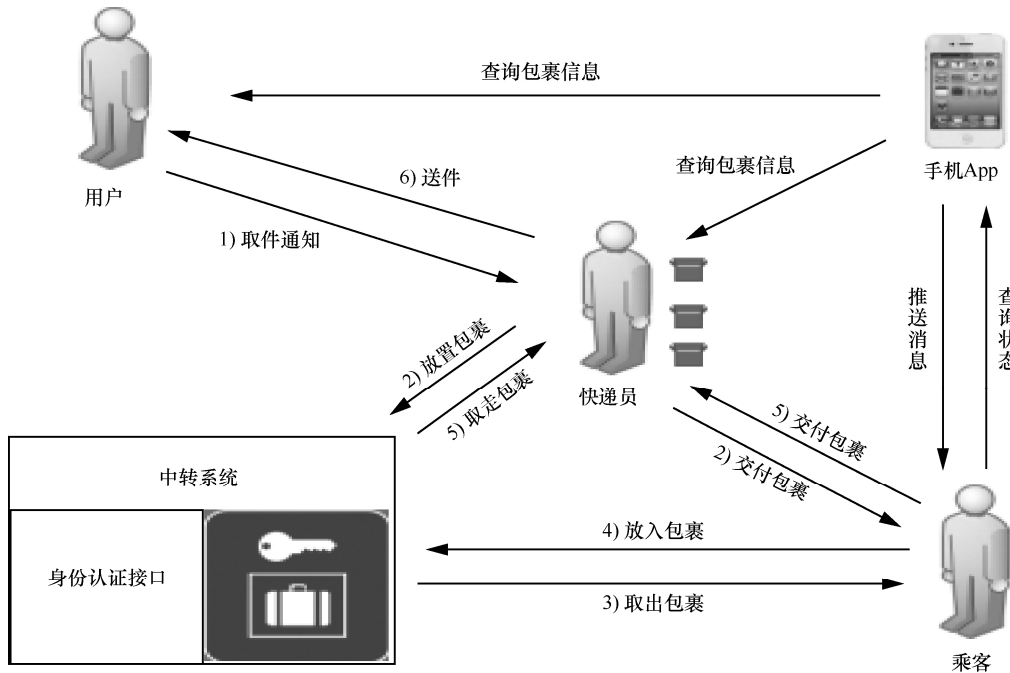


图 2 包裹投递过程

给快递员，储物柜中的包裹若已到收件地点附近则由快递员或用户取出，若仍要进行传递则跳至 3) 继续这一过程。

5) 快递员取出包裹。

6) 快递员将包裹送至收件地点或用户自己到地铁站取件。

快递传递过程中，凡是认证的用户（寄件人、收件人、乘客、快递员）都可通过 App 查看包裹的详细信息。

4 数据驱动的地铁众包快递分析

4.1 数据集

本文所用的数据是上海市 2015 年 4 月地铁一卡通刷卡数据（4 月 6 号的数据由于统计不完全从数据源中剔除）和地铁站信息。上海地铁的车站个数是 289，所以总共有 8 万多个 OD 站点对。将每天同一张卡的刷卡记录按时间排序，并将前后 2 条合并得到一个出行 OD，同时去除误差数据和异常数据（独立的单条记录和 OD 时间明显过长或过短的记录），最终得到每个工作日的地铁出行 OD 总数，如图 3 所示。

从图 3 可明显看出，上海地铁工作日的每天人流量为 420 万~480 万。巨大的人流量表明通过地铁传输货物绝不是点缀，如果每个乘客都能够传递一件包裹，那么每天通过地铁可以传递 400 万件以上包裹。

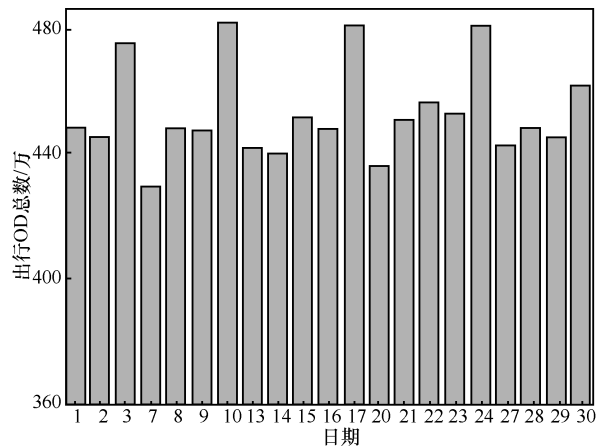


图 3 出行 OD 总数

4.2 投递时间

投递包裹最主要的衡量指标是投递时间，现有的同城快递一般提供 1 h、3 h 或当天到达的服务。地铁快递的投递时间主要受传递路径的影响，本文主要关心包裹在地铁网络中传递所需的时间，是指快递从始发站点到目的站点所用的时间。地铁中的快递投递是分段式的，每一段的传输时间主要由以下 2 个因素决定。

1) 等待时间: 放置在中转站点的包裹等待乘客提取所需的时间。

2) 运行时间: 包裹被提取到乘客再次把包裹放置到目标站点或下一个中转站点所需的时间。

为了分析等待时间和运行时间的变化规律，本文针对不同站点 OD 对在一天不同时间段的平均等待时间和平均运行时间，用地铁刷卡数据进行了估算。平均等待时间可利用车站乘客的到达速率进行估计，因为通过统计发现，每个 OD 对在一天各个不同时段内的人流量均服从泊松分布且等待下一个乘客的时间服从指数分布。

运行时间实际上可细分为进站时间、等车时间、行驶时间、出站时间，分别表示乘客携带包裹进站到等车地点的时间、乘客等待列车到达站点的时间、乘客上车到列车抵达目的地的时间、乘客下车到出站口投递包裹的时间。上述 4 种时间可用乘客同一段旅程中进站和出站刷卡的时间差来估算。

图 4 和图 5 分别给出 2 个人流量和发车频率有较大差异的站点 OD 对（莘庄站日均进站人流量超过 6 万人次，南浦大桥站日均进站人流量约 1 万人次，因此，莘庄站的发车频率要高于南浦大桥站）的平均等待时间和平均运行时间在一天内不同时段随时间变化的曲线，其中平均等待时间通过车站乘客的到达速率进行估计，如 1 h 有 20 个乘客进站，则平均等待时间是 3 min。平均运行时间用同一时段

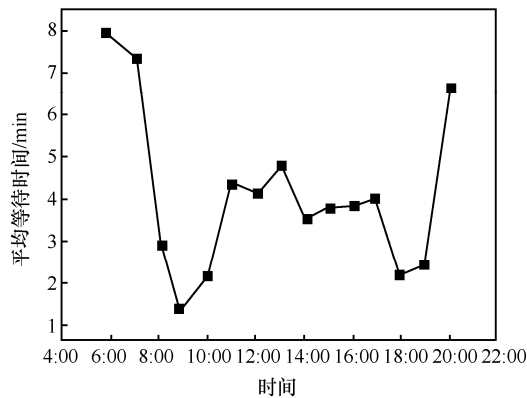
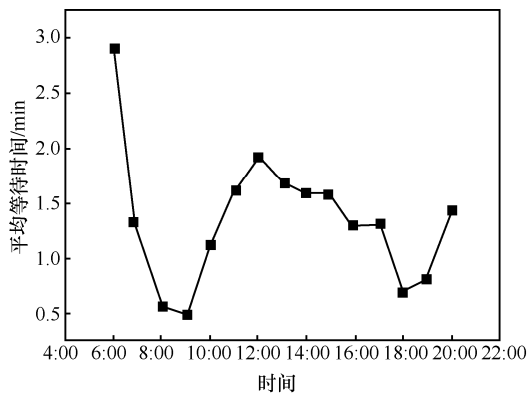
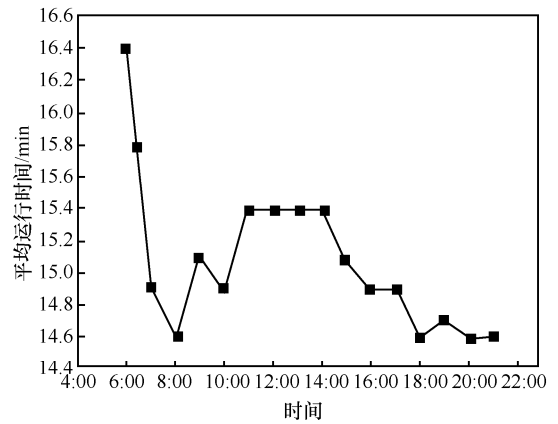
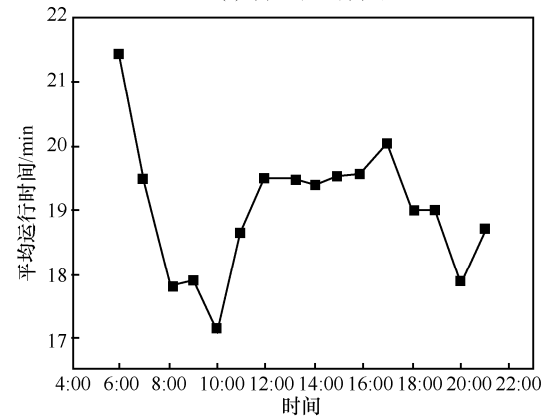


图 4 平均等待时间



(a) 莘庄到上海南站



(b) 南浦大桥到人民广场

图 5 平均运行时间

同一段旅程不同乘客进站和出站刷卡的时间差的均值来估算。从图中可看出：无论是等待时间还是运行时间均在一天在不同时段呈显著的变化，而且不同 OD 对的变化规律不尽相同。这是因为等待时间主要受人流量影响，人流量越大则等待时间越短，而运行时间的主要决定因素是站点之间的实际距离和列车的调度时间。

综上所述，等待时间和运行时间会受到人流量、列车调度方案、车站地理位置等因素的影响。所以，需要针对每个不同的 OD 单独提取随时间变化的规律，而不是简单地用一个模型描绘所有 OD 的规律。具体的设计方案参见 5.1 节。

4.3 服务质量量化

衡量同城快递服务质量的一个关键指标是投递时延，由于缺乏对现实数据的详细分析，现有的同城快递如闪送、人人快递、快速递等都只是给出一个笼统的投递时延，如 1 h 或 3 h，造成的后果是快递人员发现无法在规定时间内送到的话就选择不接单。另外包裹的投递时间实际上受到各方面的影响，如交通拥堵。相对于城市公共道路的交通设备，

虽然利用地铁进行快递投递的时延较为稳定，但仍然会有一定的波动导致投递失败。例如，由于等待合适乘客到站的时间不可预测，造成最终的投递时间波动。

图 6 描绘了从金海路到航津路在一天的不同时间段传递一个包裹的平均时间和对应的标准差。从图 6 可以看出，每个小时内的平均投递时间都有一定的偏差。如果仅用平均投递时间作为服务质量评价指标，从概率上来讲，很可能存在投递不及时的情况，而且很难刻画投递不及时的概率和时延的程度。类似地，用历史最长投递时间作为评价指标，也很难 100% 保证不会出现超时投递的状况，而且可能会导致本来大部分可以在给定时延内投递成功的快递无法通过这种方式投递。

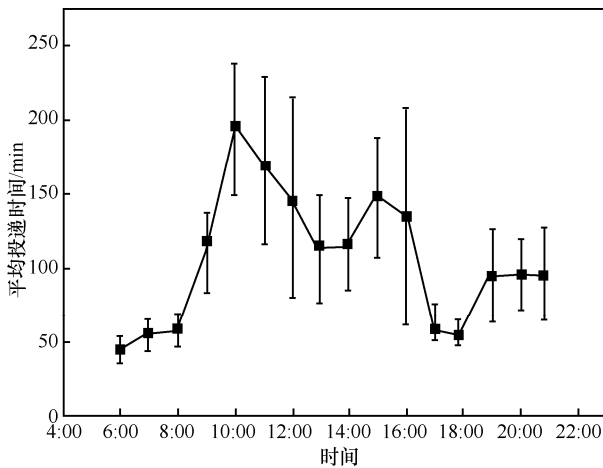


图 6 平均投递时间和标准差（金海路到航津路）

为此，在给定预计的投递时间的同时，快递公司也应该清楚该投递时间潜在的风险并进行量化，即将投递时间的波动性进行量化，这种量化可以通过一定的概率体现。实际上，根据切比雪夫不等式，投递时间的波动性与相应的成功投递的概率存在相关性

$$P(|X - E(X)| \leq \epsilon) \geq 1 - \frac{D(X)}{\epsilon^2} \quad (1)$$

其中， $E(X)$ 和 $D(X)$ 分别表示随机变量 X 的期望和方差。切比雪夫不等式描述了数据集在均值一定范围内的概率分布，对于相同的 X 取值范围，当方差越大，数据越分散，计算出的概率 P 越小；反之概率 P 越大。在快递投递的概率计算中，假定预计的投递时间是 C ，某个路径的投递时间 X 的期望是 $E(X)$ ，方差是 $D(X)$ ，如图 7 所示，只考虑 C 大于 $E(X)$ 的

路径，令 $C = E(X) + \epsilon$ ，则

$$P(X \leq C) = P((X - E(X)) \leq \epsilon) \quad (2)$$

$$> P(|X - E(X)| \leq \epsilon) \geq 1 - \frac{D(X)}{\epsilon^2}$$

基于式(2)，对于一条传输路径，在给定一个预计传输时间 C 的情况下，可以通过计算求得该条路径能够在 C 时间内成功送达的概率，并把该值作为该路径在 C 下的预计投递成功率。

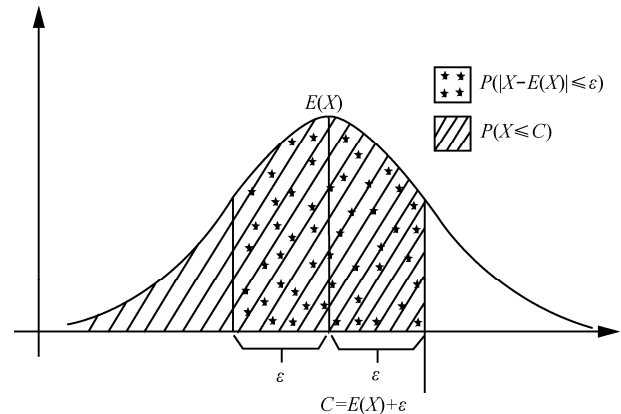


图 7 概率密度

4.4 时间可变性分析

图 6 不仅反映了传输时间的波动，也反映出传输时间是随时间变化的，高峰时期平均投递时间较低，方差也比较小，而中午时段不仅投递时间长而且波动也大。显然，在相同的预计投递时间下，高峰时的预计投递成功率要高于中午。传输时间的波动也会造成某个 OD 在选择较短时间区间内能满足投递需求，而较长时间区间内却会失败。例如，图 6 在早高峰时期段人流量较多，此时，投递时间的方差较小，能够符合传递需求。而如果扩展时间区间把一上午的记录考虑在内，由于高峰期过后的平均投递时间和方差都比较大，导致最终计算出投递时间的整体期望和方差都可能较大，不符合投递需求。因此，需按一天的时间区间给出不同的投递方案。

4.5 中转的必要性

虽然地铁网络的人流量巨大，但是实际上地铁网络的人流量主要集中在少部分的 OD 站点对中，大多数的 OD 每天的人流量非常少。也就是说，在大多数情况下，包裹直接被投递到目的地的概率很低甚至为 0。图 8 统计了上海地铁每个 OD 站点对的日均人流量，并描绘出了站点个数随人流量变化的累积分布概率。从图 8 中可以看出，大约有 95%

的站点对的日均人流量少于 250 人次，进一步观察这些人流量较少的站点对，如图 9 所示，可以看到有超过 4 万个 OD 对的人流量少于 20 人次。因此若想通过这些站点对的稀疏人流进行包裹的直接投递显然是不现实的。

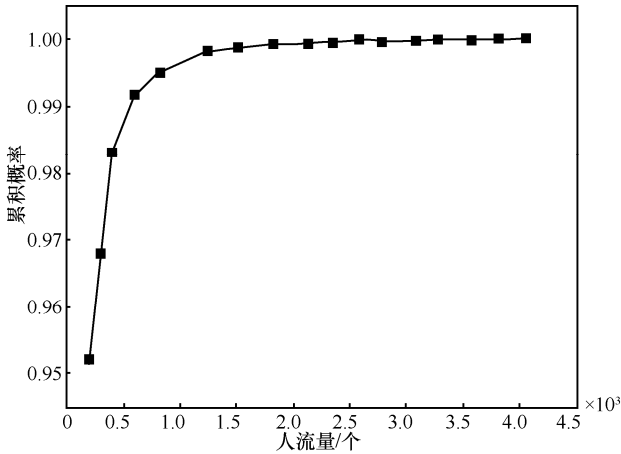


图 8 OD 站点对累积分布概率

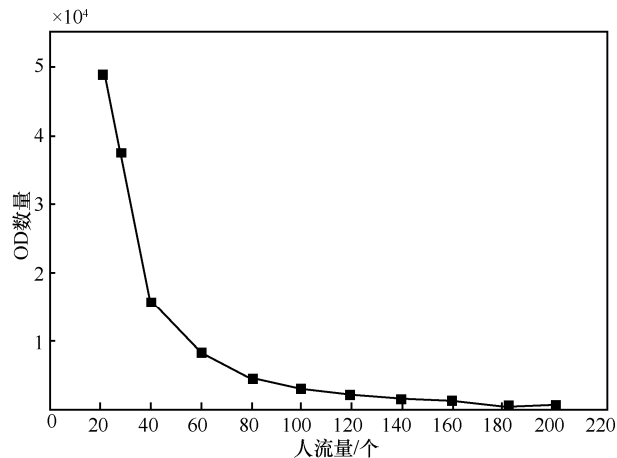
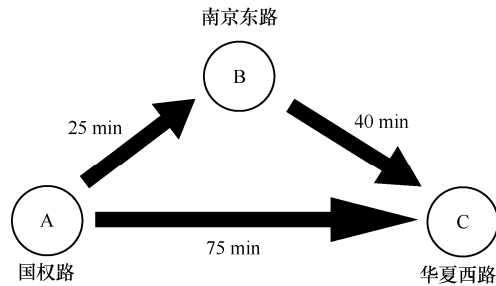


图 9 低人流的 OD 站点分布

另外，为了满足服务质量需求的量化概率指标，加入中转站点比直达的指标效果更好。图 10 显示了从国权路到华夏西路直达和通过南京东路中转的 2 种情况，由于直达的人流量过少，等待乘客投递的时间过长，导致 AC 的传输时间大于 AB+BC 的传输时间。若给定预计的投递时间为 90 min，由于每段路径是相互独立的，AB→BC 路径的期望和方差分别是各段子路径的总期望和总方差，根据式(2)可计算出 AC 和 AB→BC 的预计投递成功概率分别为 15.6%和 75.5%。不管是在预计平均时间还是预计投递成功率方面，直达的方案都明显比通过南京东路中转的方案要差。



OD 站点对	期望	方差
AC	75	190
AB	25	98
BC	40	55

图 10 中转投递

通过上述分析可以发现：对于大多数的站点对，包裹不能够在任意时刻不通过换乘直接投递。另外，由于投递时间的波动性，包裹有时候需要通过中转的方式达到更高的投递可能性。因此，必须设计一个适用于地铁人群投递系统的中转策略。

5 中转投递决策

5.1 中转投递的目标

如第 4.5 节所述，某些情况下中转可能比直达更能满足服务质量需求。而本文系统每次传递的激励金额固定，不受传递距离的影响，一个包裹每增加一次中转，就要多占用一次中转系统的空间资源和增加一次激励成本。所以对快递公司而言，为了节省成本，每个包裹的中转次数应该越少越好。同时为了保证服务质量，投递成功的概率应该不能过低。投递成功的概率和预计的投递时间也有很大的关系，如果给定的预计投递时间越短，根据式(2)计算出的投递成功率会越小；反之，成功率较高，但不能体现地铁快递系统快速送达的优势。所以为了能够同时保证服务质量和成本，快递公司在计算传递路径时应该遵循以下两个准则。

- 1) 尽可能地减少中转次数。
- 2) 在保证一定的投递成功率的同时尽可能减少预计传输时间。

5.2 中转决策算法

5.2.1 算法设计

根据上述的分析，给出中转决策系统的设计方案，如下所示。

首先，由于投递时间期望和方差随时间变化的特性，必须将一天划分成不同的时间区间并分别计

算路径。考虑时间区间反映的信息以及期望和方差的计算，区间长度选取要适当。本文根据历史数据 0.5 h 为区间划分时间，统计每个 OD 在不同区间 Z^{od} 的投递时间的期望和方差。 $Z^{od} \cdot s$ 和 $Z^{od} \cdot e$ 分别是区间起始和结束时间； $Z^{od} \cdot we$ 、 $Z^{od} \cdot wd$ 、 $Z^{od} \cdot re$ 和 $Z^{od} \cdot rd$ 分别表示该区间内等待和运行时间的期望和方差；区间投递时间的期望 $Z^{od} \cdot de$ 和方差 $Z^{od} \cdot dd$ 分别可用 $Z^{od} \cdot we + Z^{od} \cdot re$ 和 $Z^{od} \cdot wd + Z^{od} \cdot rd$ 表示； $Z^{od} \cdot r$ 表示区间的传递路径。如果某个区间的 $Z^{od} \cdot we$ 过大，即人流过少，可以认为当前区间内不可达并从中删除。这样生成的结果是所有能够一跳直达的 OD 的区间路径信息，称为 *directZoneRoute*。

其次，为了尽可能减少预计传输时间，从 1 h 开始逐渐增加预计传输时间并重新搜索，直到某个预计传输时间能得到一个满意的投递成功率。

最后，为了最小化中转次数，可以通过逐跳增长的方式计算更多跳数的每个区间的路径。由于每一层的计算都可以基于前一层少一跳的路径信息，例如，计算 3 跳可达的路径可在 2 跳的基础上扩展一跳。如果一个区间在较少跳数搜索中已经满足需求，则可不再次搜索。

综上所述，地铁快递网络中的中转方案设计问题可以定义为给定一个起始站点 O 和目的站点 D，预计的传输时间 C，如何选择合适的传递路径使包裹能够以一定的概率 P 在预计的时间内送达目的地。问题的输出是满足给定约束条件的一天中所有可以送达的时间段以及相应的中转路径。

5.2.2 算法过程与复杂度分析

根据上述的分析，给出中转决策算法的详细过程，如算法 1 所示。

算法 1 中转决策算法

输入 预计传输时间 C，预计投递成功率 P，中转站点集合 S，最大跳数 *max_HOP*，直达的区间路径信息 *directZoneRoute*。

输出 每个 OD 满足需求的所有时间区间及其对应的传递路径。

初始化 首先根据 C、P 和式 (2) 去除 *directZoneRoute* 中预计投递成功率小于 P 的区间，并定义前一层的路径存储结构 *baseZoneRoute*，初始化为 *directZoneRoute*。

- 1) $hop \leftarrow 2$;
- 2) while $hop < max_HOP$ do:
- 3) 定义当前跳数的路径存储结构

currentZoneRoute;

- 4) 筛选出 *baseZoneRoute* 还未包含的区间集合 \mathcal{Z} ;
- 5) for \mathcal{Z} 中的每一个区间 Z^{od} do:
- 6) $P_m = 0$ // 表示当前最大的预计成功率
- 7) for S 中的每一个中转站点 m do:
- 8) if *baseZoneRoute* 包含 Z^{om} and $Z^{om} \cdot s == Z^{od} \cdot s$ and $Z^{om} \cdot e == Z^{od} \cdot e$ and $Z^{om} \cdot r \cdot length == hop$ then:
- 9) 计算 Z^{om} 到达 m 的时间 t;
- 10) if *directZoneRoute* 包含 Z^{md} and $Z^{md} \cdot s \leq t \leq Z^{md} \cdot e$ then:
- 11)
$$P' = 1 - \frac{Z^{om} \cdot dd + Z^{md} \cdot dd}{\left(C - (Z^{md} \cdot de + Z^{md} \cdot de)\right)^2}$$
;
- 12) If $P' \geq P$ and $P' \geq P_m$ then:
- 13) $P_m = P'$;
- 14) $Z^{od} \cdot de == Z^{om} \cdot de + Z^{md} \cdot de$;
- 15) $Z^{od} \cdot dd = Z^{om} \cdot dd + Z^{md} \cdot dd$;
- 16) $Z^{od} \cdot r = Z^{om} \cdot r + Z^{md} \cdot r$;
- 17) end if
- 18) end if
- 19) end if
- 20) end for
- 21) if $P_m \geq P$ then:
- 22) 把 Z^{od} 加入到 *currentZoneRoute* 中;
- 23) end for
- 24) 把 *baseZoneRoute* 中的所有区间加入到 *currentZoneRoute* 中;
- 25) $baseZoneRoute \leftarrow currentZoneRoute$ ，重置前一层的路由信息，以便下一层搜索;
- 26) $hop \leftarrow hop + 1$;
- 27) end while
- 28) return (*baseZoneRoute*)

算法在初始化阶段去除了 *directZoneRoute* 中不满足需求的所有区间，处理后的 *directZoneRoute* 保存了所有直达的满足需求的区间。算法在每次的循环中依次增加最大跳数限制，每次循环称为一层迭代过程。为了降低算法的时间复杂度，每一层迭代中满足需求的所有区间将会保留，并在下一层迭代中用于计算剩余未满足需求的区间。

在每一层的迭代中，首先筛选出所有 OD 当前还未满足需求的所有时间区间。如果站点 src 在时

间区间 $beginZone$ 通过 2 跳能到达中转站点 mid ，到达的时间区间为 $midZone$ ，同时中转站点 mid 能够在时间区间 $midZone$ 直达到 dst ，那么可以说 src 能在时间区间 $beginZone$ 经过 mid 到达 dst ，跳数为 3，并计算新路径的期望、方差以及预计投递成功率。如果一个区间在一次搜索中有多条路径能够从 src 到 dst ，算法只保留预计投递成功率最大的路径。

算法每次迭代都把 $baseZoneRoute$ 中的所有区间加入 $currentZoneRoute$ 是为了方便最终的输出，每次迭代都会增加 $baseZoneRoute$ 的记录，当程序结束之后 $baseZoneRoute$ 中就保存了所有区间的路径。由于在生成新路径的同时也要检查跳数，所以增加旧的区间不会对最终结果有影响。

对于一个 OD 的某个区间的搜索，每次都是在前一层的基础上增加一个中转站点并重新计算整条路径是否满足需求。如图 11 所示，乘客乘车是相互独立的事件，即

$$\begin{aligned} Z^{ac} \cdot de &= Z^{ab} \cdot de + Z^{bc} \cdot de \\ Z^{ac} \cdot dd &= Z^{ab} \cdot dd + Z^{bc} \cdot dd \end{aligned} \quad (3)$$

在给定预计投递时间为 180 min 的情况下，A→B→C 的传递路径能够达到 92.2% 的预计投递成功率。

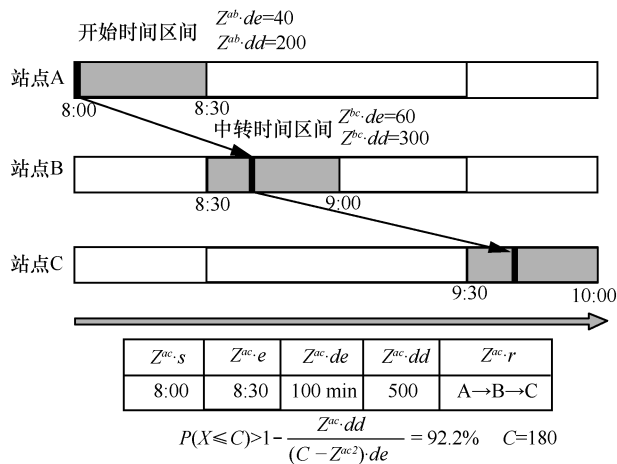


图 11 区间的路径搜索和概率计算

决定中转决策算法复杂度的主要因素是中转站点集合的大小 $\|S\|$ 和车站站点集合的大小 $\|\bar{S}\|$ 。地铁快递系统的起点只能是中转站点，总的 OD 个数为 $\|S\| \cdot \|\bar{S}\|$ ，由于在一层迭代过程中对每个区间的搜索都需要获得预计成功率最大的线路，复杂度为 $O(\|S\| \cdot \|\bar{S}\|)$ 。虽然算法整体的复杂度较高，但是由于中转站点集合的规模较小，实验表明 10 个中转

站点就能达到较好的效果，同时，总的车站个数只有 289，在实际的计算过程中延迟较小。更为重要的是，本文提出的中转决策算法是一种基于历史出行记录的离线计算算法，通过提前进行线路计算并缓存的方式能够使其应用于实时系统中。

5.2.3 后续处理

经过上述的计算， $baseZoneRoute$ 中保存了不同 OD，能够满足时间和概率需求的所有时间区间的路径信息。一些相邻时间区间的路径相同，可以将这样的区间合并成一个大的区间。如

07:00:00 07:30:00 国权路 南京东路 高青路

07:30:00 08:00:00 国权路 南京东路 高青路

合并之后变为一个长度为 1 h 的区间

07:00:00 08:00:00 国权路 南京东路 高青路

最终一个 OD 站点对的输出如表 1 所示。

表 1 路径输出

时间段	传递路径
07:00:00 08:00:00	国权路 南京东路 高青路
15:00:00 17:30:00	国权路 陆家嘴 高青路
17:30:00 19:00:00	国权路 南京东路 高青路

6 实验

地铁快递系统的效益受到多方面因素的影响，其中主要的影响因素有：中转站点的个数，参与投递的乘客比例和给定的预计投递成功率。本节将根据这 3 个参数对地铁快递系统的效益影响进行实验分析，主要的评价指标有以下 3 个参数。

1) 平均中转次数 V_{tran} ：侧面反映了成本。

2) 平均覆盖时间 V_{time} ：平均每个 OD 站点对在一天内有多长的时间能够成功传递包裹。如果一个 OD 站点对一天内只有半小时可以成功投递，显然没有多大意义。越长的平均覆盖时间表示系统越大的可用性。

3) OD 覆盖百分比 V_{ratio} ：能够通过地铁快递系统进行传递的 OD 站点对占有所有 OD 站点对的百分比，该值越大表示地铁快递系统能够覆盖的城市范围越广。由于 OD 的起始站点必须是中转站点，这里的所有 OD 站点对只包含起始站点是中转站点的 OD。

设中转站点集合为 S ，车站集合为 \bar{S} ，可满足的区间集合为 Z ，上述 3 个指标的计算规则分别如式(4)、式(5)和式(6)所示。

$$V_{tran} = \frac{\sum Z^{od} \cdot r \cdot length - 1}{Z \cdot size} \quad (4)$$

$$V_{\text{time}} = \frac{\sum Z^{od} \cdot e - Z^{od} \cdot s}{\sum f(od)} \quad (5)$$

$$V_{\text{ratio}} = \frac{\sum f(od)}{S \cdot \text{size}(\bar{S} \cdot \text{size} - 1)} \quad (6)$$

$$f(od) = \begin{cases} 1, \exists Z^{od} \in Z \\ 0, \forall Z^{od} \notin Z \end{cases}$$

$$o \in S, d \in \bar{S}, o \neq d$$

6.1 中转站点的个数

为了对比不同中转站点个数对地铁快递投递的影响，需要固定乘客参与率和投递成功率。假设有 10% 的乘客参与投递，最低的投递成功率是 95%。

首先，把 OD 的人流量作为选取中转站点的决定因素，并在这些站点的基础上进行扩展，根据地理位置信息额外选取一部分中转站点。这些增加的中转站点的主要功能是增加覆盖范围，以减少快递公司到中转车站的成本。

根据 OD 的人流量，分别选取了 10、20、30、40 和 50 个中转站点测试地铁快递系统的效果。对于每组中转站点，分别统计能够满足 1~6 h 送达的不同需求，预计投递成功率不低于 95% 的所有 OD 站点对，实验结果如图 12 所示。其中，横坐标均表示 1~6 h 的预计送达时间，纵坐标分别表示对应某个送达时间的平均中转次数、平均覆盖时间和 OD 覆盖百分比。

从图 12 中可以看出，仅选择 10 个中转站点和选择 50 个中转站点的实验结果非常类似，只是在平均中转次数上有略微的差异；3 h 的预计送达时间下的平均中转次数最高；近 90% 的 OD 对能够在 4 h 内送达，此时的平均覆盖时间达到了 10 h。

上述实验结果还反映出中转决策算法选择出的中转站点主要集中在少部分站点，快递公司在初始阶段建立 10 个中转站点就能够达到比较好的效果，但是这些站点主要分布在城市的中心区域，覆盖的城市范围较小。因此，快递公司还必须根据车

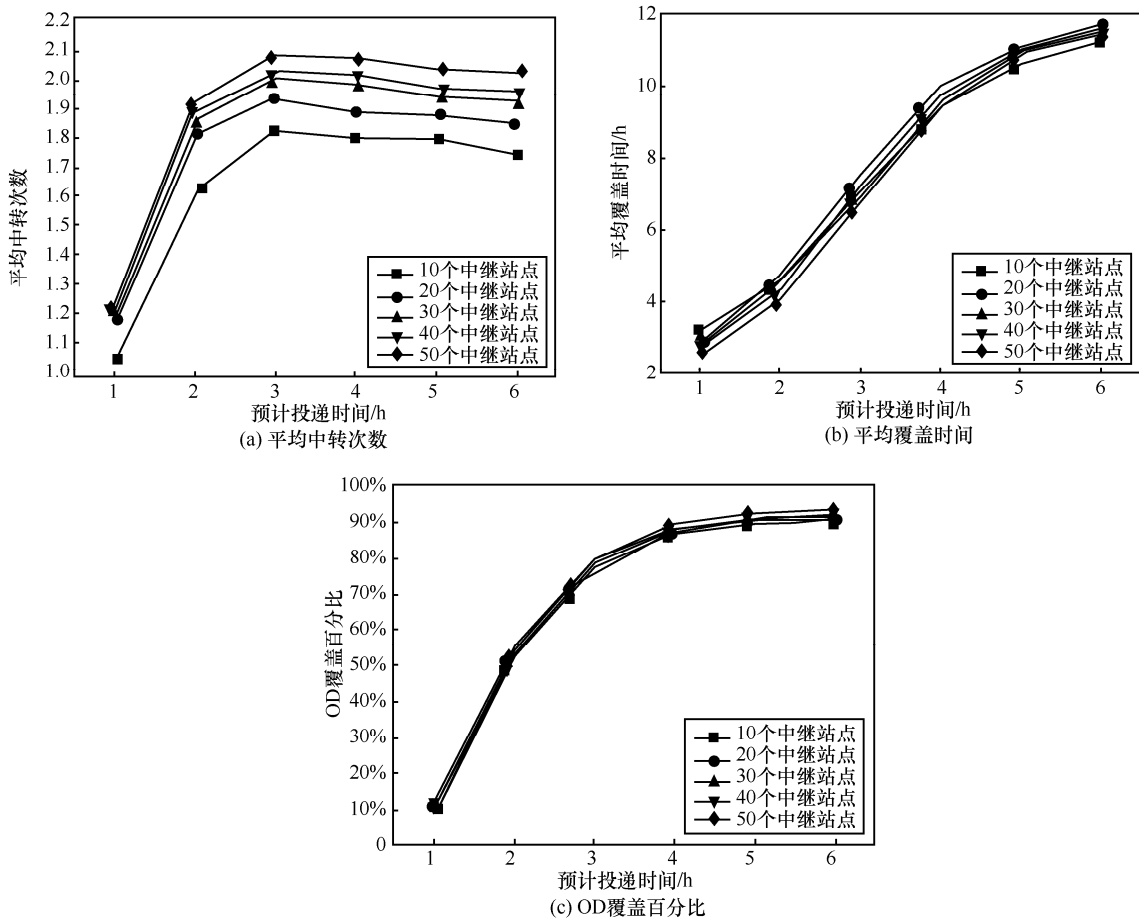


图 12 中转站点个数的影响

站位置信息增加一些中转站点以提高覆盖范围。本文综合考虑车站位置和客流量额外选取了 10 个中转站点，最终选取的 20 个中转站点，覆盖了上海主城区的大部分区域。由于上海南部目前还未建设地铁车站，此区域没有可供选择的中转车站，这些区域的包裹必须通过一定的路面运输过程才能使用地铁快递系统。

6.2 乘客参与率

为了评估乘客参与率对地铁快递系统的影响，对原始数据分别以 0.05、0.1、0.2 和 0.5 的概率进行采样，模拟不同的乘客参与率，地铁人流和投递时间的规律统计以及路由计算只能基于这些被采集的数据。在给定中转站点个数为 20，预计的投递成功率为 95% 的情况下的实验结果如图 13 所示。从图 13 中可以看出，越多的乘客参与投递在提升平均覆盖时间和 OD 覆盖率方面效果较为明显。由于增加了 10 个人流量较少的站点，路由系统的效益相对图 12 有所降低。在平均中转次数方面，0.1 和 0.2 的乘客参与率的实验结果较为相似。而如果

只有 5% 的人参与投递，在较短的预计时间内能够送达的 OD 站点数量较少，同时平均覆盖时间也相对较低，虽然此时提高预计投递时间能够提高一定量的覆盖百分比，但也会造成平均中转次数的急速上升。综合考虑，选取乘客参与率 (ν) 为 0.1，进行下面的实验。

6.3 预计投递成功率

预计投递成功率是指快递有多大的可能性能够在预计的时间内送达。快递公司在投递失败的情况下不仅需要损失一定的收益，长此以往还会降低快递公司的信誉，所以快递公司给定的预计成功率不能过低。本文分别测试了 80%、85%、90%、95%、99% 的成功率的情况下路由算法的效果，中转站的个数和乘客参与率分别固定为 20 和 0.1，最终的实验结果如图 14 所示。从图中可以看出，预计的投递成功率对平均覆盖时间和 OD 覆盖率有较为明显的影响。80% 到 95% 的 3 个实验指标较为类似；而如果预计成功率高达 99% 时，路由系统的性能降低。综合考虑，快递公司可以优先给出 95% 的预计成功率下的投递方案。

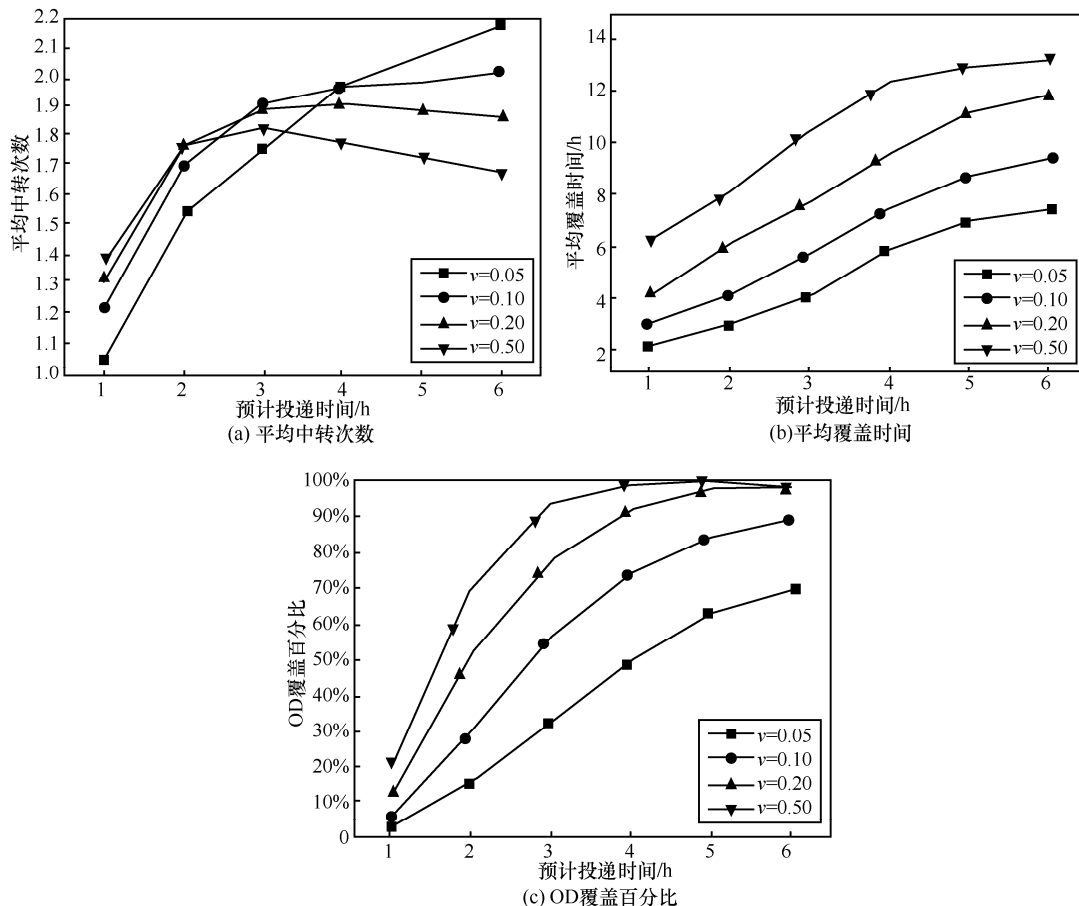


图 13 乘客参与率的影响

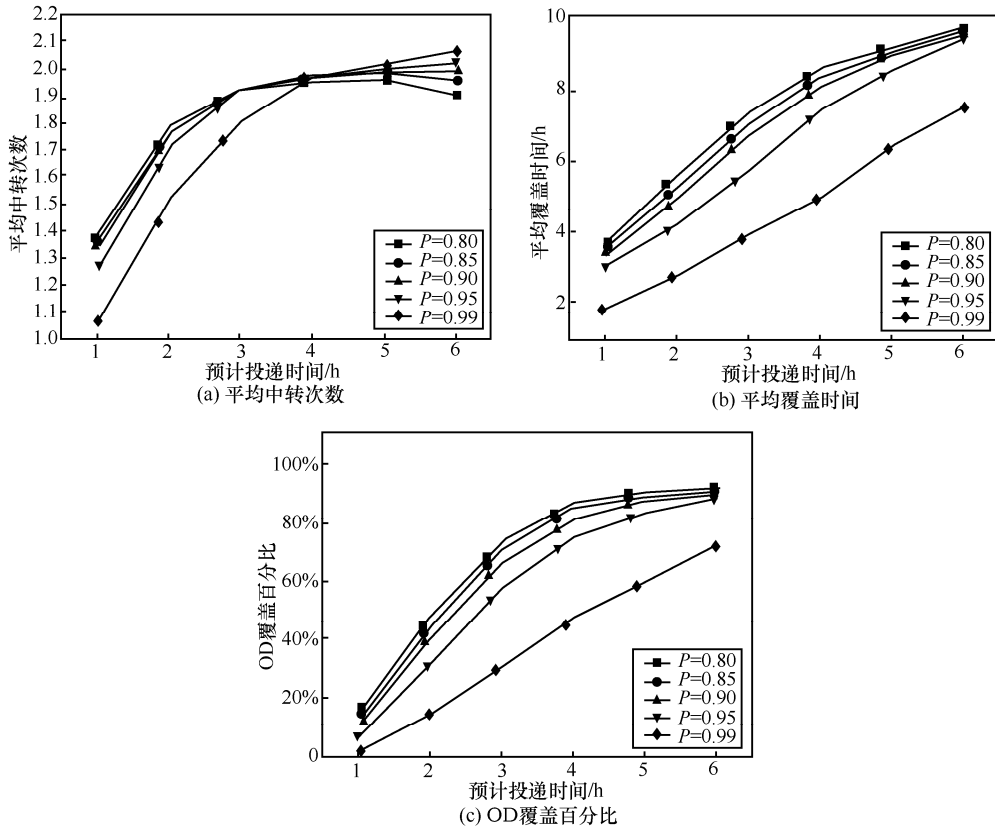


图 14 预计投递成功率的影响

6.4 日投递包裹数

为了验证地铁快递系统的巨大投递量，根据统计数据生成上海居民的快递请求，并读取乘客出行数据进行仿真投递。根据国家统计局统计上海 2013 年的年同城快递量达到了 3.3 亿件，粗略的估计出日均快递量在 90 万件左右，本文分别模拟了每天快递需求量分别为 5 万、10 万、20 万、50 万下的快递投递情况。假设这些包裹均由国内快递业务量最大的圆通公司进行投递，同时为了简化问题，快递的发送者被设定为快递公司官网可查询的快递代理点。包裹的起始站点按照以下原则分配：首先，将包裹平均分配到各个代理点，然后，将包裹分配给代理点附近的中转车站。包裹的目的站点按照以下原则分配：首先，将包裹按照各行政区域的人口比例分配到各个行政区，然后，在行政区域内将包裹按各个车站的人流量比例分配到特定的目的站点。每个包裹请求的生成时间是早上 7:00~19:00 的随机值，同时假设每个乘客每次只能投递一个包裹，中转车站有足够的空间放置包裹，乘客的参与率和预计投递成功率分别设定为 0.1 和 0.95，包裹在每个小时的投递数量随时间变化的实验结果如

图 15 所示。快递的投递量变化呈现出和地铁人流量类似的变化规律。如果假定乘客一次能携带多个包裹，最终得到在不同最大携带包裹数情况下的日投递包裹数如图 16 所示。横坐标表示乘客最多能同时携带 1~5 个包裹，纵坐标表示不同快递量需求下的日投递包裹数。从图 16 中可以看出，如果快递公司考虑适当的包裹打包策略，可以较为明显地提高地铁快递系统的传输量。

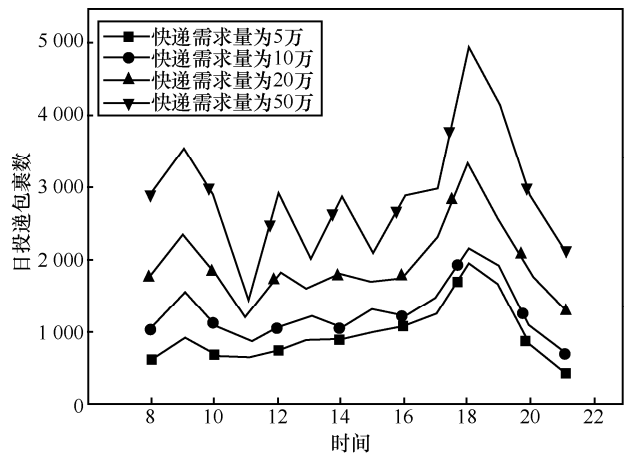


图 15 包裹投递数量的变化

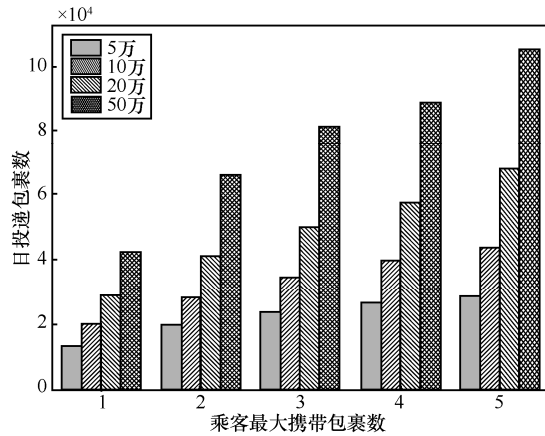


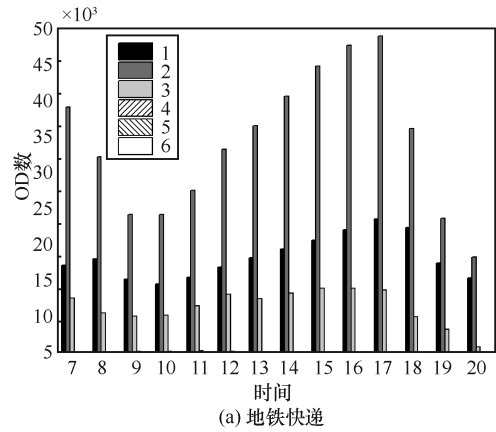
图 16 总投递包裹数

6.5 现有快递方案的对比

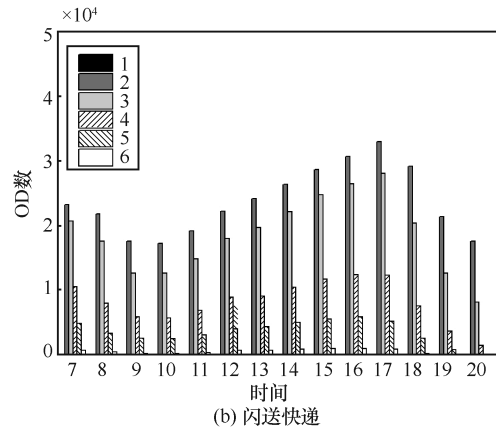
国内时效性最强的顺丰快递公司在上海市内的当日达快递服务的预计送达时间是 20 点，同时在大部分地区的 14 点之后不再提供该项服务，预计投递时间在 6 h 左右，部分行政区之间甚至不提供即日送达的服务，难以提供时延要求较高的快递服务。目前可以查询到较为详细的预计送达时间的众包模式快递公司只有闪送快递，闪送承诺的送达时效是 5 km 内 1 h 送达，每增加 5 km 送达时间增加 0.5 h。图 17 给出了地铁快递在一天不同时间段中预计送达时间的分布，同时，统计了这些可达地铁站之间的距离，并给出了闪送快递相应的预计送达时间的分布。其中，横坐标表示时间线，纵坐标表示每个时段对应不同预计投递时间的 OD 数量，此时，乘客参与率和预计投递成功率分别为 0.1 和 0.95。从图 17 中可以看出，可达的 OD 数在高峰期时段有较为明显的上升；两者在 2 h 内送达的 OD 占比均为最高；地铁快递在较短时间下的投递服务明显好于闪送快递。

7 结束语

通过分析上海地铁旅客出行数据，本文提出了一个通过地铁乘客以众包的方式帮助投递快递的系统，分析了系统的可行性，量化了地铁快递系统的服务质量。同时还提出了一种中转投递方案，能够有效地利用在地铁快递系统中，最后，针对地铁快递系统不同的参数进行了实验分析。在以后的研究中，本文作者计划分析并解决以下问题：找到一种合适的包裹打包策略增加传输量，同时，寻找有效的激励方式鼓励更多的人参与投递；协同不同的快递公司共同建设基础设施，并根据业务量设置合适的储物柜容量；包裹在传递过程中的安全性和隐私性问题。



(a) 地铁快递



(b) 闪送快递

图 17 不同预计投递时间下的包裹数量变化

参考文献:

- [1] MA S, ZHENG Y, WOLFSON O. Real-time city-scale taxi ridesharing[J]. IEEE Transactions on Knowledge & Data Engineering, 2015, 27(7):1782-1795.
- [2] ZHANG D, LI Y, ZHANG F, et al. CoRide: carpool service with a win-win fare model for large-scale taxicab networks[C]//ACM Conference on Embedded Networked Sensor Systems. 2013:1-14.
- [3] MA S, ZHENG Y, WOLFSON O. T-share: a large-scale dynamic taxi ridesharing service[C]//IEEE International Conference on Data Engineering. 2013:410-421.
- [4] CHEN L, ZHANG D, PAN G, et al. Bike sharing station placement leveraging heterogeneous urban open data[C]//ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015: 571-575.
- [5] YANG Z, HU J, SHU Y, et al. Mobility modeling and prediction in bike-sharing systems[C]//The International Conference. 2016.
- [6] LI Y, ZHENG Y, ZHANG H, et al. Traffic prediction in a bike-sharing system[C]//Sigspatial International Conference on Advances in Geographic Information Systems. 2015.
- [7] ZHANG D, JIANG R, WANG S, et al. EveryoneCounts: data-driven digital advertising with uncertain demand model in metro networks[C]// IEEE International Conference on Big Data. 2015:898-907.

- [8] ZHANG D, HUANG J, Li Y, et al. Exploring human mobility with multi-source data at extremely large metropolitan scales[C]//International Conference on Mobile Computing and Networking. 2014: 201-212.
- [9] ZHANG S, QIN L, ZHENG Y, et al. Effective and efficient: large-scale dynamic city express[C]//Sigspatial International Conference on Advances in Geographic Information Systems. 2015.
- [10] JEONG J, GUO S, GU Y, et al. TBD: trajectory-based data forwarding for light-traffic vehicular networks[C]//International Conference on Mobile Computing & Networking. 2009:231-238.
- [11] ZHANG L, YU B, PAN J. GeoMobCon: a mobility-contact-aware geocast scheme for urban VANETs[J]. IEEE Transactions on Vehicular Technology, 2015:1.
- [12] BURLEIGH S, HOOKE A, TORGERSON L, et al. Delay-tolerant networking: an approach to interplanetary Internet[J]. IEEE Communications Magazine, 2003, 41(6):128-136.
- [13] CASTEIGTS A, FLOCCHINI P, QUATTROCIOCCI W, et al. Time-varying graphs and dynamic networks[J]. International Journal of Parallel Emergent & Distributed Systems, 2010, 27(5):346-359.
- [14] MULLANGI R P. Efficient reachability analysis for evolving graph structured data[D]. Georgia, USA: University of Georgia, 2015.
- [15] B. BUI XUAN, A. FERREIRA, A. JARRY. Computing shortest, fastest, and foremost journeys in dynamic networks[J]. International Journal of Foundations of Computer Science, 2011, 14(2).

作者简介:



张坤芳(1994-),女,河南三门峡人,中南大学硕士生,主要研究方向为物联网、大数据、深度学习。



鲁鸣鸣(1978-),男,湖南长沙人,中南大学副教授,主要研究方向为大数据分析、深度学习、群智感知。



郑林(1992-),男,江西赣州人,中南大学硕士生,主要研究方向为物联网、大数据。