Using the Crowd of Taxis to Last Mile Delivery in E-commerce: a Methodological Research

Chao Chen[1] and Shenle Pan[2]

[1]Chongqing University
College of Computer Science
144 Shazheng Street, 400044 Chongqing, China
cschaochen@cqu.edu.cn
[2]MINES ParisTech, PSL Research University
CGS - Centre de Gestion Scientifique
60 Bd St Michel 75006 Paris, France
shenle.pan@mines-paristech.fr

Abstract. Crowdsourcing is garnering increased attention in freight transport area, mainly applied in internet-based services to city logistics. However, scientific research, especially methodology for application is still rare in the literature. This paper aims to fill this gap and to propose a methodological approach of applying crowdsourcing solution to Last Mile Delivery in E-commerce environment. The proposed solution is based on taxi fleet in city and a transport network composed by road network and customer self-pickup facilities that are 24 hours shops in city, named as *TaxiCrowdShipping* system. The system relies on a two-phase decision model, first *offline taxi trajectory mining* and second *online package routing and taxi scheduling*. Being the first stage of our study, this paper introduces the framework of the system and the decision model development. Some expected results and research perspectives are also discussed.

Keywords: Last Mile Delivery, Crowdsourcing, Taxi Trajectory Data Mining, Freight Transport, City Logistics

1 Introduction

In E-commerce environment, Last Mile Delivery (hereafter LMD) is the problem of transport planning for delivering goods from e-retailers' hub to the final destination in the area, for example the end consumers' home, see [1] and [2]. Speed and cost are the two crucial success factors to LMD. Faster shipping while with lower cost is the major challenge; nevertheless, it is also a paradox to a certain extend. Indeed, when customers are given a choice between fast and cheap delivery, most of them choose the cheap one, observed by a recent report [3]. The report also infers that that low-cost, speedy two-day delivery corresponds to most customers' expectation, opposite to the one-day delivery policy pursued by giant e-retailers such as Amazon and Alibaba etc. This fact may open up new opportunities to innovative freight transport models [4] for LMD aiming at reducing delivery cost while respecting shipping time,

adfa, p. 1, 2011.

nevertheless, not necessarily aiming at minimizing shipping time. Being our topic here, crowdsourcing is one of such solutions getting more and more attention [5].

In the literature, crowdsourcing has been usually seen as "an interesting idea" for freight transport before seriously moving to real applications. Despite the existence of some internet-based services, scientific research, especially methodology for application is rare in the literature [5]. This paper aims to fill this gap by providing a methodological approach of applying the crowdsourcing solution proposed and to assess its performance. In this paper, the crowd studied is taxi fleet in city, supported by a transport network composed by road network and customer self-pickup facilities such as 24 hours shops in city, named as *TaxiCrowdShipping* system. The system relies on a two-phase decision model, first offline taxi trajectory mining and second online package routing – taxi scheduling. As the first stage of our research, this paper introduces the objective and the framework of the *TaxiCrowdShipping* system, as well as to define the function of system.

The reminder of the paper is organised as follows. Section 2 consists of a relevant literature review. Then, Section 3 presents the *TaxiCrowdShipping* system. After introducing some basic concepts and assumptions, we focus on the two steps approach being the decision support tool for the system. Some expected results are also discussed. Finally, Section 4 concludes this paper by giving some research perspectives for future works.

2 Related Works

Recently some innovative solutions have been studied for city logistics and LMD in E-commerce environment, for example those involved in our study like interconnected city logistics enabled by Physical Internet [6, 7], self-service parcel station (e.g., DHL PackStation, LaPoste Pickup Station etc.), new tools for LMD (bicycle, motor, electric vehicle etc.), Smart city logistics [8], and crowdsourced delivery [5]. Due to the space limitation, here we focus on the works regarding crowdsourcing in Freight Transport.

Being firstly discussed in [9], crowdsourcing has been increasingly studied as a solution to freight transport. It can be simply defined as "*outsourcing a task to the crowd via open call*" [9]. One the practice side, it occurs mainly in the form of internet-based services for example *imoveit.co.uk* and *zipments.com*, where the crowd is undefined. Thus, both professional (e.g., carriers) and non-professional (e.g., inhabitants) service providers may answer the calls. In 2014 Amazon has launched a project to explore taxi deliveries in San Francisco and Los Angeles[1]. The idea is similar to our study, though, their methodology and results are not yet published to our knowledge. Moreover, the package deliveries are completed by ordering free taxis, while our proposed solution leverages the *hitchhiking rides* provided by occupied

---

[1] http://www.engadget.com/2014/11/05/amazon-is-exploring-taxi-deliveries-in-san-francisco-and-los-ang/

taxis when they are sending passengers, thus our solution is more green and economic. On the scientific side, only few relevant works in the area of logistics can be found. A case study of applying crowdsourcing to library deliveries in Finland is conducted in [5]. They study a system called *PiggyBaggy* to assess the sustainability and adaptability of such solution. A taxi-based solution for the waste-collecting or product return problem (i.e., reverse logistics) in metropolitan area is discussed in [10], without considering goods delivery. Some other relevant works can be also found in the area of data science. Data scientists are mainly interested at mining the taxi trajectory data to understand the city dynamics, and developing various smart services for taxi drivers, passengers, as well as the city planners [11, 12]. However, almost all the current research related to taxi data mining focuses on the people or public transport [13, 14], little attention has been paid to freight transport.

From the literature we can see that crowdsourcing in freight transport usually occurs in the form of internet-based services in practice, and it is usually investigated via case study in the literature. Methodology for application is not well addressed. Besides, no attention has been paid to crowd selection or definition. People in city are often regarded as eligible crowd. Following the previous work [10] dealing with reverse flows, this paper focus on a methodology approach for the LMD problem, where logistics constraints and decision model are different.

3 *TaxiCrowdShipping* System

3.1 Basic Concepts and Assumptions

To ease the description, we define the related concepts based on Fig. 1, and also make some assumptions.

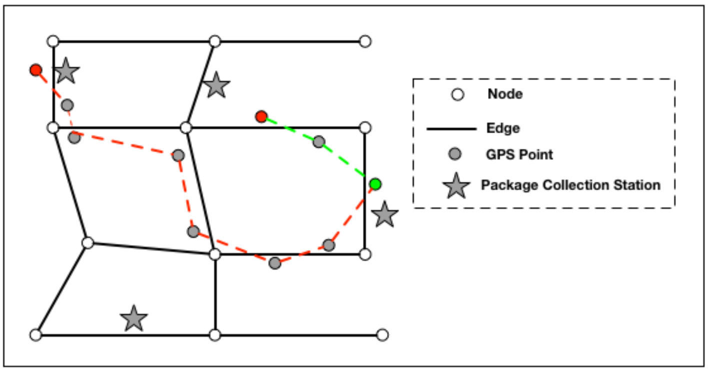

Fig. 1. Illustration of some basic concepts.

Definition 1. (Road Network) A road network is a graph *G(N, E)*, consisting of a node set *N* and an edge set *E* (as shown in Fig.1), where each element *n* in *N* is an intersection and is associated with a pair of longitude and latitude degrees *(x, y)* representing its spatial location. Edge set *E* is a subset of the cross product *NxN*. Each element *e(u, v)* in *E* is a street connects node *u* to node *v*, which can be one-way or bidirectional, depending on real cases.

Definition 2. *(Taxi Trajectory)* A taxi trajectory is a sequence of time-stamped GPS points. Each GPS point $p_i=(t_i, x_i, y_i, ind_i)$ consists of a time-stamp t_i, a longitude x_i, a latitude y_i, and an indicator ind_i showing whether the taxi is occupied or not. A pick-up point is a special GPS point, with the indicator changing from 0 to 1 (the red-coloured circle in Fig.1); while a drop-off point is the one with the indicator changing from 1 to 0 (the green-coloured circle in Fig.1). Thus we can further define a *passenger-delivery trajectory* is the GPS sequence from the pick-up point to the followed drop-off one (the red dashed line in Fig.1); a *passenger-hunting trajectory* is the GPS sequence from the drop-off point to the followed pick-up point (the green dashed line in Fig.1).

Definition 3. *(Package Pickup Station)* A package pickup station is a Point of Interest (POI hereafter) near roadside that is responsible for storing packages waiting for consumer pickup, (the star in Fig.1). Here we select *24-hour opening convenience stores* near roadside as the package pickup stations.

Definition 4. *(Package Delivery Request)* A package delivery request is defined as a triple $<o_p, d_p, t_p>$, where o_p and d_p refer to the origin and the destination of the package respectively, and t_p refers to the time when the user submits the request. The request is generated by users who need the package express delivery service.

Definition 5. *(Real-time Taxi Ordering Request)* A real-time taxi ordering request is defined as a triple $<o_t, d_t, t_t>$, where o_t and d_t refer to the passenger's origin and the destination respectively; t_t refers to the time when the passenger submits the request. The request is made by passengers who need taxi service.

Assumption 1. All selected POI is open 24/7, without capacity issue and with good accessibility to taxi.

Assumption 2. The taxi drivers are willing to accept the assigned package delivery tasks.

Assumption 3. The package can be *trackable*. Since the birth time, the package is either stored at the pickup station or carried by the scheduled taxi. Each pickup station is authorized and has a unique ID; each taxi is registered in taxi management department and also has a unique ID.

3.2 Problem Description

To help understand how our proposed solution works to handle the LMD, we intentionally design a simple running example. Suppose in Fig.1 the leftmost and rightmost star are the origin and the destination of the package respectively. After the generation of package delivery request, there happens to be a passenger who makes a real-time taxi ordering request, intending to go to the same destination. At that time, we can assign the package delivery task to the taxi which has responded the passenger's request. Finally, the package will be also delivered, with a hitchhiking ride provided by the taxi while sending the passenger. The solution can be featured as economic and eco-friendly one since it almost does not incur extra labour cost and energy as well as CO_2 emissions.

Accordingly, the taxi-based crowdsourcing solution to LMD consists of the shortest path finding problem for packages and the scheduling problem for taxis; and it can be described as follows.

Given:

- A road network and a set of package pickup stations in the studied city;
- A set of taxi trajectory data in the studied city in history (e.g., last month);
- A set of package delivery requests, and a set of real-time generated taxi ordering requests. Note that these requests come in stream.

Objective:

For a given package delivery request, find its optimal package delivery path which can minimize the total package delivery time (i.e. maximize the delivery speed). Once determined the path, we can schedule the next coming taxi to delivery the package having the same destination. Note that in one of the scenarios in this study the path can be re-planned according to the Real-time Taxi Ordering Request.

Constraints:

Only taxis which response the taxi ordering requests after the package delivery request can be scheduled.

Once a taxi is involved into a delivery task, it can be available again to be scheduled to participate only after completing the current task (i.e. sending the package to the predefined pickup station). In another word, a taxi can carry at most one package when sending passengers.

3.3 The *TaxiCrowdShipping* System

The *TaxiCrowdShipping* system contains two components, i.e., the *Offline-Trajectory Mining and the Online-Package Routing* respectively, which will be detailed as follows and shown in Fig 2.

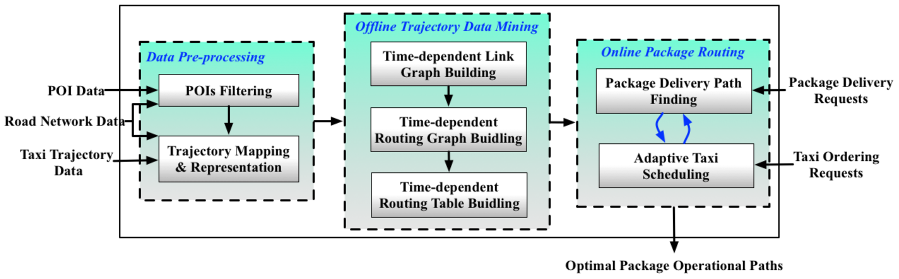

Fig. 2. The *TaxiCrowdShipping* system architecture.

Offline-Trajectory Mining

The objective of the Offline-Trajectory Mining is to estimate the direct package delivery time from one pickup station to another one, by taking a single hitchhiking ride. The time cost mainly includes two parts: the time cost on waiting for the hitchhiking rides, which is related to the frequency of taxi rides, and the time spent on driving on the roads. Here, we propose a two-step procedure to estimate the time cost.

Step 1: From Trajectory Data to Passenger Flow. From the given taxi trajectory data, it is not difficult to compute the passenger flow between any two pickup stations. Specifically, to compute the passenger flow from cs_i to cs_j, the trajectories meeting Eqns. 1~2 will be counted. Then, the passenger flow from cs_i to cs_j is just the number of trajectories satisfying the requirements. Note that there may be no passenger flow between some pickup station pairs, and the passenger flow is different at different time slots. For the purpose of future research, we divide the time into three time slots for a day in advance, i.e., night-time hours, day-time hours and rush hours.

$$Ddist(Tr_i.o, loc(cs_i)) < \delta \tag{1}$$

$$Ddist(Tr_i.d, loc(cs_j)) < \delta \tag{2}$$

where $Tr_i.o$ and $Tr_i.d$ are the original and destination points of Tr_i, respectively; $loc(\cdot)$ gets the latitude and longitude location of the given pickup station; δ is a user-specified parameter. $Ddist(a \cdot b)$ calculates the driving distance from point a to b.

Step 2: From Passenger Flow to Time Cost. To estimate the time cost, we need to estimate two parts, i.e. the waiting time and the driving time. The waiting time is defined as the time cost on waiting for the suitable hitchhiking ride event of passenger taking taxis, to help deliver a package from cs_i to cs_j directly (with no transhipment). Here, we employ the *Non-Homogeneous Poisson Process (NHPP)* to model the behaviour of passenger taking taxis [15]. According to the passenger flow, we can estimate the waiting time of packages at different time slots at the pickup stations. Under the Poisson hypothesis within a time slot, we could derive the probability distribution of the waiting time for the next suitable hitchhiking ride event (i.e. t_{next}, the event of a passenger taking taxi from cs_i to cs_j), which can be expressed in Eq. 3:

$$\begin{aligned} P\{t_{next} \leq t\} &= 1 - P\{t_{next} > t\} \\ &= 1 - P\{N(t) = 0\} \\ &= 1 - e^{-\lambda \cdot t} \end{aligned} \tag{3}$$

Here $N(t)$ represents the number of event occurring within t, and $P\{N(t) = k\} = e^{-\lambda \cdot t \cdot \frac{(\lambda \cdot t)^k}{k!}}$. Then the probability density function (pdf) of t_{next} is just the derived function of $P\{\}$, as can be seen in Eq. 4.

$$p(t) = \lambda \cdot e^{-\lambda \cdot t} \tag{4}$$

Thus, we can deduce the expectation of t_{next} (i.e. the waiting time for the hitchhiking ride event occurring):

$$E[t_{next}] = \int_0^{\infty} t \cdot \lambda \cdot e^{-\lambda \cdot t} \cdot dt = \frac{1}{\lambda} \tag{5}$$

Note that λ in the model is the frequency of passenger taking taxis from cs_i to cs_j (i.e. the passenger flow from cs_i to cs_j), which can be easily estimated by the Eq. 6.

$$\hat{\lambda} = \frac{\bar{N}}{\Delta T} \qquad (6)$$

where $\bar{N}$ is the average number of passengers taking taxis from cs_i to cs_j during the studied time slot in the observed days; ΔT is the time duration of the that time shot.

Therefore, the waiting time from cs_i to cs_j is:

$$waiting\ time = \frac{1}{\hat{\lambda}} = \frac{\Delta T}{\bar{N}} \qquad (7)$$

For each passenger-delivery ride from cs_i to cs_j, it is easy to derive its time spent on driving on the roads. The driving time is simply the average one of all such rides, as shown in Eq. 8.

$$driving\ time = \frac{\sum_{i=1}^{N} Tr_i.(te - ts)}{N} \qquad (8)$$

where N is the number of passenger-delivery rides during the studied time slot in the observed days. $t_e - t_s$ is the time cost of the corresponding taxi ride.

Finally, the time cost is just the sum of waiting time and driving time, as shown in Eq. 9.

$$tc = waiting\ time + driving\ time = \frac{\Delta T}{\bar{N}} + \frac{\sum_{i=1}^{N} Tr_i.(te - ts)}{N} \qquad (9)$$

Note that the time cost will be $+\infty$ if there was no passenger flow on the respected pickup station pair.

Online- Package Routing and Taxi Scheduling

The objective of the Online- Package Routing and Taxi Scheduling is to schedule the specific taxis to help delivery the packages with the determined optimal path, according to the real-time coming taxi ordering requests. Here, we also propose a two-step procedure to complete, detailed as follows.

Step 1: Find the Optimal Pickup Station Sequence. For a package delivery request, with the estimated time cost values in the last component, it is trivial to find the best pickup station sequence from the origin to the destination of the package, in terms of the total time cost, by applying the classical shortest path finding algorithms.

Step 2: Schedule the Taxis. After obtaining the optimal pickup station sequence for a package delivery request, we schedule the taxis according to the real-time taxi ordering requests. In more detail, from the origin of the package to the followed pickup station in the optimal pickup station sequence, we wait for the taxi which will pick up a passenger at the origin, heading to the followed station (the information is included in the real-time taxi ordering requests), and assign the package delivery task to that taxi. After that, the origin of the package will be also updated. The procedure will be repeated until the package arrives at its destination.

3.4 Expected Results

Following the framework proposed here some results are expected in the next steps. First, we will conduct study to assess the implementability of the *TaxiCrowdShipping* system proposed. A large city in China, namely Hangzhou city is selected to be the test field, thanks to some available data sets there such as Open data of taxi trajectory, may of city shops' and road network etc. However, the data of package delivery request is still to be completed. Second, a set of algorithms for package routing and taxi scheduling problem will be developed and examined. Then a set of scenarios is expected to the study in order to assess the performance of the system as well as its sensibility of setting.

4 Conclusion

In this paper we aim to propose a methodological approach of applying crowdsourcing solution to Last Mile Delivery (LMD) problem in city logistics. To this end, a system called *TaxiCrowdShipping* is proposed, whose objective is to use the taxi fleet and shops in city for LMD purpose. The framework of such system is discussed in this paper, as well as the two-phase decision model. Except the expected results discussed above, this study opens some research perspectives. First, the Physical Internet-based active container described in [16] may be adapted to this study. Suppose that such containers are able to actively publish a delivery request on web (as a passenger calls a taxi), the taxi scheduling could be more efficient and responsive. It could also provide a good real time tracking and tracing technique to the system. Second, the study can be extended to the automated self-service parcel station implantation problem. Coupling such station placement and crowdsourcing solution is still rarely studied in the literature.

Reference

1. Lee, H.L. and S. Whang, *Winning the last mile of e-commerce.* MIT Sloan Management Review, 2001. **42**(4): p. 54-62.
2. Punakivi, M., H. Yrjölä, and J. HolmstroÈm, *Solving the last mile issue: reception box or delivery box?* International Journal of Physical Distribution & Logistics Management, 2001. **31**(6): p. 427-439.
3. Gibson, B.J., C.C. Defee, and R. Ishfaq, *The state of the retail supply chain: Essential Findings of the Fifth Annual Report*. 2015: RILA, Dallas, TX.
4. Meyer-Larsen, N., J.B. Hauge, and A.-S. Hennig, *LogisticsArena—A Platform Promoting Innovation in Logistics*, in *Logistics Management*. 2015, Springer. p. 505-516.
5. Paloheimo, H., M. Lettenmeier, and H. Waris, *Transport reduction by crowdsourced deliveries – a library case in Finland.* Journal of Cleaner Production, 2015.
6. Sarraj, R., et al., *Analogies between Internet network and logistics service networks: challenges involved in the interconnection.* Journal of Intelligent Manufacturing, 2014. **25**(6): p. 1207-1219.

7. Crainic, T.G. and B. Montreuil, *Physical Internet Enabled Interconnected City Logistics*, in *1st International Physical Internet Conference*. 2015: Québec City, Canada.
8. Neirotti, P., et al., *Current trends in Smart City initiatives: Some stylised facts.* Cities, 2014. **38**: p. 25-36.
9. Howe, J., *The rise of crowdsourcing.* Wired magazine, 2006. **14**(6): p. 1-4.
10. Pan, S., C. Chen, and R.Y. Zhong. *A crowdsourcing solution to collect e-commerce reverse flows in metropolitan areas*. in *15th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2015*. 2015. Canada, Ottawa: Elsevier.
11. Castro, P.S., et al., *From taxi GPS traces to social and community dynamics: A survey.* ACM Computing Surveys (CSUR), 2013. **46**(2): p. 17.
12. Zheng, Y., *Trajectory data mining: an overview.* ACM Transactions on Intelligent Systems and Technology (TIST), 2015. **6**(3): p. 29.
13. Chao, C., et al., *B-Planner: Planning Bidirectional Night Bus Routes Using Large-Scale Taxi GPS Traces.* IEEE Transactions on Intelligent Transportation Systems 2014. **15**(4): p. 1451-1465.
14. Liu, Y., et al. *Exploiting Heterogeneous Human Mobility Patterns for Intelligent Bus Routing*. in *Data Mining (ICDM), 2014 IEEE International Conference on*. 2014: IEEE.
15. Qi, G., et al. *How Long a Passenger Waits for a Vacant Taxi--Large-Scale Taxi Trace Mining for Smart Cities*. in *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*. 2013: IEEE.
16. Sallez, Y., B. Montreuil, and E. Ballot, *On the Activeness of Physical Internet containers*, in *Service Orientation in Holonic and Multi-Agent Manufacturing Control*, T. Borangiu, A. Thomas, and D. Trentesaux, Editors. 2014, Springer Berlin Heidelberg. p. 235-244.