A Data-driven Approach to Design Feeder Bus Network based on **Aggregated Cellphone Data and Open GIS Tool**

Yaojun Wang, Yue Liu (corresponding author) **Department of Civil and Environmental Engineering University of Wisconsin - Milwaukee**



Abstract

A dedicated feeder bus service of being satisfied with commuters will greatly prompt them to shift travel mode to public transit. This research focuses on designing feeder bus network through big data mining and the application of Open GIS.

Research Framework

A feeder bus system is proposed to transport passengers to urban rail transit station. Such a system benefits from using cellphone data to extract real distribution of passenger demand which is expected to efficiently solve the issue of demand uncertainty in traditional feeder bus design problem. Moreover, an open GIS tool is further developed to offer traffic status and topological features within study area. A mixed-integer programming is formulated to select the most appropriate locations of bus stops, and guide passengers from demand points to their associated stops. Those key components can be better illustrated in Figure 1.

The Open GIS tools help to obtain the travel information (walking distance, driving time, traffic status) for each OD pairs.





Data Processing

The cellphone database will record the cellphone subscribers' activities when they carry a cell phone and make a call, text, or have network connections. Figure 2 shows the raw data from the database. The moving path of the cell phone subscriber can be retrieved (figure 3) and the matrix of original-Destination (OD) pair is able to be generated (figure 4).

Figure 5. The application of Open GIS tools.

Case Study

The residential area around Jiandingpo Station at Metro Line 1 in Chongqing, China is selected as the case study. More than 3.5×10^8 records were generated in 2 days. By applying the data processing methodology explained, 25 demand points containing 513 passengers who catch Line 1 at Jiandingpo Station in morning peak hour are located at the map shown in figure 6.



		MSID	TIMESTAMP	LAC	CELLID	EVENTID	CAUSE	FLAG	MSCID	BSCID	CAUSETYPE
1	1	599d66977a78b2320a418c70892b3baf	2014-03-03 00:01:31.522	33589	44510	1	9	000	188422	189957	1
	2	6f8f64d8e8e6e6b6e1d6eec81e7897d0	2014-03-03 00:01:33.647	33575	50260	8	9	000	188422	189957	1
	3	1b28c4819349441f6db8e3a8cf85e16f	2014-03-03 00:01:33.217	33589	51591	8	9	000	188422	189957	1
	4	8c6cdbfcc1c6119ed4f3f31b569b3464	2014-03-03 00:01:33.247	33589	46171	8	9	000	188422	189957	1
1	5	0406dbb65272ff654c60af136e871d86	2014-03-03 00:01:33.470	33576	48812	8	9	000	188422	187653	1
	6	307ff986d223466c75c0786f817a2e8b	2014-03-03 00:01:33.712	33576	44995	8	9	000	188422	187653	1
	7	e86f3abb07ca36a6aeead8c66f7cf88d	2014-03-03 00:01:33.747	33575	52600	8	9	000	188422	189957	1
4	8	2ce203cbc1a23cb523989bc14487fd72	2014-03-03 00:01:33.784	33569	42465	8	9	000	188422	187653	1
-	9	19cb1618f1ce1183d98a77d867b7e279	2014-03-03 00:01:32.942	33576	48130	7	9	000	188422	187653	1
	10	f03f6f220dedb124ea6992acba8362da	2014-03-03 00:01:32.630	33575	46882	8	9	000	188422	189957	1
	11	155d305b82ef2b42b8d3ecdd551749d0	2014-03-03 00:01:33.371	33576	40060	8	9	000	188422	187653	1
	12	68710f1c8b62ef7b63c6e94ae41c41ed	2014-03-03 00:01:33.336	33589	50300	8	9	000	188422	189957	1
	13	cd0d6a20c08246021e4c766532bd5390	2014-03-03 00:01:33.357	33589	50271	8	9	000	188422	189957	1
- C	• •	1 +1 103004 470 00 ACC 0 ACC		2017	10000	•	0	000	100400	100057	. 1

Figure 2. The signaling data from the base station.





Demand point Candidate bus stop

Figure 6. Spatial distribution of demand using Jiandingpo Station and candidate bus stops

Result



Using the demand collected from cellphone data exploration, distance and time matrix generated from Open GIS tool, the proposed model was able to be solved in CPLEX optimality in 390 seconds.

Conclusion

Different from existing studies, the proposed methodology features in:

- introducing a big data processing technology for extracting 1) aggregated-level spatial distribution of demand with using cellphone dataset
- retrieving map-based travel distance and time information to 2) include the network characteristics and traffic status by using Open GIS tool









