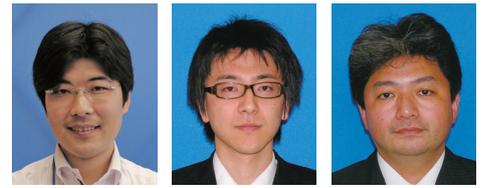


プローブ旅行速度のデータクレンジング手法の開発



道路交通研究部 道路研究室

交流研究員 **松島 敏和** (主任研究官 博士(工学)) 室長 **橋本 浩良** (博士(学術)) **高宮 進**

(キーワード) プローブデータ、旅行速度、データクレンジング、ETC2.0 プローブ情報

1. はじめに

近年、カーナビなどによるプローブデータが大量に収集され、渋滞状況の分析などに活用されている。プローブデータをリンク別の旅行速度に加工したプローブ旅行速度には、個人の運転嗜好や路側での駐停車などによる特異な高速度や低速度のデータが存在するため、分析にあたっては特異なデータの除去(データクレンジング)が必要である。

2. データクレンジング手法の概要と特徴

開発中のデータクレンジング手法の特徴は、個別車両間の所要時間差に着目する点である。図1の例のような駐停車のデータは交通状況分析には適さないため、除去などの措置をとるべきと考えるが、個別車両の旅行速度のみでは混雑による低速度走行のデータと区別ができない。駐停車車両は他の車両との所要時間差が大きく、混雑による低速度車両は他の車両との所要時間差が小さいことに着目し、当該リンクの所要時間が同一時間帯における最小所要時間から一定程度遅れる車両を特異な走行とみなし除去することを考えた。

3. データクレンジングの試行結果のその考察

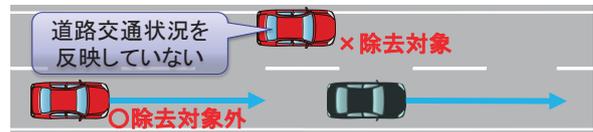
ETC2.0プローブ情報(2015年4月~6月全国分)のDRMリンク別上下方向別の旅行速度データに対して、最小所要時間からの遅れにあたる時間閾値を600秒に設定してデータクレンジングを試行した。対象件数は約274百万台リンク、うち除去対象件数は約0.7百万台リンク、除去対象件数の割合は約0.32%である。同一時間帯の個別車両に着目し、相互比較をした。その結果、路側での駐停車など特異な走行と想定されるデータを除去できていることが確認できた。

4. おわりに

大容量データを活用していく上では、分析目的に

照らしたデータクレンジングが重要である。今後、データクレンジング手法をさらに吟味し、マニュアル化などの汎用化策について検討する。

送迎・荷物の積み下ろしなどによる駐停車



混雑による低速度の走行

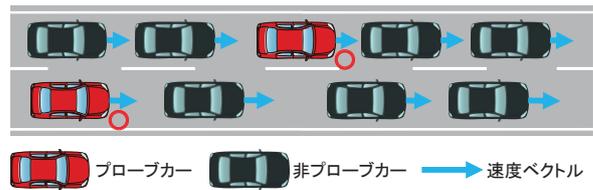


図1 除去対象とする走行の例

表1 データクレンジング手法の概要

項目	内容
①閾値によるデータの除去	低速度側1km/h未満、高速度側150km/h以上となるデータを除去
②同一時間帯のデータの除去	同一時間帯における最小所要時間のサンプルから600秒以上遅れるサンプルを除去。ただし、高速道路では80km/h、一般道では30km/hの走行時の所要時間を最小所要時間の下限とする

表2 データクレンジング前のデータ件数

速度ランク\DRMリンク延長	500m未満	500m以上1000m未満	1000m以上2000m未満	2000m以上3000m未満	3000m以上	合計
1km/h未満	101,128	1,730	28	0	0	102,886
1km/h以上10km/h未満	11,992,068	381,337	61,583	7,295	1,913	12,444,196
10km/h以上20km/h未満	21,784,312	1,393,434	182,147	21,286	8,499	23,389,678
20km/h以上40km/h未満	39,880,142	5,829,960	1,169,011	105,362	47,875	47,032,350
40km/h以上60km/h未満	63,059,274	8,611,434	2,945,802	497,285	206,778	75,320,573
60km/h以上80km/h未満	36,702,063	7,951,405	4,152,835	1,162,003	861,972	50,830,278
80km/h以上100km/h未満	23,351,446	7,013,708	5,187,359	2,015,993	1,852,373	39,420,879
100km/h以上120km/h未満	10,918,100	3,441,007	2,727,682	1,246,404	1,332,974	19,666,167
120km/h以上150km/h未満	3,262,987	600,510	427,468	180,034	189,023	4,660,022
150km/h以上	601,744	44,377	17,343	5,410	4,609	673,483
合計	211,653,264	35,268,902	16,871,258	5,241,072	4,506,016	273,540,512

(台リンク)

表3 データクレンジングの除去対象件数の割合

速度ランク\DRMリンク延長	500m未満	500m以上1000m未満	1000m以上2000m未満	2000m以上3000m未満	3000m以上	合計
1km/h未満	100.00%	100.00%	100.00%	-	-	100.00%
1km/h以上10km/h未満	0.63%	10.10%	29.08%	71.17%	80.71%	0.75%
10km/h以上20km/h未満	0.02%	0.00%	0.01%	6.14%	38.98%	0.02%
20km/h以上40km/h未満	0.00%	0.00%	0.00%	0.00%	0.81%	0.00%
40km/h以上60km/h未満	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
60km/h以上80km/h未満	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
80km/h以上100km/h未満	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
100km/h以上120km/h未満	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
120km/h以上150km/h未満	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
150km/h以上	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
合計	0.35%	0.24%	0.21%	0.23%	0.22%	0.32%

凡例: 割合 大 小
 ※除去対象件数の割合(%) = 除去対象データ対象件数 / データクレンジング前のデータ件数