

ラムダアーキテクチャによるダイナミックマップにおける プローブデータクレンジングの検討

Probe data cleansing in dynamic map system by lambda architecture

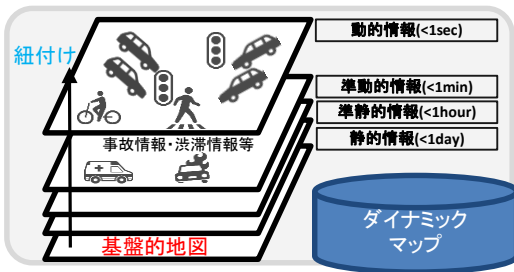
赤間 優樹・暗号・認証分科会・情報セキュリティ大学院大学

Abstract :

Dynamic map is a collection of data obtained by superimposing dynamic information on static map. Dynamic map contains information generated by using the probe data. Probe data is information collected through vehicle sensors and measuring devices. However, abnormal data may be input in dynamic map system due to sensor failure or data falsification. Cleaning of abnormal data is necessary to satisfy accuracy and real-time property which is a requirement to be satisfied by dynamic map. In this research, the accident location of semi-dynamic information is taken as an example of probe data cleansing. To ensure the accuracy of data in time constraints, I propose a combination of cleansing using two types of outlier detection based on the idea of lambda architecture. Also, I evaluate the proposed method.

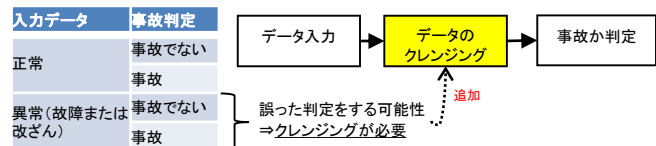
ダイナミックマップとは

- 自動走行システムや安全運転支援システムのための利用が想定されている、静的な地図に動的な情報を重畳したデータの集合体
- 地図の更新時にはリアルタイム性が求められ、データ基盤としては正確性が求められる



背景と目的

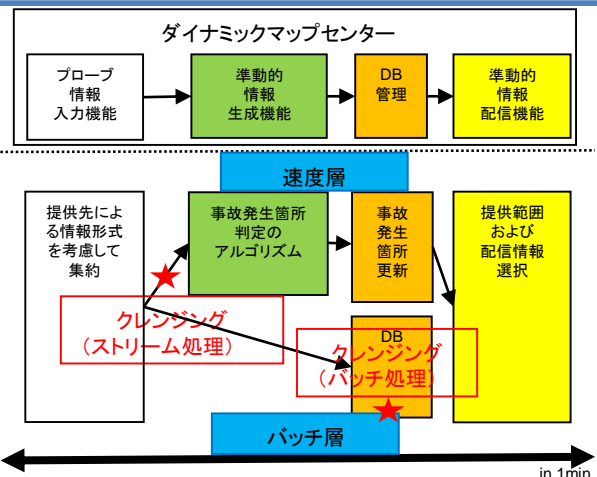
ダイナミックマップの情報には**プローブデータ**（車両のセンサ・計測装置を通して収集される情報）を利用して生成されるものがある（例、事故発生箇所）
しかし、センサの故障や改ざんにより異常なデータが入力される可能性があり、誤った情報を生成する恐れがある
当該データに対して、リアルタイム性と正確性を満たすように**クレンジング**を行う方法について、事故発生箇所を例として検討する



提案手法

- データ改ざんの検出には送信側の検証情報を受信側が入手することが抜本的な対策となるが、機器に取り付けるコストや設備更改までの期間を考慮すると直ちに対策を行うことは容易ではない
- 入力されたデータに対して**外れ値検知**の手法を利用してクレンジングを行う
- リアルタイム性と正確性を両立するために、速度層およびバッチ層による処理を組み合わせた**ラムダアーキテクチャ**の考えを用いる
- 速度層ではストリーム処理でのクレンジングとして、事故疑いのあるデータに対して山西ら[1]のオンライン外れ値検知エンジン **SmartSifter** を適用する
- バッチ層ではバッチ処理によるクレンジングとして前回クレンジング以降の入力データに対して、Hawkins[2]の**RNN(Replicator Neural Network)**を用いた**外れ値検知手法**を適用する。初期段階においては、アナリストによる結果のモニタリングも検討する

クレンジング箇所	タイミング	処理データ	クレンジングアルゴリズム
速度層	即時	事故疑いのあるデータのみ	SmartSifter オンライン外れ値検知
バッチ層	定期	新しく入力された全データ	RNNを用いた外れ値検知



評価

- 速度層において、適用するSmartSifterのアルゴリズムの計算量から、クレンジングにかかる処理時間が軽微であることを確認した
- バッチ層において、速度層では発見できない異常データを確認可能であることを定性的に確認し、理論的には許容される異常データの割合以内に抑えることが可能である

まとめ

本稿ではダイナミックマップにおけるプローブデータに対し、リアルタイム性および正確性を満たしながら、クレンジングが追加可能であることを検討した。故障時のデータの挙動が完全に把握できないこと、機器の検証情報を利用しない仮定を置いていることにより、全ての異常データを検出することは不可能であるが、実験等により提案手法の有効性を確認することが今後の課題である。

[1] K.Yamanishi, J.Takeuchi, G.Williams, P.Mine, "On-Line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms", Data Mining and Knowledge Discovery, Volume 8, Issue 3, pp 275-300 (2004/03)
[2] S.Hawkins, H.He, G.Williams, R.Baxter "Outlier Detection Using Replicator Neural Networks 4th International Conference on Data Warehousing and Knowledge Discovery, Springer-Verlag, pp170-180(2002).