

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

赤間 優樹[†] 後藤 厚宏[‡]

(株)アーク情報システム[†] 情報セキュリティ大学院大学[‡]

概要

ダイナミックマップ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である。地図としてのダイナミックマップは高精度かつ高鮮度（地図が最新の状態に更新されている）であることが求められ、データ基盤としてのダイナミックマップはデータの正確さが求められる。ダイナミックマップにはプローブデータを用いて生成される情報があるが、センサの故障やデータの改ざんにより異常なデータが入力される恐れがある。ダイナミックマップの要件を満たすために、2種類のクレンジングを組み合わせた構成およびアルゴリズムを検討する。

1. 背景と目的

近い将来普及していくことが予想されている自動走行システムでは、自動走行技術とデータ基盤技術を通じた交通データ等の利活用が相乗的に発展していくことが想定されており、データ基盤技術としてのダイナミックマップの役割が重要になっていくと考えられている。

ダイナミックマップは更新頻度別に静的、準静的、準動的、動的情報の4階層からなるデータの集合体であり、更新頻度が高い情報はデータが入力されてから地図を更新する時間的制約が強い。なお、現状ダイナミックマップにおいて更新が検討されている最も更新頻度の高いのは準動的の情報である。また、ダイナミックマップが生成する情報の中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）を利用するものが存在する。自動車の台数と各車両に設置されているセンサの数を考慮すれば、センサが故障している可能性は無視できない。また、センサデータを送信する過程で改ざん等により正常でないデータが発生することも想定される。

本稿の目的は当該データが発生した場合にダイナミックマップの運用に影響を与えずに取り除くこと（クレンジング）が可能かを検討することである。

2. 応用例

内閣府 SIP-adus の報告書[1]による、ダイナミックマップにおけるプローブデータを利用して生成する準動的情報の一つである「事故発生箇所」を例に挙げる。

「事故発生箇所」の情報生成アルゴリズムは、車両側とクラウド側での処理に分かれている。車両側では、加速度のモニタリングを行い、閾値を超えた場合に急加速度情報（加速度と位置情報）をクラウドにアップロードする。クラウド側では、過去N分間の急加速度情報と同一車両や周辺の平均車速を取得し、急に平均車速が遅くなった場合に事故発生箇所として警告する。

図1に[1]をもとに作成した「事故発生箇所」までの流れを示す。上記のアルゴリズムに基づき、加速度が閾値を超えた車両が、対象車両のメーカーを経由して、クラウド（ダイナミックマップセンター）に急加速度情報を送信する。クラウド側では事故発生箇所を判定し、DBを更新、生成した情報を対象メーカーに配信する。

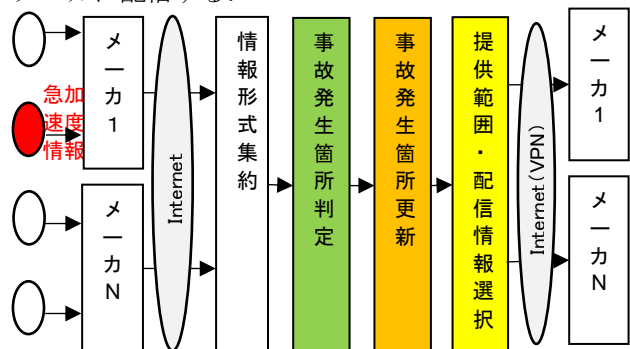


図1 「事故発生箇所」生成の流れ

ここで、センサの故障やデータ送信経路での改ざん等による異常なデータが入力されることは想定されていない。入力データが不正であれば「事故発生箇所」の判定が正確に行えずに誤った情報を生成してしまう。そのようなデータに対してのクレンジングが必要である。また、クレンジングにあたり情報生成時はリアルタイム性を損ねてはならず、データ基盤として参照される際には正確性を欠いてはならない。これらの理由から、特徴の異なる2つのクレンジングを併用する構成を考える。

Examination of probe data cleansing in dynamic map

† Yuki Akama

‡ Yuki Akama, Atsuhiko Goto

† Ark Information Systems

‡ Institute of Information Security

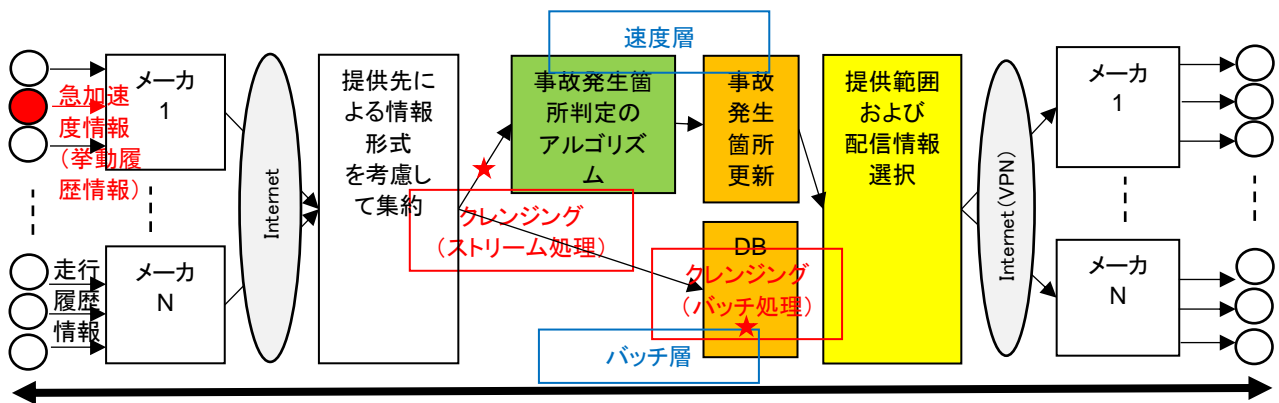


図2 ラムダアーキテクチャによるクレンジング構成

in 1min

3. プローブデータのクレンジング構成

2. で述べたクレンジングに関する要件を満たすために、ラムダアーキテクチャ[3]の考えを用いたクレンジング構成を図2に示す。図の速度層では、車両から送られてきた急加速度情報に対してリアルタイムでストリーム処理によるクレンジングを行う。他方、バッチ層では、前回のクレンジング以降入力された全データに対してバッチ処理によるクレンジングを行う。後者のクレンジングにより異常データが発見された場合はDBの修正と、必要に応じて生成した情報の訂正を行うことができる。

4. クレンジングアルゴリズム

アルゴリズムの選定のために、クレンジングの対象となるプローブデータを国土交通省[2]の「ITSスポットサービス」を参考に仮定する。プローブデータは基本情報、走行履歴情報、挙動履歴情報により構成される。基本情報には車両の搭載する無線機およびカーナビのメーカー、型番である。走行履歴情報は時刻、緯度、経度、高度および速度であり、前回蓄積時から200m走行時または、進行方向が45度以上変化した時点を最大約80km相当分蓄積する。挙動履歴情報は表1に示す通りであり、閾値を超えた時点のデータを記録するイベント型記録方式を取る。

表1 プローブデータにおける挙動履歴情報

挙動履歴情報	センシング周期	分解能	閾値
時刻	1sec	1sec	-
緯度	1sec	10 ⁻⁵ 度	-
経度	1sec	10 ⁻⁵ 度	-
方位	1sec	16方位	-
前後加速度	<0.3sec	0.01G	-0.25G
左右加速度	<0.3sec	0.01G	±0.25G
速度	<0.3sec	1km/h	-

この仮定のもと、速度層におけるクレンジングアルゴリズムとして、山西[4]の外れ値検知エンジン

SmartSifterの適用を検討している。SmartSifterはオンラインで統計的モデルを適応的に学習し、そのモデルに対する各データの異常度合いをスコアリングするものであり、閾値を超えたデータという特徴を持った入力データに対して有効だと考えられる。バッチ層におけるクレンジングアルゴリズムに関しては速度層と比較して時間的制約を考慮しなくても良いため、選択自由度は高い。適用例としてSmartShifterとの比較が文献内で行われているHawkinsのニューラルネットワークを用いた外れ値検知手法[5]を検討している。処理時間をかけられる点を活用し、大量のデータの学習により精度が向上することを期待している。

5. まとめと今後の課題

準動的情報の「事故発生箇所」を生成する際のプローブデータクレンジングに関する構成およびアルゴリズムを検討した。クレンジングアルゴリズムの選択や具体的な適用方法、データ量の見積もり、実験による結果の検証が今後の課題である。

参考文献

[1] 内閣府, http://www.sip-adus.jp/wp/wp-content/uploads/cao_2016_caol-01_01.pdf, (2017年7月10日参照)
 [2] 国土交通省, 『プローブデータに関する日米共同研究 評価報告書』, 2015年1月
 [3] N.Marz, "Big Data: Principle and best practices of scalable realtime data systems" Manning Publications Co.
 [4] 山西, 『情報論的学習理論の現状と展望』, 2001年1月
 [5] S.Hawkins, "Outlier Detection Using Replicator Neural Networks" Springer-Verlag Berlin Heidelberg 2002