

スマートフォンとCANとソーシャルメディアの統合に基づく 道路行政支援

Supporting Road Governmental Operations by Integrating Big-Data from Smart Phone, CAN Scanner, and Social Media

阿部 秀尚*1
Hidenao Abe

森田 武史*2
Takeshi Morita

田中 壘*3
Rui Tanaka

山口 高平*3
Takahira Yamaguchi

*1 文教大学
Bunkyo University

*2 青山学院大学
Aoyama-Gakuin University

*3 慶應義塾大学
Keio University

In this paper, we introduce a framework for supporting Japanese road governmental operations by integrating massive data from open governmental data, smart phones, OBD scanners, and social media. The scheme consisting of such massive data and integration process is called "big data" analysis in these days. For understanding user generated information about traffic problems on social media more concretely, we have constructed two liked data with RDF for road structural information. On the other hand, by using sensor data gathering tools on each car, we carried out an analysis for detecting particular action by using a data mining method. By integrating these layers of analyzed information, the framework aims to provide more concrete issues based on more voices not only from human beings but also from so many sensors.

1. はじめに

近年、情報通信技術の進展と共に、個人が扱える情報端末の性能が向上し、スマートフォンと呼ばれる各種センサを搭載した多機能型端末が一般に普及してきている。また、自動車における情報化は、CAN(Control Area Network)による車載コンピュータからの各種車載センサデータの活用とともに、車載型映像機器による情報の活用が求められている。これらのセンサデータや映像データについて、スマートフォンによる通信機能を利用することにより、リアルタイムでのデータ収集と分析が可能となることが期待される。

本研究では、スマートフォンとCANによって得られる自動車車載センサデータと道路状況に関するソーシャルメディア上のテキスト情報を統合し、道路行政施策の策定支援を目的とする。これにより、迅速な交通事故対策や改良計画立案などの行政施策に繋げていくため、時系列センサデータから特徴的な運転行動を検出するパターンを抽出し、利用者が日々発信する情報から道路を利用する上で危険を感じる箇所の客観的な提示の実現を目指す。

本稿では、道路行政のための道路管理基盤データの構築、および車載型センサデータ収集システムとデータマイニング技術を利用した分析事例について述べる。

2. ビッグデータの紐付けに基づく道路管理情報基盤の構築

本研究では、異なる形式・種類のデータおよびデータから得られるパターンを統合し、道路管理における価値の創造を実現する情報基盤を構築する。このため、行政機関が個別の目的毎それぞれに収集したデータの統合のために必要となる基礎データを構築し、共有と活用の方法を開発する。この基礎データを基に従来は不可能であった各種管理作業のニーズをビッグデータと称される高頻度・大容量・多様な形式として収集され

たデータからデータマイニング手法の適用によりパターンを統合することで可能にする。

従来、道路行政の現場では、業務に応じて部署毎に独自かつ個別に各種のデータを収集し、蓄積してきた。基礎データは、このように個別化されたデータから、共通機能を標準化し、関係部署間でデータの相互活用を可能とすることで時間的・空間的・コスト的制約を低減する。

基礎データの構築により、走行履歴情報基準に基づいた道路の実態把握管理や車輦情報の把握管理が可能になると考えられる。例えば、道路の構造物に関する基礎データと車輦重量区分毎の走行履歴データを用いることで、各構造物のより正確な疲労度や補修計画立案、日常の道路管理作業の計画立案の支援が想定される。

さらに、基礎データと走行履歴データ、ソーシャルメディア上での一般利用者の投稿したテキストや画像、動画データを紐づけることで、道路への要望や不満、あるいは道路構造物の異常や道路構造に関連する不具合事例の収集を可能にする。このため、事故の危険が感じられる地点名や運転者の特徴的な運転行動に関連する特徴語をソーシャルメディア上のテキスト情報からテキストマイニング手法で抽出し、各地点での走行履歴データからデータマイニング手法を用いて、特徴的なパターンを抽出する。これらのパターンと基礎データを照合し、より多くの箇所でのヒヤリハット事例を道路構造と合わせて、道路管理者に提示することを想定している。

本研究がこれまでの先行研究事例 [佐々木 12] などと大きく異なる点は、行政機関がもつデータを基に道路利用者に情報提供を一方的に行うのではなく、行政がもつ道路管理基盤データと道路利用者の要望や自発的な提供データをビッグデータ分析の枠組みを用いて統合し、情報としての価値を付与して応えることを目指している点である。

3. 道路管理基盤データの構築

従来、行政機関におけるデータの収集は、各部署に割り当てられた業務に依存し、それぞれがどのようなデータを収集しているかの把握が困難であった。これらのデータを有効に活用するため、共通の機能を標準化し、統合することが必要と

連絡先: 阿部秀尚, 文教大学情報学部情報システム学科, 〒253-8550 神奈川県茅ヶ崎市行谷 1100, 0467-53-2111, hidenao@a.email.ne.jp

なる．共通機能を標準化した公開データは Open Data と呼ばれ，Data.gov (米国)，data.gov.uk (英国) といった事例があるように各国政府が行政データとリンクしたデータ活用手法の開発を積極的に推進している．Open Data の利活用では特にセンサデータや情報システムに蓄積されるログデータとの連携が価値を高める上で重要な役割を果たすと期待されている．

3.1 既存公開データからの Linked Open Data の構築

森田らは，これまで日本語 Wikipedia を対象として，豊かな意味情報を持ち，他の Linked Open Data と関連付けが可能な Linked Open Data によるオントロジーを構築してきた [玉川 11]．Linked Open Data (LOD) は，セマンティックウェブの記述形式である RDF によって記述され，SPARQL などの問い合わせ言語を用いることで検索可能な仕組みで公開されるデータである．本節では，表記式のデータとして既に公開されている道路構造に関するデータを Linked Open Data としていくため，RDF 形式への記述形式の変換と SPARQL による問合せ受付を行う SPARQL エンドポイントの構築結果について述べる．

3.1.1 道路交通センサデータの Linked Open Data 化

道路交通センサデータは，「全国道路・街路交通情勢調査」の結果を国土交通省道路局が公開を行っているデータであり，CSV 形式では 10 個の基本属性，43 個の道路構造関連属性，26 個の交通量などに関する属性の計 89 属性の値が道路を一定区間に収集されている*1．我々は，本データを LOD 化するため，まず，区間の存在する市町村コードを総務省統計局が公開している標準地域コード*2からのプロパティとして関連付けを行い，RDF 形式への変換を行った．各区間 ID に対応した道路交通センサデータから RDF への変換の概観を図 1 に示す．以上の変換を全ての区間インスタンスに対して実行し，8,252,736 トリプルからなる RDF ストアを構築した．

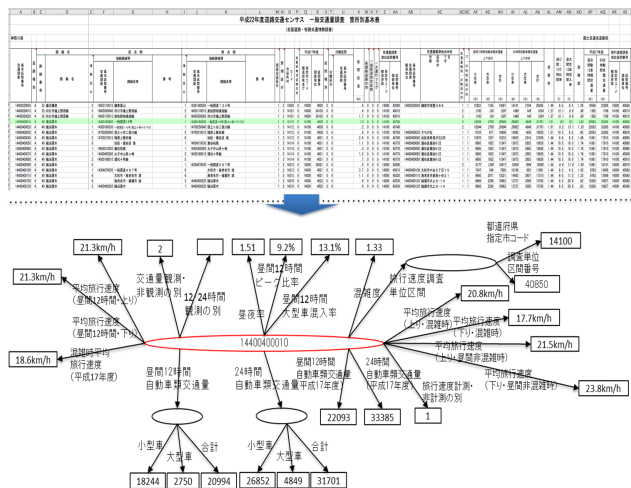


図 1: 道路交通センサデータの RDF への変換の概観．図下部のグラフは楕円がリソース，リンクがプロパティ，四角形がリテラルを表す．

*1 平成 22 年度 全国道路・街路交通情勢調査 (道路交通センサ) 一般交通量調査 : <http://www.mlit.go.jp/road/census/h22-1/index.html>
 *2 統計に用いる標準地域コード : <http://www.stat.go.jp/index/seido/9-5.htm> .

さらに，本形式で LOD 化した道路交通センサデータについて，問い合わせ言語 SPARQL を用いて問合せが可能な SPARQL エンドポイントを構築した．サーバ上での SPARQL クエリの実行例を図 2 に示す．

SPARQL

```
select ?id ?name ?value
where {
  ?id <http://zest.comp.ae.keio.ac.jp/road_census_id/> ?value.
  ?id <http://zest.comp.ae.keio.ac.jp/road_census_id/> ?name.
}
ORDER BY desc(?value)
```

(混雑度が高い順に区間IDと車線名を検索)

```
<?sparql xmlns="http://www.w3.org/2005/sparql-results#">
  <?head>
    <?variable name="id"/>
    <?variable name="name"/>
    <?variable name="value"/>
  </?head>
  <?result>
    <?binding name="value">
      <?literal datatype="http://www.w3.org/2001/XMLSchema#double">92.89</?literal>
    </?binding>
    <?binding name="id">
      <?uri>
        http://zest.comp.ae.keio.ac.jp/road_census_id/41400170120
      </?uri>
    </?binding>
    <?binding name="name">
      <?literal>久留米基山筑紫野線</?literal>
    </?binding>
  </?result>
  <?result>
    <?binding name="value">
      <?literal datatype="http://www.w3.org/2001/XMLSchema#double">46.25</?literal>
    </?binding>
    <?binding name="id">
      <?uri>
        http://zest.comp.ae.keio.ac.jp/road_census_id/6602320020
      </?uri>
    </?binding>
    <?binding name="name">
      <?literal>板谷米沢停車場線</?literal>
    </?binding>
  </?result>
```

図 2: 道路交通センサ LOD に対する SPARQL クエリの実行結果の一例．

3.1.2 DRM データの Linked Open Data 構築

DRM (Digital Road Map) データは，一般財団法人日本デジタル道路地図協会が提供している道路の区間 ID テーブルに関するデータである*3．本データは道路の区間 ID を利用した位置参照を提供することを目的としていることから，Linked Open Data 化することにより，道路交通センサ LOD の活用可能性を大きく広げることができる．各区間のデータは，区間 ID を識別子 (主語) とし，それ以外の値を目的語，属性を述語としたトリプルと呼ばれる RDF の構成単位に変換される．図 3 に DRM データを RDF に変換した LOD の一部を示す．

```
<?rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <?rdf:Description rdf:about="http://zest.comp.ae.keio.ac.jp/drm_id/64436400001">
    <?道路種別 xmlns="http://zest.comp.ae.keio.ac.jp/drm_id/">一般国道</?道路種別>
    <?リンク先 xmlns="http://zest.comp.ae.keio.ac.jp/drm_id/" rdf:datatype="http://www.w3.org/2001/XMLSchema#int">457</?リンク先>
    <?終点参照点ID xmlns="http://zest.comp.ae.keio.ac.jp/drm_id/">
      <?rdf:resource="http://zest.comp.ae.keio.ac.jp/drm_id/64436400001"/>
    <?起点参照点ID xmlns="http://zest.comp.ae.keio.ac.jp/drm_id/">
      <?rdf:resource="http://zest.comp.ae.keio.ac.jp/drm_id/64436400005"/>
    <?経度 xmlns="http://zest.comp.ae.keio.ac.jp/drm_id/" rdf:datatype="http://www.w3.org/2001/XMLSchema#double">143.56047</?経度>
    <?緯度 xmlns="http://zest.comp.ae.keio.ac.jp/drm_id/" rdf:datatype="http://www.w3.org/2001/XMLSchema#double">43.21211</?緯度>
    <?参照点種別 xmlns="http://zest.comp.ae.keio.ac.jp/drm_id/">次差点</?参照点種別>
  </?rdf:Description>
</?rdf:RDF>
```

図 3: DRM LOD の一例．

*3 道路の区間 ID テーブル : <http://www.drm.jp/etc/roadsection.html> .

現在、DRMデータのLODを利用することにより、区間位置のGoogle Map上への表示、および最短経路検索機能の実装、CANを利用した走行履歴データやソーシャルメディア上の投稿(テキスト)からの位置情報の同定方法の開発を進めている。

3.2 道路管理基礎データとセンサデータとの連携

本節で示した様な道路構造や交通状況に関するLODを道路管理基礎データとして、車載センサや位置情報をデータとして収集し、運転行動に関するパターンの発生位置の可視化やソーシャルメディア上での目撃情報などとの対応付けが可能となる。センサデータからの運転行動パターンと位置情報との対応付けとソーシャルメディア上での投稿情報は、一致して発生することは稀であることが考えられる。

しかし、道路構造や交通状況に関する基礎データを利用することにより、それぞれの発生箇所を地図上に表示することなどが可能となる。このため、センサデータからのパターン抽出とソーシャルメディアでの指摘情報は、相互に補完してより多くの問題箇所の指摘を行うこととなる。

4. スマートフォンとCANによる自動車センサデータの収集と分析

本研究では、自動車の挙動に関するセンサから得られるデータとして、スマートフォンによる位置情報、CANに接続されたDBDスキャナから得られる車載センサの各データを収集し、これらのデータを基に特定の運転行動を分析する。このため、スマートフォンを利用した位置情報収集ツールを作成し、実データでの運転行動の分析をデータマイニング手法によって行った。

4.1 スマートフォンとOBDスキャナによる運転行動データの収集

SmaphoProbeは、iOS上に実装された位置情報データ収集・表示ツールである[SmaphoProbe]。図4にシステムの構成を示す。本ツールの特徴は、図4に示すデータ項目がスマートフォンの各種センサから得られ、ユーザが設定するサーバに蓄積し、利用することを可能にした点にある。現在、ユーザ向けのデータ利用方法としては移動速度履歴のマップ上での可視化を提供している。

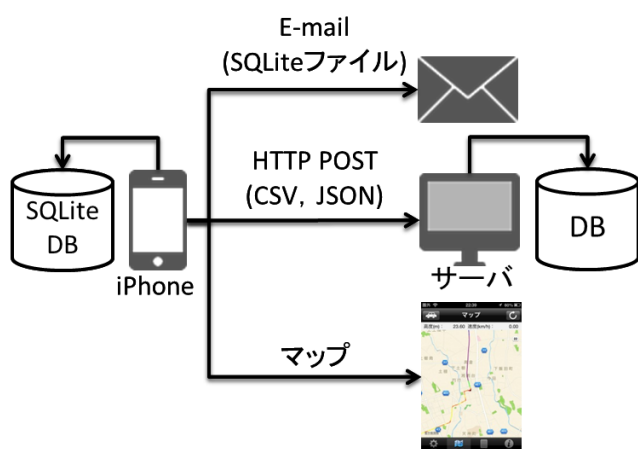


図4: スマートフォンを利用したプローブデータ収集システムの構成。

車載コンピュータからCANを通じて各種センサデータは、OBD(OBD-II)スキャナと呼ばれる機器を用いて取得することができる*4。これらのデータを統合し、位置情報と車載センサからの値を時刻毎に1データとすることにより、多項目の時系列データを収集する。

4.2 車載センサデータからの運転行動の分析

本節では、スマートフォンとOBDスキャナから得られた各種センサデータから、特定の運転行動として高速道路上での車線変更を対象とした分析について述べる。

本分析事例では、表1に示すように、実際に高速道路上での走行を行い、各種センサデータからの値を収集した。

表1: データ取得方法と取得データの内容。

データ取得日	2012年12月10日
データ取得区間	新東名高速道路 (沼津駿河湾SA~清水SA)
データ取得時間	15:17~15:39
取得センサ項目数	11
取得時点数	1,151

このとき同時に撮影したビデオ映像を基に各時点での車線変更の有無(車線変更の方向も区別)を人手によってラベル付けした。表1に示した各センサからの値に基づいて、各時点における車線変更の有無を説明するため、前後5秒間の時間窓から4種類の区間特徴量を図5に示すように抽出した。それぞれの区間特徴量について、各センサの時系列データ全体の平均からの外れ度合いを値とした。

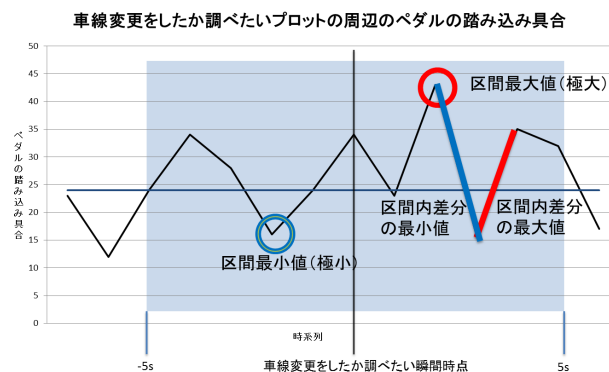


図5: 時点毎スライドする時間枠での区間特徴量。

以上の前処理をセンサから得られる11項目のうち、緯度・経度を除く9個の時系列データに適用した。この結果、35属性*5を持つデータセットを生成し、Weka[Witten00]に実装された決定木学習アルゴリズムJ4.8(C4.5)[Quinlan92]を用いて、車線変更の方向を分析する決定木の構築を行った。本データセットについて、属性選択・サンプリングともに行わなかった場合、クラス配分をサンプリングにより変更した場合[Abe07]、クラス配分を調整した遺伝的アルゴリズムを用いた属性選択

*4 海外では規格化され、コネクタの装着が義務づけられている場合もあるが、日本国内では早急な導入が検討されている段階である[自動車整備高度化検討会]。

*5 方位については、変化量のみを採用し、34属性を説明変数とし、クラス1属性を加える。

[Vafaie92]を行った場合、の3条件下での決定木構築を行った。クラス配分の調整は、最多クラスの割合を最少クラスの N 倍とするようにランダムサンプリングにより削減している。10回交差検証による各クラスの F 値・適合率・再現率を表 2 に示す。

表 2: クラス配分の調整および属性選択を行った場合の決定木学習アルゴリズムの実行結果 (10 回交差検証による各クラスの F 値・適合率・再現率)。

訓練データ	クラス	F値	適合率	再現率
全体のデータ (35属性、 1151データ)	車線変更無し	0.978	0.978	0.978
	右へ車線変更	0.707	0.725	0.69
	左へ車線変更	0.647	0.629	0.668
サンプリング (最多クラス倍 率8倍)	車線変更無し	0.944	0.944	0.944
	右へ車線変更	0.747	0.756	0.738
	左へ車線変更	0.836	0.824	0.848
サンプリング (最多クラス倍 率8倍) +属性選択	車線変更無し	0.941	0.934	0.948
	右へ車線変更	0.800	0.842	0.762
	左へ車線変更	0.758	0.758	0.758

この結果、最多クラスを最少クラスの 8 倍のデータ数となるようにランダムサンプリングにより削減したクラス配分のデータセットにおいて、各クラスについて最も正確な分類が行われた結果となった。このとき、最も多くの右車線への車線変更のデータを被覆する代表的なルールは以下の通りである。

```
IF      高度の変化量が減少方向に極端に大きい
AND    右への加速度が高い
AND    ペダルの踏み具合は極端に多いとは言えない
THEN  「右へ車線変更」
```

条件節の“極端に”という記述は、標準化したそれぞれの値である (z 値) が標準正規分布 $N(0, 1)$ に従うと仮定した際の上側確率 $p(z)$ が $p(z) < 0.05$ の値であったことを意味している。

本分析の結果から、スマートフォンを利用した車載データ収集ツールを利用することにより、各種車載センサーデータの値から特定の運転行動が発生した箇所が検出可能であることが示された。

5. おわりに

本稿では、センサーデータから収集されるデータ、ソーシャルメディアから得られるデータを行政機関がもつデータと対応付けを可能にし、行政施策の策定を支援する情報基盤の構築について述べた。このうち、道路構造情報基盤の整備では、公開された複数の道路構造情報を統合し、RDF を用いて大規模な Linked Data として構築した。一方、自動車の位置情報や各自動車の運転状況を車載センサーデータから収集するツールを作成し、高速道路路上での運転行動が時系列データの処理と決定木学習によって分析可能であることを示した。

今後は、道路構造情報に位置情報を付加し、ソーシャルメディア上で得られるテキストで指摘された危険な状況が道路構造や天候・時間帯などの多面的な側面からの原因分析を可能とすることが課題と考える。これにより、ソーシャルメディア利用者が危険を感じた箇所の道路構造や自動車の運転行動を重ね

ることで、より利用者に身近な交通問題を解決できる道路行政を支援するための基盤の構築を行っていく必要がある。

参考文献

- [佐々木 12] 佐々木 卓ほか 10 名: 持続可能な生活交通情報フィードバックシステムにおける時空間融合交通情報基盤の検討, 第 11 回 ITS シンポジウム 2012, 1-A-05 (2012).
- [玉川 11] 玉川 奨, 森田 武史, 山口 高平: 日本語 Wikipedia からプロパティを備えたオントロジーの構築, 人工知能学会論文誌, Vol.26, No.4, pp.504-517 (2011).
- [SmaphoProbe] Smapho Probe: <https://sites.google.com/site/smaphoprobe/>.
- [自動車整備高度化検討会] 自動車整備技術の高度化検討会: http://www.mlit.go.jp/jidosha/jidosha_fr9_000009.html.
- [Witten00] Witten, I. H. and Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann (2000)
- [Quinlan92] Quinlan, J. R.: Programs for Machine Learning, Morgan Kaufmann (1992)
- [Abe07] 阿部 秀尚, 津本 周作: ルール評価支援手法における有効な学習モデル構築に必要な最小訓練事例数の見積もり法, 人工知能学会第 3 回データマイニングと統計数理研究会予稿集 (A603), A603-10 (2007)
- [Vafaie92] Vafaie, H. and Jong, K. D.: Genetic algorithms as a tool for feature selection in machine learning, in Proceedings of the fourth International Conference on Tools with Artificial Intelligence pp. 200?204 (1992)