

## 論文の内容の要旨

論文題名 道路変化イベントの把握のための機械学習を利用した統合的システムに関する研究

氏名 小林 亘

本研究は、道路変化イベントの早期かつ網羅的な把握の実用化に向けて、道路変化イベントと工事情報を組み合わせるための統合的システムを構築し、機械学習を利用した道路変化イベントの予測と情報の流通の促進に取り組み、以下の内容を明らかにした。

第1章では、道路変化イベントを取り巻く現状と問題を研究の背景として概説した。道路管理者による情報の提供、そして、地図調製者による情報の収集が実施されているにも関わらず、これらがかみあっておらず、道路利用者がカーナビゲーションの地図の更新に不満を示しているという状況を示した。

次に、この問題の改善を図るための先行研究の調査から、主に3点の課題を抽出した。その一つ目は、情報利用の準備のためのリードタイムを得るには道路変化イベントに先立つ工事情報が有効であり、工事情報を活用するには道路変化イベントと対照する方法が必要であること。二つ目は、道路変化イベントの予測には、道路変化の種類、予測時期(2年後)というリクワイヤメントに応える予測方法、簡便な予測方法が必要であること、三つめは、実用化には情報の利用を容易にする流通環境が必要であることである。これらを本研究の課題として設定し、研究の取り組みの方向性を示した。

第2章では、最初に、一つ目の課題である、工事情報と道路変化イベント(これらをまとめて原典情報と呼んだ)を対照させる方法を提案した。原典情報は場所に固定したものであり、そして、時間的な順序を有する事柄を示す情報であることから、本研究では、(a)場所と路線による集約、(b)時期による整理、を用いる方法を提案した。場所と路線による集約においては、地名辞典として郵便番号データを選定し、これに改良を加えて2010年度の道路工事データにおいて87%程度の場所を大字レベルの地点に変換できた。また、原典情報の起点と終点の経由地点を補間する方法を合わせて提案した。時期による整理においては、公示日、供用日、開通日、工事完了日を用いて事業の流れにあわせて原典情報を点過程として整理する方法を提案した。

次に、この提案した方法が実現可能であることを検証するため、加えて、情報を可視化して確認できるように、(a)データベースシステムと(b)WEB マッピングシステムを中核とする統合的システムをデザインし構築した。

第3章では、統合的システムに実際の原典情報を投入し、供用開始情報、道路開通情報、道路区域情報、工事情報の実態を調査分析した。これによって道路変化イベントの把握の方向性を整理した。

調査分析の結果、道路区域に関する情報の告示が供用開始の告示に先立って実施されている傾向は大きくは見られなかった。このため、道路区域情報を道路変化イベントの予測に利用することは現時点では考えにくいことが分かった。道路区域情報には対象区間の幅員と延長が示されている。他方の供用開始情報には供用開始日は必ずあるが、対象区間の規模を明示する項目は（道路法施行規則では示されていないので）有るとは限らない。調査分析から、道路区域情報の約4割は供用開始情報と同日同地点で公表されている傾向が見られた。このため、道路区域情報と供用開始情報とを組み合わせると、道路変化イベントの規模と時期が得られる可能性があることが分かった。

次に、工事情報と道路変化イベントの関係について調査分析を行った。ただし、工事の対象は契約額2500万円以上である。道路変化イベントのうち、道路開通情報の該当地点では工事が行われた割合が高く、しかも工事件数が他の地点に比べて多いことが分かった。供用開始の該当地点では、道路開通情報の地点には及ばないものの、工事が行われている地点が、公共事業費の多い自治体ほど高い割合で観測された。

第4章では、工事情報から道路変化イベントを予測することの利点、評価方法といった予測に共通的なことを最初に示した。利点として、(a)リードタイムの確保、(b)道路法の対象外の道路の把握、(c)道路変化イベントに関係する工事の絞り込みが得られる、ことを示した。評価方法として、道路変化イベントを把握できた割合である真陽性率（1に近いほど良い）、道路変化イベントの無い場所を誤って有ると予測した偽陽性率（0に近いほど良い）、これらを総合的に示すROC曲線とAUCを採用することを述べた。

次に、工事完了件数から道路開通を予測する方法を提案し、実際のデータにより検証した。この予測は、公共工事から道路工事を、主として発注者、工事件名、工事場所によって抽出し、その工事場所と工事完了日を第2章で提案した関連づけ手法で整理し、一定期間内の工事件数から開通を予測するものである。この予測方法は、工事の内容を入手する必要が無く、簡便に行えるという利点がある。その予測結果は、3つの自治体の契約額2500万円以上の工事を用い、1年度に2件以上の工事がある地点を開通と予測するルールによって真陽性率0.69、偽陽性率0.43が得られた。同じ条件で500万円以上の工事を用いたときには真陽性率0.92、偽陽性率0.35であった。このことから、予測の適用の範囲は開通の短期予測（数箇月）と制限はあるものの、簡便な方法によっても予測が可能であることを示した。

第5章では、道路変化イベントに対する多様な情報収集のリクワイヤメントに応えるために、工事内容を利用した予測の方法について述べた。工事内容とはアスファルト舗装や

橋梁上部工などの工事概要あるいは工法・型式などに示されているものである。これらの種類は多岐にわたるため、予測に用いる説明変数（素性，特徴量）も多変数となる。したがって、高次のデータ空間を対象として予測を行える機械学習（教師あり学習）を適用した。

本研究では、機械学習によって予測するクラスを、規模別に道路開通，供用開始の2種類とし、それぞれについて時期別に短期（工事期間の数箇月程度）と長期（道路変化イベントの2年前）の2種類の組み合わせの合計4種類の予測を行うこととした。

機械学習の準備として、工事内容を説明変数に次元圧縮する必要性と方法を整理した。次に、規模と時期の異なる予測クラスごとに教師データとテストデータを生成する方法を提案した。予測は、教師データから予測アルゴリズムによって学習モデルを生成し、テストデータにこのモデルを適用することで行う。機械学習に用いるアルゴリズムには、未知の（教師データと異なる）工事データに対する予測誤差（汎化誤差）を評価するため、誤差に対して特性が異なる線形分類器（ナイーブベイズ分類器，AODE），非線形分類器（決定木，ランダムフォレスト）を使用した。工事情報を集積する期間（情報収集期間）は3種類（6箇月，4箇月，2箇月）とした。

実験の結果，道路開通は供用開始に比べて成績が良いこと，短期予測は長期予測より成績が良いこと，情報収集期間による成績の差には決まったルールが発見できないこと，非線形分類器では過学習により真陽性率が低下することなどの知見を得た。これらをもとに予測の利用フローを作成した。

予測結果の真陽性率は供用開始，長期予測ではおおよそ0.6以下であり，予測の正確さの向上が課題である。データの質・量の充実のほか，説明変数の改良，スパースデータに適したデータマイニングアルゴリズムの適用，工事の順序関係（マルコフ連鎖）を利用する方法を課題への対応案として示した。

第6章では，研究の実用化に必須な事項である，情報の流通に関する取り組みを述べた。情報の流通には，情報の提供者である道路管理者の理解が必要であると考えた。そのため，道路管理者に対してアンケートを実施した。調査の意図は，研究の実用化において，(a)対象とする原典情報の拡張の必要性，(b)(c)道路管理者における情報流通の障害の有無，(d)システムの活用可能性を調査することである。(a)道路変化イベントにおける有用な情報に関する調査では，本研究で扱った原典情報以外に効果的なものは見つからなかった。(b)道路変化イベントの公表への意思に関する調査から，73%の道路管理者が公表の意思を持っていることがわかった。(c)情報の流通に対する問題の有無では，79%の道路管理者から工事データの利用について概ね受け入れられた。(d)本研究で構築したシステムの有用性の評価については，既に庁内システムを除く道路管理者から有用と評価された。

なお，道路管理者への調査は，道路管理者へ道路変化イベントを取り巻く問題点に理解を促すこと，研究の意図を浸透させ，システムの認知を高めることも意図している。

第7章では、総括として本論の成果をまとめた。加えて本研究の独自な点として次の3点について述べた。

- (a) 情報に含まれる時間（時期）を利用したこと。
- (b) 統合的で共通して利用できるシステムを構築したこと。
- (c) 情報の流通への取り組み

また、本研究の有用な点として次の3点について述べた。

- (a) 原典情報の関連づけ手法
- (b) 工事情報を用いた予測
- (c) 統合的システムとフォーマット

最後に、今後の展望として予測結果と詳細調査との連動による利便性の向上、道路変化イベント以外の予測への拡張について述べた。

工事データから道路変化イベントを把握することにより、道路管理者にとっては供用した道路について、道路利用者・地図データ等の制作者等への情報提供を効率的に行うことができるほか、情報公開請求への対応が軽減できる。道路利用者にとっては、利用できる道路の情報が、管理者のサイト、地図やカーナビゲーションを通じてこれまでより早く入手できるため、効率的な道路の利用が可能となる。地図データ等の制作者にとっては、全国の道路管理者への問い合わせを効率化できる。これらによって社会的な便益が向上する。