

# 工事实績情報を用いた道路供用の予測に関する研究

小林 亘<sup>1</sup>・柴崎 亮介<sup>2</sup>・関本 義秀<sup>3</sup>

<sup>1</sup>正会員 建設情報研究所次長 (一般財団法人) 日本建設情報総合センター  
(〒107-8416 東京都港区赤坂 7-10-20)

E-mail: kobayashi\_w@jactic.or.jp

<sup>2</sup>正会員 東京大学教授 空間情報科学研究センター (〒277-8568 千葉県柏市柏の葉 5-1-5)

E-mail: shiba@csis.u-tokyo.ac.jp

<sup>3</sup>正会員 東京大学特任准教授 空間情報科学研究センター (〒277-8568 千葉県柏市柏の葉 5-1-5)

E-mail: sekimoto@csis.u-tokyo.ac.jp

道路地図やカーナビゲーションの地図データにおいて、道路の供用に整合して更新がなされることは、道路の利便性の向上に寄与する。しかし、道路の供用の情報を得るには、道路管理者毎の公示内容等を調査しなければならない、手間と時間を要している。これを改善するため、公共工事の品質確保のために収集されている工事实績情報から道路の供用を予測できないだろうか考えた。研究によって、工事实績情報から一定の期間内に行われた工事を選定し、使用された工法・型式を説明変数に変換し、さらに場所ごとに集約することで予測用データを生成し、これに教師あり機械学習を適用する手法を考案した。これを埼玉県  
の工事实績情報、道路開通情報により検証し、予測が可能であることを示した。

**Key Words:** Construction-Records, Road update Information, Machine Learning, Decision Tree

## 1. はじめに

建設工事受注動態統計調査によると、公共機関による道路工事は2009年度には工事件数で56,280件、請負契約額で29,380億円であった<sup>1)</sup>。道路工事には維持補修も含まれているが、一般社団法人日本道路建設業協会のまとめによると、2009年度において2,884億円分、道路の全舗装費のうち45.2%が舗装新設に充てられていた<sup>2)</sup>ことから、これに見合う道路が新たに供用されたと考えられる。供用された道路の交通・輸送機能を十分に発揮させるには、道路地図やカーナビゲーションの地図データが実際の道路と整合し、最新の状態に更新されることが必要である。

しかし、道路地図等の更新に必要な道路の供用に関する情報は道路管理者毎に管理されており、多くは道路管理者毎に提供されているため、情報の収集には多くの労力と時間が必要となっている<sup>3), 4)</sup>。

道路地図等の更新に必要な情報が一元的に管理されていない一方で、公共工事の実績情報は、その多くが「工事实績情報」として一元的に集積され、公共

工事の品質を確保するため、発注機関による企業や技術者の実績確認に利用されている<sup>5)</sup>。

近年、データを活用して価値ある情報を生み出す動きが活発化している。公共工事の分野においても工事实績情報に見られるようにデータの蓄積は進んでおり、その収集、蓄積に要した費用、時間、労力は少くないことから、これを社会に役立てられるよう検討することは意義のあることだと考える。

このような観点から工事实績情報を捉えると、工事实績情報には工事と道路地図等の更新との関係を明示的に登録する項目は無いが、データマイニングによって道路地図等の更新に関する道路の新設(開通)、拡幅、大規模な交差点改良などの「道路の供用」に関する情報を得ることができれば、道路地図等の更新作業を改善することができる。

そのため、工事实績情報に対して機械学習によるデータマイニングを適用し、道路の供用を工事実施時に推測することの可能性について研究を行った。

表-1 工事实績情報の例 (一部の項目のみ)

項目	内容
自社が請け負った一体的先行契約工事の有無	一体的先行契約工事無し
工事件名	(一)三刀屋佐田線 吉野工区 社会資本整備総合交付金(改良)工事
路線・水系名等	一般県道三刀屋佐田線
請負金額	XX,XXX,XXX(円)
工期	2010年08月05日~2011年02月28日
発注機関	島根県出雲県土整備事務所
受注形態	単独
VE対象工事	VE対象工事ではない
ISO対象工事	ISO対象工事ではない
工事分野	道路
工事の業種	土木一式工事
工事種別	一般土木工事
工種/工法・型式	土工事/掘削または切土工 水路・管路工事/水路・管路工 法面工事/法面工 アスファルト舗装工事/アスファルト舗装工 軟弱地盤処理工事/表層混合 処理工法
施工場所	島根県出雲市佐田町吉野地内
施工地域	その他の地域
夜間工事の有無	夜間工事無し
交通規制の有無	交通規制有り
規制道路の種類	都道府県道
規制道路の交通量(台/日)	5,000未満
規制車線数	規制車線数1/全車線数1
交通の確保手段	通行止 片側交互通行
近接施工の有無	近接施工無し

## 2. 工事实績情報

公共工事の発注機関では、企業や技術者の工事实績を確認するため、公共工事の工事实績に関する情報(工事实績情報)を収集、蓄積、提供する「工事实績情報システム(コリンズ)」が利用されている。工事实績情報の例(一部の項目のみ)を表-1に示す<sup>6)</sup>。1993年、「中央建設業審議会」において工事实績情報のデータベース整備の必要性が建議され、1994年から工事实績情報システムの運用が開始されて以来、工事实績情報の蓄積は順調に進み、2011年3月における登録件数は約345万件、利用機関は、国土交通省、47都道府県、19政令市、791市区町村など合計1030となっている<sup>5)</sup>。これらの機関では一般的に、契約に使用する土木工事共通仕様書<sup>7)</sup>にて、「請負金額500万円以上の工事について、契約、契約変更、工事完成のそれぞれの時点で10日以内に発注機関の確認を受け、工事实績情報を工事实績情報システムへ登録すること」を受注者に求めている。工事实績情報の公共工事における捕捉状況を確認するため、統計法に定められた基幹統計である「建設工事施工

統計調査」<sup>8)</sup>との比較を行った。2006年度から2010年度までの建設工事施工統計調査における元受完成工事高に対して、工事实績情報システムに登録された工事实績情報のうち、建設工事施工統計調査と同じ発注機関<sup>9)</sup>を対象とした工事の請負金額(各年度に完了登録されたもの)の比率は、107.4%~98.7%であった。このことから、多くの公共工事は工事实績情報システムに登録されており、従って、工事实績情報の活用の有用性も高いと考えられる。

## 3. 道路の供用に関する情報

道路法第十八条では、道路の供用について、「・・・道路管理者は、道路の供用を開始し、又は廃止しようとする場合においては、国土交通省令で定めるところにより、その旨を公示し、かつ、これを表示した図面を道路管理者の事務所において一般の縦覧に供しなければならない。・・・」と定めている。公示によって情報を収集するには、全国の行政機関等の公示内容を調査し、道路の供用に関するものを抽出しなければならない。これに対し、工事实績情報から道路の供用を推測することが可能となれば、公示に先立って、工事契約後10日以内の登録の時点において供用の場所を予測できることとなり、計画的に道路地図等の更新の準備を進められる。また、工事实績情報は電子化され、一元化されているので、公示に比べて容易に情報を収集できる。

## 4. 本研究において採用した手法

### (1) 工事实績情報と道路の供用

工事实績情報には、件名、請負金額、工期、発注機関、契約形態、請負者、工事の分野、建設業許可業種、入札参加資格区分、工種、工法・型式、施工場所、技術者などが登録されている<sup>10)</sup>。しかし、工事实績情報は工事の品質を確保するために、「工事」という生産プロセスにおける請負会社と技術者の実績を登録することを主な目的としている。このため、工事の生産物や目的の登録は任意である<sup>11)</sup>。

さらに、道路の供用の登録について考えると、後述するように供用に必要な工事は一つとは限らないことに加えて、それら全てが完了し、関係機関との調整や所定の行政事務手続きを経なければ供用とはできないことから、工事の受注者が工事関係図書から登録することはできないと考えられる。

このように、工事实績情報から道路の供用の情報

表-2 「工種」と「工法・型式」の一例

工種の選択肢	工法・型式の選択肢
トンネル工事	NATM, 矢板トンネル, TBM
	開放型シールドトンネル
	密閉型シールドトンネル
	開削トンネル・半地下
	沈埋トンネル
	非開削特殊工
軟弱地盤処理工事	置換工法
	プレローディング工法
	表層混合処理工法
	バーチカルドレーン工法
	サンドコンパクション工法
	ロッドコンパクション工法
	パイロフロテーション工法
	石灰パイル工法
	深層混合処理工法
	DC工(動圧密工法)
サーチャージ工法	
軽量土工法	
軟弱地盤グラウト工事	薬液注入工法
ボーリンググラウト工事	ボーリンググラウト工
コンクリート構造物工事	カルバート工
	鉄筋コンクリート橋
	橋梁下部工
	鉄筋コンクリート床版工
	キャブ工
	コンクリート擁壁工
	フーチング工
	情報ボックス
	電線共同溝

を直接的に得ることはできない。

(2) 採用する手法

工事実績情報から道路の供用に関する情報を推測することを、明示されていない有用な情報をデータから抽出するデータマイニングの応用の一つと考えた。そこで、本研究では、多くのデータからコンピュータ(機械)によって未知のルールを発見することに適した手法である機械学習によってデータマイニングを行うこととした。中でも、目的が「道路の供用の有無の識別」と明確であるため、目的を明示した訓練データによる学習である「教師あり機械学習」を利用することとした。

5. 説明変数

(1) 説明変数の選定

本研究は、道路の供用を予測するために使用する工事実績情報のデータを「説明変数」とし、道路の供用の有無を「目的変数」として、目的変数別に説明変数を分類する問題と整理される。説明変数には、工事について技術的な情報が得られること、全ての工

事に必ず登録されていること、語句の揺らぎが少ないこと、という条件を課した。「工種」、「工法・型式」はいずれも技術的な項目であり、全ての工事実績情報に最低1種類、最大5種類の登録がなされており、登録されるデータには選択肢が用意されている。「工法・型式」の選択肢は137であり、「工種」はこれが集約され77となっている<sup>12)</sup>。本研究の目的に合致するように説明変数を整理するには、集約される前の細かな区分であることが望ましいため、「工法・型式」を説明変数の候補とした。

(2) 説明変数の次元の圧縮

説明変数から工事による生産物の機能(工学的な意味での工事目的)や工事の段階を捉えたい。ここで、同じ目的や段階の工事でも、工事現場の条件によって工法・型式が異なるものに個別に説明変数を割り当てることは有害となる<sup>13)</sup>。このため、区別の必要のない工法・型式をまとめ、説明変数を削減する。これは次元の圧縮と呼ばれ、次の効果が得られる。

a) 計算量の軽減

一つの工事には1以上5種類以下の工法・型式が登録されるので、137種類の選択肢を説明変数に使用すると工事当たりの組み合わせは、「登録なし」を138番目の選択肢と考えて、 ${}_{138}C_5 - 1 = 387610811$ (約3.9億通り)となる。仮に、説明変数を元の選択肢137種類の約1割まで減らして14種類とした場合には ${}_{15}C_5 - 1 = 3003$ と、組み合わせ数は約13万分の1に軽減される。予測には多数の工事を組み合わせた計算が必要となるので、計算量の差はより大きなものとなる。

b) 汎化能力の向上

工事現場の条件の違いの影響を取り除くことにより、特定の条件に過剰に適合した学習を防ぎ、汎化能力、すなわち新たなデータを用いた時の予測性能を向上させることができる。

(3) 次元圧縮の具体的な方法

工事の体系によって「工種」と「工法・型式」とを対比すると、工種である「トンネル工事」の下位に6種類の工法・型式(NATM, 密閉型シールドトンネル等)が、「軟弱地盤処理工事」の下位に12種類の工法・型式(置換工法, 深層混合処理工法等)があるなど、一つの工種に複数の工法・型式が属するものが見られる。表-2にその例を示す。上述したトンネル工事, 軟弱地盤処理工事の例では、工法・型式は工事現場の条件などによって異なるが、工事の目的や段階の大要は工種が示しており、工種を説明変数とすることが合理的である。さらに軟弱地盤処理工事で



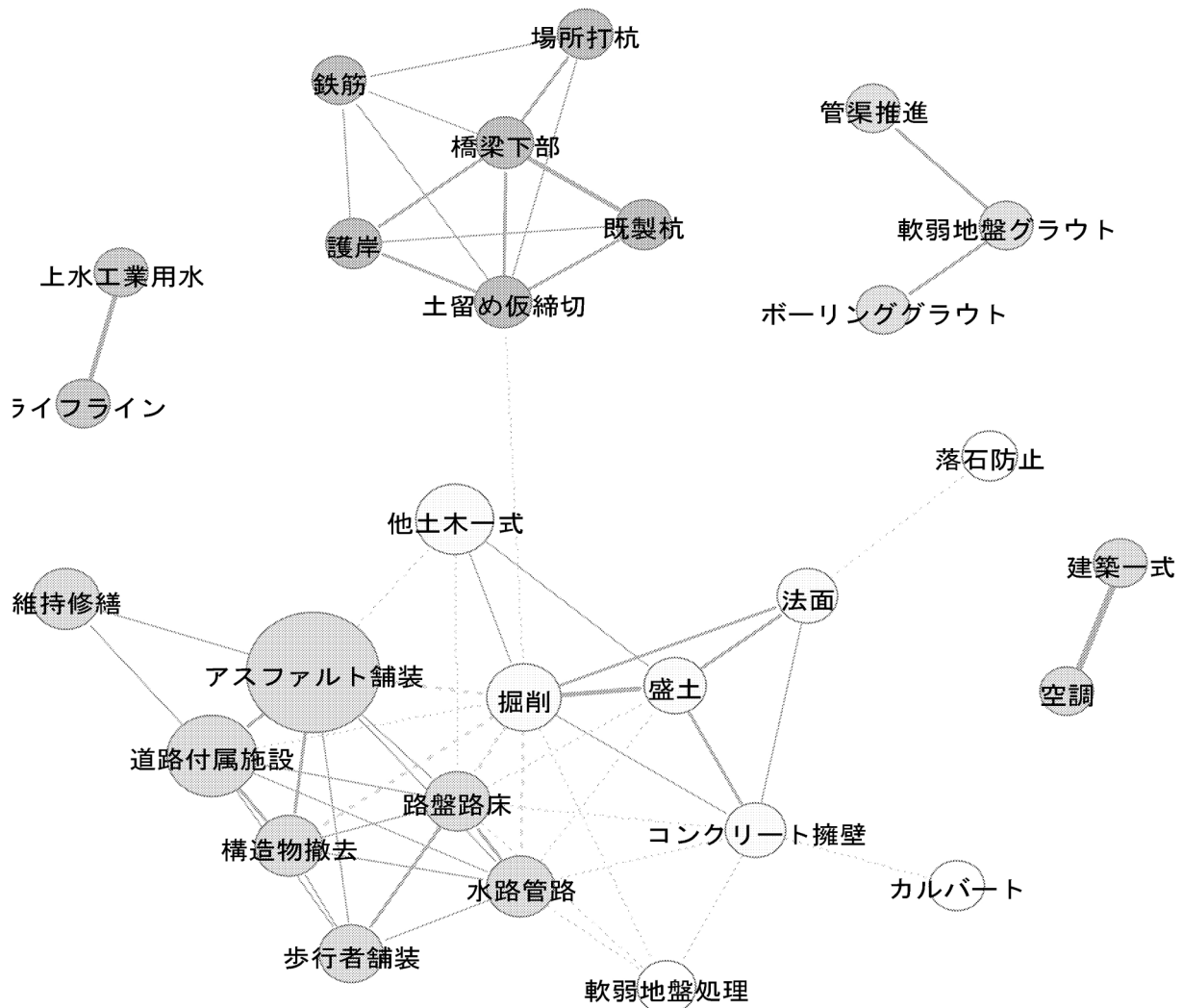


図-1 工法・型式の共起ネットワーク

は、目的から軟弱地盤グラウト工事、ボーリンググラウト工事などの他の工種も統合できる。

他方、工種「コンクリート構造物工事」に属する工法・型式は、橋梁下部工、鉄筋コンクリート床版工、コンクリート擁壁工、フーチング工、情報ボックスなどであり、これらの工法・型式の目的や段階は異なるため、工種で代表させることはできない。個々に他の工種の工法・型式との統合について検討が必要である。

さらに、道路の供用に関係の無いと考えられる工法・型式、例えば「海洋、波力等発電工」などについては削除できる。

以上の考え方によって、工法・型式の施工上の目的や段階を体系的に整理した資料<sup>14), 15), 16), 17)</sup>に基づき、さらに実際に各工事実績情報に登録されている工法・型式の出現回数、工法・型式の共起状況、クラスター分析による類似度を確認しつつ、工事の目的や段階から工法・型式を21の説明変数に編成した。図-1に確認に用いた工法・型式の共起状況の一

例を、表-3に最終的に策定した説明変数を示す。

## 6. 予測計算の対象単位

工事毎に道路の供用の有無を識別することが適当であるならば、工事毎に作成されている工事実績情報毎に予測計算を行える。しかし、工事実績情報によって、一つの工事では、道路建設のある部分、ある工程だけを行っているものがあることが明らかである。後述の図-2はその一例である。

このため、複数の工事を集約し、それを予測計算の対象単位とすべきであると考えた。そして、集約の単位は、工事が行われる「場所」が適当であると考え、説明変数を場所を単位として集約し、予測計算を行うこととした。

表-3 説明変数

No.	説明変数	工法・型式
1	Foundation	カルバート工、置換工法、プレローディング工法、表層混合処理工法、パーチカルドレーン工法、サンドコンパクション工法、ロッドコンパクション工法、パイプロフレーション工法、石灰パイル工法、深層混合処理工法、DC工(動圧密工法)、サーチャージ工法、軽量土工法、薬液注入工法、ボーリンググラウト工、土留め・仮締切工、既製杭工、場所打杭工、深礎杭工、地中連続壁工
2	Earth	土地造成工掘削または切土工、盛土または埋戻し工
3	Slope	法面工
4	Wall	コンクリート擁壁工
5	WaterWay	水路・管路工、伏せ越し工
6	RoadBed	路盤路床工、鉄筋コンクリート床版工
7	AsPavement	アスファルト舗装工
8	WaterProof	防水工
9	BrigdeBase	橋梁下部工、鋼製橋脚工
10	BridgeSS	PC橋上部工、鋼橋上部工
11	SmallBridge	鉄筋コンクリート橋
12	Tunnel	NATM・矢板トンネル・TBM、開放型シールドトンネル、密閉型シールドトンネル、開削トンネル・半地下、沈埋トンネル、非開削特殊工
13	CoPavement	セメントコンクリート舗装工
14	CCBox	キャブ工、情報ボックス、電線共同溝
15	PedestrianPV	歩行者系舗装工
16	RoadEquipment	道路付属施設工、建設電気設備工、トンネル非常用設備工、トンネル換気設備工、通信工
17	PC	PC構造物工
18	Works	その他のとび・土工・コンクリート工、上部・場所打コンクリート工
19	Prime	その他土木一式工
20	Removal	構造物撤去工
21	Maintenance	維持修繕工、道路清掃作業

(1) 工事实績情報における工事場所の登録

工事实績情報では「施工場所」中に次の4種類の工事場所が登録できる<sup>11)</sup>。

- ・起点もしくは代表地点の住所(以下、「起点住所」)
- ・終点の住所(以下、「終点住所」)
- ・起点もしくは代表地点の座標(緯度経度)(以下、「起点座標」)
- ・終点の座標(緯度経度)(以下、「終点座標」)

上記のうち、起点住所の登録は必須であるが、起点住所と同一の場合の終点住所の登録は不要であり、また、座標の登録は起点、終点とも任意となっている。起点住所、終点住所の登録の説明には、「都道府県名 - 市区町村名 - 丁目番地等」との例示がされている。これをより正確に表現すると、「都道府県 - (町村の場合には郡) - 市区町村・特別区 - (政令指定都市の区) - 町・字 - 番地」となる。ここ

で、「町・字」は地方自治法第260条における「市町村区域内の町又は字の区域」である。以下、この表記を「住所表記方式」と呼ぶ。座標は度分秒(60進法)での登録であり、1秒を最小単位としている。

(2) 場所の正規化とその識別能力

工事場所毎の集約には、工事場所を統一された体系を有する地名辞典などで正規化する必要がある。正規化に必要な場所の識別能力、すなわち、同一とみなせる場所、異なるとすべき場所の大きさについて、住所表記方式に沿って考える。

平均的な市町村(東京都特別区を市扱い)の面積は、日本の総面積(388,930km<sup>2</sup>)を市町村の総数(1,742)で除し、223.3km<sup>2</sup>と得られる。これが市町村を正規化の水準とした場合の平均的な識別能力となり、この区域を正方形と考えるとその一辺は14.9kmとなる。

市区町村より小さな区域を示す町・字相当を正規化の水準とした場合、国土交通省の大字・町丁目レベル空間位置参照情報<sup>18)</sup>の大字・町丁目の総数253,818を用いると一辺が1.24km、日本郵便の郵便番号の7桁の番号が表す範囲(町域)<sup>19)</sup>を後述の方法で精査したものの総数118,155を用いると一辺が1.81kmの区域となる<sup>20)</sup>。

町・字相当の水準であれば、同一地域の道路において、関連の無い工事が同時期に行われる可能性は少ないと考え、本研究ではこれを正規化の水準として用いることとした。

なお、座標の最小単位1秒は緯度30度の地点で27m(緯度45度では22m)に相当し、町・字相当の水準に比べて識別能力が高いが、座標は全ての工事に登録されていないため、住所を補助する情報としてリバースジオコーディング(緯度経度座標から住所表示方式へ変換)したうえで、住所表示方式と同じ町・字の水準で扱うこととした。

(3) 使用する地名辞典

町・字相当の水準で正規化を行うにあたり、地名辞典の候補として次の3種類を比較した。

- ・郵便番号データ(日本郵便)<sup>19)</sup>; 118,155地名
- ・大字・町丁目位置参照情報(国土交通省)<sup>18)</sup>; 253,818地名
- ・街区レベル位置参照情報(国土交通省)<sup>18)</sup>の大字・町丁目; 220,992地名

日本郵便(郵便事業株式会社)の郵便番号は、町域(町名から〇丁目を除く部分、および大字)に設定されており、小字または通称には原則として設定されていない。大字・町丁目レベル位置参照情報は、日

表-4 工事実績情報の工事場所と正規化

年度	2006	2007	2008	2009	2010
工事件数(件)	85,984	82,285	77,894	87,077	83,872
起点住所に登録あり(工事件数比%)	100.00	100.00	100.00	100.00	100.00
終点住所に登録あり(同)	17.26	18.52	18.78	36.35	43.17
路線に登録あり(同)	28.86	28.28	27.43	52.95	63.88
始点座標に登録あり(同)	0.02	0.03	0.09	31.80	42.55
終点座標に登録あり(同)	0.02	0.02	0.06	20.59	28.50
郵便番号データにて正規化(工事件数比%)	71.78	74.37	75.02	81.46	84.00
大字・町丁目レベル位置参照情報にて正規化(同)	56.17	58.20	58.85	63.83	66.67
街区レベル位置参照情報(大字・町丁目)にて正規化(同)	45.52	48.29	49.23	52.80	54.28
郡, 都道府県, 路線名, 路線番号による補完後, 郵便番号データにて正規化(同)	74.87	77.34	78.00	84.29	87.09
地名辞典に無い町・字相当地名を含めた(同)	83.42	85.88	86.31	90.16	91.71

本における住所体系のうち, 市, 町, 村, 区, 特別区の直下に属す行政区である「大字」, 「町丁目」, 自治体によっては「町字」を示す住所代表点とされている。街区レベル位置参照情報は, 大字・町丁目レベル位置参照情報のデータ整備年度が2009年であるのに対して2011年度となっているが, 都市計画区域相当範囲に限られている。

2006年度から2010年度に完了した道路関連工事における工事場所の登録状況と, これら3種類の地名辞典を用いて工事場所の正規化を行った結果を表-4に示す。この結果から, 最も多くの工事場所が正規化され, 場所を特定することができた郵便番号データを地名辞典として使用することとした。以下, 「町域」とは郵便番号データにおける町域を表す。

(4) 工事場所の特定

a) 郵便番号データ

工事場所の正規化に用いた郵便番号データ118,155地名は, 2012年6月22日の全国郵便番号

データに2009年9月30日更新分から2012年6月29日更新分までの削除された郵便番号データ行を加え, さらに次の処理を行ったものである。

- ・「以下に記載の無い場合」などの町域が示されていないデータ行, 高層ビルの各階毎に割り当てられたデータ行を削除する。
- ・(町村) 一円といった具体的な町域名を表さない要素を削除する(ただしデータ行は削除しない)。
- ・複数行にまたがって一つの括弧をなすデータ行も含め, 全ての括弧を閉じた形にした上で削除する。
- ・読点(,)や範囲指定(~)を分解し, 省略されている部分を補って独立したデータ行として追加する。
- ・重複を排除し, 最長一致を行うために文字列長の降順で並べ替える。

b) 工事実績情報

工事実績情報については, 都道府県名と町村名が存在し, 「郡名」を欠くデータについて「郡名」を補完する。次に, 市区名または郡と町村名の組が存在し, 都道府県名を欠くデータについて「都道府県名」を補完する。ただし, 同一名の伊達市(北海道, 福島県)と府中市(東京都, 広島県)については補完を行わない。

c) 工事実績情報と地名辞典の照合

工事実績情報, 地名辞典の双方に対して, 「ツ」と「ッ」, 「ケ」と「ヶ」の表記の揺らぎを統一し, 空白の除去などのデータクレンジングを行い, 市区町村名と町・字名の間の「字」, 「大字」の揺らぎに対応したうえで最長一致の原則で照合を行う。

複数の住所が含まれている場合には, 個別に取り出し, 起点住所, 終点住所, 起点座標, 終点座標で重複する場所がある場合は取り除く。

d) 路線名, 路線番号の活用

これまでの作業によって得られた場所に, 路線名, 路線番号を組み合わせて, 市区町村名と路線名に対応する町域, 市区町村名と路線番号に対応する町域のデータの組み合わせのリストを作成し, これらを用いて, 市区町村名と路線名または路線番号を有する工事について町域名の推定を行う。

e) 地名辞典に存在しない町・字相当の地名

郵便番号データの町域名に存在しない市区町村以下の地名については「他」や「地先」などの地名以外の文字列を除去し, そのまま住所として採用する。



場所名 ; [ 埼玉県日高市高萩 ]

							+ ; 開通日 2010/03/27	
2006. 4	2007. 4	2008. 4	2009. 4	2010. 4	2011. 4	2012. 4		
- ---9---1--	---9---1--	---9---1--	---9---1--	---9---1--	---9---1--	---9---1--	---	---
		S		+	E	**	RoadEquipment	
		S		+	***		AsPavement	
		S		+	****		AsPavement, RoadEquipment, Prime	
		S	*****		E		採用 = AsPavement	
		S	*****		E		採用 = RoadEquipment, AsPavement	
		S	*****		E		採用 = PedestrianPV	
		S	**		E		採用 = AsPavement	
		S	**		E		採用 = AsPavement, RoadEquipment, Maintenance	
		S	*****		E		採用 = BridgeSS, AsPavement, WaterProof	
		S	*****		E		採用 = Prime, AsPavement	
		S	*****		E		採用 = AsPavement, RoadBed, Earth, WaterWay, Wall	
		S	***	+	E		採用 = AsPavement	
		S	*****	+	E		採用 = Wall	
		S	*****	+	E		採用 = Wall	
		S	****	+	E		採用 = AsPavement	
		*****		+	E		採用 = BridgeBase, Wall, Foundation, Earth	
	*****S			+	E		Earth, Foundation	
	*****S			+	E		Foundation, BridgeBase, BridgeSS, Works, PedestrianPV	
*****		S		+	E		Wall	
*****		S		+	E		Earth, RoadBed, AsPavement, WaterWay, RoadEquipment	
***		S		+	E		AsPavement	

S:[ 開通日-18 ヶ月 ] E:[ 開通日+6 ヶ月 ]

図-2 予測に用いるデータの例

(5) 工事場所の特定結果

以上による工事場所の特定の結果を表-4にあわせて示した。場所が特定できないものは、「管内一円」や複数の都道府県に及ぶ広い範囲の工事で町・字が示されていない工事などであった。

(6) 工事期間の条件を加えたデータの作成

予測計算の対象データは、場所毎の集約に加えて、工事の開始と完了という工事期間の条件を同一にすべきと考えた。データの作成の具体的な方法を以下に示す。

a) 教師 (訓練) データ

教師あり機械学習の教師 (訓練) データとして、場所毎の道路の供用の有無が付加されたデータを作成する必要がある。

「道路供用有り」の教師データを作成するため、道路の供用開始があった場所 (供用場所) において、供用日以前に工事が開始され (複数年に及ぶ工事があるため、供用日以前に開始された工事であればその長さは問わず良いものとした)、供用日のNヶ月前 (遡り調査期間) から供用日の6ヶ月後の間に工事完了日を迎えた工事全ての工法・型式を説明変数に変換して同種の説明変数の重複を除いて集約した。Nは実データによる検証において12から30まで6ヶ月を単位として4種類の値を使用した。

「道路供用無し」の教師データは、道路の供用開

始が無い場所 (非供用場所) において、上記と同じ考え方で供用日を基準日に置き換え、基準日前に工事が開始され、基準日のNヶ月前から基準日後6ヶ月の間に工事完了日を迎えたものを採用した。

b) 予測対象 (テスト) データ

道路の供用の情報が無い予測対象場所については、基準日に置き換えて予測日を設定し、上記と同様な条件で選定した工事の工法・型式から説明変数を集約しテストデータを作成する。

(7) 予測に用いるデータの例

図-2は、供用された場所の2006年度から2010年度までに完了した工事を完了日の順に並べ、工事期間をアスタリスクで、その右側に工法・型式から変換した説明変数を示し、遡り調査期間 (N) を18ヶ月としたときに採用されるものを表したものである。採用された説明変数の集計に当該場所の供用の有無を加えたものを教師 (訓練) データとし、供用の有無が付加されていないものをテストデータとする。

7. 具体的な予測計算手法

教師あり機械学習には、決定木分析、ナイーブベイズ分類器、サポートベクターマシン、アンサンブル学習などがあり、いずれの手法によっても予測計

表-5 検証に使用した工事数と地点数

得られた 工事数と場所(町域)数		遡り調査期間			
		12ヶ月	18ヶ月	24ヶ月	30ヶ月
非開通場所	採用工事数	947	1381	1669	1806
	場所数	559	689	778	813
	基準日 2008.9.30	場所当りの 採用工事数	1.69	2.00	2.15
非開通場所	採用工事数	1092	1491	1798	2232
	場所数	628	754	844	932
	基準日 2009.9.30	場所当りの 採用工事数	1.74	1.98	2.13
非開通場所	採用工事数	1236	1645	1913	2312
	場所数	714	835	907	1007
	基準日 2010.9.30	場所当りの 採用工事数	1.73	1.97	2.11
開通場所	採用工事数	219	272	319	349
	場所数	35	35	35	35
	基準日	場所当りの 採用工事数	6.26	7.77	9.11

採用工事: 供用又は基準日以前に工事が開始され、対象期間内に工事の一部又は全部が完了し、説明変数として有効な工法・型式を有する工事

算は可能である。今まで、工事発注見通し情報の案件名称等を用いて道路更新情報を抽出する試みはあったが<sup>3)</sup>、工法・型式から作成したデータと道路の供用との関係を調査した研究は無い。この関係を分析できるよう、析出ルールを明示的に出力できる決定木分析を使用した。

(1) 決定木分析

決定木分析では、供用場所、非供用場所の混在したデータに対して、説明変数を一つずつ適用して、説明変数の有無に従って分岐を行うことを繰り返して供用場所と非供用場所を分類していく。分岐に用いる説明変数の適用順序は、当該目的変数を判定に用いることによって減ぜられる不確かさの大きさ(多様性指標)によって決定される。後述する検証で用いたC4.5アルゴリズムの多様性指標は「情報利得比」であり、これは、分割によるエントロピーの平均的な減少量である「情報利得」を「分割情報量」で除したものである<sup>13)</sup>。

(2) 不均衡データ問題

工事場所の中で供用場所は非供用場所に比べて少ない。このため、供用場所に関する情報利得比は小さなものになり、予測精度が低下する(不均衡データ問題)<sup>21)</sup>。そこで本研究においては、供用場所数に対する非供用場所数の比の直近上位の整数倍によって供用場所数のデータを非供用場所数と同等程度に複製する補正(オーバーサンプリング)を行っ

た。

8. 埼玉県道路開通を対象とした検証

これまで述べた方法によって、埼玉県が建設した道路における道路の供用の予測を試みた。埼玉県では道路の開通情報をインターネットで公開しており、これを教師データとしたため、供用の中でも開通のみを対象としていることとなる。

(1) 使用したデータと予測手法

a) 工事实績情報

2006年度(2006年4月1日)から2010年度(2010年3月31日)の間に工事が完了した埼玉県発注の埼玉県内の道路に関する工事实績情報を4,475件収集した。道路に関する工事は、予算費目によって事業分野を示す項目「工事の分野」を用いて抽出した。

埼玉県では土木工事共通仕様書において工事金額500万円以上の工事の工事实績情報への登録を求めているため<sup>22)</sup>、道路建設に関する工事は網羅されていると考えられる。

b) 道路供用(開通)情報

埼玉県がインターネットで「道路開通情報」として公表している2007年度~2010年度の道路開通を16件収集し、その開通区間から開通場所として35の町域を得た<sup>23), 24), 25), 26)</sup>。

c) データの作成

前述の方法によって作成したデータ数を表-5に示す。非開通場所の基準日は、2008年9月30日、2009年9月30日、2010年9月30日の3種類、遡り調査期間は、12ヶ月前、18ヶ月前、24ヶ月前、30ヶ月前の4種類とした。これにより、12種類の予測計算用データセットを作成した。遡り調査期間を長くすると採用工事数、場所(町域)数、場所当りの採用工事数が増加している。また、開通場所の場所当り採用工事数は非開通場所に比べて約4倍となっていることが表-5から読み取れる。

d) 機械学習手法

データマイニングツールWEKAのC4.5アルゴリズムによる決定木「weka.classifiers.trees.J48」によって、10分割交差検証(全データをランダムに10分割し、1つをテスト事例、残る9個を訓練事例とし、10回検証)を行った。



表-6 開通場所の予測計算の評価

評価指標	遡り調査期間(基準日:2010.9.30、2009.9.30、2008.9.30の平均)			
	12ヶ月	18ヶ月	24ヶ月	30ヶ月
適合率	0.257	0.196	0.253	0.251
再現率	0.928	0.965	0.948	0.965
F値	0.402	0.325	0.399	0.399
予測された開通町域数	126	173	131	135
(参考) 決定木の大きさ(分岐数)	5.7	8.7	8.3	9.3
(参考) 不均衡への補正を実施しない場合				
適合率	0.194	0.056	0.370	0.119
再現率	0.048	0.010	0.095	0.048
F値	0.077	0.016	0.151	0.068

## 9. 検証結果と考察

### (1) 予測の性能評価について

適合率、再現率、F値を用いて予測の性能を評価した結果を表-6に示す。これは、12種類のデータセットの10-分割交差検証結果に対して、4種類の遡り調査期間の評価結果を比較するため、遡り調査期間毎に3種類の基準日の評価値の平均をとったものである。

再現率は、全開通場所数35のうち予測で得られた場所数の割合である。例えば、予測によって35のうち33が予測された場合の再現率は、 $33 / 35 = 0.943$ となる。このため、遡り調査期間12ヶ月における再現率0.928から、予測で得られた場所のうち実際に開通であった場所数を算出する場合は、 $35 \times 0.928 = 32.5$ となる。適合率は、開通と予測された場所のうち実際に開通であった場所(正解)の割合である。例えば、予測によって開通と判定された場所が125ヶ所であって、そのうち32ヶ所が正しい場合の適合率は $32 / 125 = 0.256$ となる。このように、遡り調査期間12ヶ月の予測で得られた開通場所において、再現率が0.928であったことから、そのうち実際の開通場所は32.5であったことがわかり、さらに適合率が0.257であったことから、予測で開通と判定された場所数は、 $32.5 / 0.257 = 126$ であったことがわかる。多くの場所を開通と判定することにより再現率は上昇するが、適合率は低下する。した

がって、再現率と適合率を総合した指標として、再現率と適合率の調和平均であるF値が評価に用いられることがある。

公表されている開通町域数が35であるのに対して、表-6にあるように開通と予測された町域数は126~173であった。埼玉県内の町域数は、精査後の郵便番号データによると2,943であり、工事实績情報を用いた予測計算によって対象となる町域数を、県内の全町域数の4%から6%程度に絞り込んだこととなる。そして、再現率により、全開通場所の約95%を正しく予測していることが分かる。

表-6の上部は、開通場所のデータ数と非開通場所のデータ数の不均衡をオーバーサンプリングによって補正して分析を行い、その結果をオーバーサンプリング前の実際のデータ数に換算したものである。これに対して、不均衡への補正を実施しなかった場合の評価を表の下部に参考として示した。補正を行わない場合は、開通場所のデータ数が非開通場所に比べて少ないため、開通を特徴づける説明変数の情報量得比が補正時に比べて低く、その結果、開通に対する予測計算の評価指標の値が低い。

表-6に、参考に、決定木の大きさを分岐数によって示した。今回は予測精度の向上よりも析出されるルールの分かりやすさを優先したため、決定木の大きさを控えめなものとしている。今後、予測精度を重視する手法の採用や設定の改良によって、性能をさらに向上できる可能性があると考えている。

### (2) 予測できなかった開通場所について

開通場所を予測できなかった原因には次のことが考えられる。

#### a) 工事に起因する原因

- ・ 工事現場の条件により、他の開通場所で一般的に行われる工法・型式が無かった
- ・ 開通時期より相当以前に工事が行われていた

#### b) 登録に起因するもの

- ・ 施工場所あるいは工法・型式が正しく登録されていなかった

### (3) 開通と予測した非開通場所について

非開通場所を開通と予測した原因には次のことが考えられる。

- ・ 道路の開通に匹敵する大規模な工事(大規模な拡幅工事や災害復旧工事など)が行われた。
  - ・ 開通予定であるが未公表の場所を予測した
- ただし、開通以外でも大規模な道路改良(拡幅等)は道路地図の更新に重要であることから、実用性の観点から誤りとは言えない可能性がある。

表-7 説明変数の出現回数

説明変数	12種類のデータセット全体	12種類の分析結果		
		決定木中	第一分岐	第二分岐
AsPavement	16225	4		2
RoadEquipment	13308	12	6	6
Earth	9762	12		6
RoadBed	8687	11	6	4
Prime	8280	8		
WaterWay	7593	3		
Foundation	7061	6		
PedestrianPV	5472	2		
Wall	5464	3		
Removal	5157	9		
BridgeSS	4312	8		3
Maintenance	3520	2		
WaterProof	2083	7		
Works	2019			
BridgeBase	1762			
Slope	1600			
PC	147			
CoPavement	95			
CCBox	66			
SmallBridge	28			
Tunnel	4			

```

=RoadEquipment = yes
| ② RoadBed = yes: Open = 24, Norm = 50
| ② RoadBed = no
|
| ③ WaterProof = yes: Open = 2, Norm = 5
| ③ WaterProof = no
| ① | ④ Earth = yes: Open = 5, Norm = 41
| | ④ Earth = no
| |
| | ⑤ Prime = yes: Open = 2, Norm = 13
| | ⑤ Prime = no: Norm = 123, Open = 1
=RoadEquipment = no: Norm = 396, Open = 1
    
```

遡り調査期間; 12ヶ月 基準日; 2009年9月30日  
図-3 決定木の出力結果(例1)

```

=RoadBed = yes
| =RoadEquipment = yes: Open = 29, Norm = 79
| =RoadEquipment = no: Norm = 51
=RoadBed = no
| =Earth = yes
| | =Removal = yes: Norm = 55
| | =Removal = no
| | | =BridgeSS = yes: Open = 2, Norm = 2
| | | =BridgeSS = no
| | | | =RoadEquipment = yes: Open = 3, Norm = 25
| | | | =RoadEquipment = no: Norm = 36
| =Earth = no: Norm = 659, Open = 1
    
```

遡り調査期間; 24ヶ月 基準日; 2010年9月30日  
図-4 決定木の出力結果(例2)

(4) 析出された予測ルールと推定の可能性について

分析に使用した3種類の基準日、4種類の遡り調査期間による12種類のデータセットには、合計で延べ19,085の場所で行われた工事(次元圧縮の結果、有効な説明変数を持たない工事を除き、開通・非開通間のデータ数をオーバーサンプリングによって補正したもの)の説明変数が含まれている。表-7に、このデータセット全体に含まれていた説明変数の出現回数、析出された12種類の決定木に含まれていた説明変数の出現回数、その第一分岐、第二分岐に含まれていた説明変数の出現回数を示した。

図-3、図-4に得られた12種類の決定木から2例を示した。図-3中の丸付き数字は分岐数を表している。左上から順に説明変数の有無(yes/noで示される)とその条件下での開通・非開通(Open/Normで示される)の場所数を示している。下線のOpen/Normは、決定木がそれぞれ開通場所(Open)あるいは非開通場所(Norm)と判定したことを示す。開通場所数が非開通場所数より少ないにも係らず開通と判定されている結果がある。これは、オーバーサンプリングによって開通データを増やして分析を行い、その結果をオーバーサンプリング前の実際のデータ数に換算したことによって生じている。

これらは、埼玉県の評価対象期間中の道路関連工事における維持補修工事、橋梁工事、トンネル工事等の割合と開通との関連、開通日と採用された工法・型式の時期の差を反映したものとなっている。今回の検証の決定木において分岐に対する評価が高かった(第一分岐に採用された)説明変数は、遡り調査期間を18ヶ月以下とした場合は「RoadEquipment(道路付属物等)」であり、24ヶ月以上の場合には「RoadBed(路盤路床、コンクリート床版)」であった。これは道路の建設の工程と整合している。

予測の再現率が95%であり、予測ルールが合理的であることから、工事实績情報から道路の供用の推定は可能と考えられる。

10. おわりに

本研究は、工事实績情報からの道路供用の推定可能性を調査するため、工事場所、工事開始日、工事完了日、工法・型式から生成したデータに対し、教師あり機械学習を適用した。埼玉県発注工事の道路開通による検証によって推定が可能と考えられる。今後、他の道路管理者、他の種類の道路供用に対して検証し、汎化の検討を進めたい。

参考文献

- 1) 総務省統計局：日本統計年鑑 第9章建設業 9-3 発注者及び目的別工事分類別公共機関からの受注工事件数及び請負契約額, <http://www.stat.go.jp/data/nenkan/zyouhyou/y0903000.xls>, 2012. 10. 13. アクセス.
- 2) 一般社団法人 日本道路建設業協会：統計資料 道路事業費と舗装費 10年間の推移(データ), [http://www.dohkenkyo.com/shiryog\\_pdf/costs/h02.pdf](http://www.dohkenkyo.com/shiryog_pdf/costs/h02.pdf), 2012. 10. 13. アクセス.
- 3) 関本義秀, 中條覚, 南佳孝, 山口章平, 山田晴利, 布施孝志：工事発注見通し情報を用いた全国における道路更新情報の自動抽出に向けた試み, 土木学会論文集 D3(土木計画学), Vol. 68, No. 3, pp.117-128, 2012.
- 4) 特定非営利活動法人 ITS Japan：安全・環境に資する走行支援サービス実現のための道路情報整備と流通へ向けた提言, <http://www.its-jp.org/wp-content/uploads/2010/09/2d241ee556cc6bcfb250a3130e642658.pdf>, 2008. 4, 2012. 10. 13. アクセス.
- 5) 一般財団法人 日本建設情報総合センター：新しいコリンズ・テクリスの概要, JACIC 情報 102, Vol. 26, No. 2, pp.25-26, 2011. 7.
- 6) 一般財団法人 日本建設情報総合センター：工事実績情報公開, <http://www.ct.jacic.or.jp/teikyoku/search.html>, 2012. 10. 13. アクセス.
- 7) 例えば, 国土交通省関東地方整備局：土木工事共通仕様書(平成23年度) 1-1-5 CORINSへの登録, [http://www.ktr.mlit.go.jp/ktr\\_content/content/000011773.pdf](http://www.ktr.mlit.go.jp/ktr_content/content/000011773.pdf), p.7, 2012. 10. 13. アクセス.
- 8) 国土交通省：建設工事施工統計調査 完成工事高等時系列表, <http://www.e-stat.go.jp/SG1/estat/Xlsdl.do?sinfid=000012903080>, 2012. 3, 2012. 10. 13. アクセス.
- 9) 国土交通省：調査の概要・用語の定義, [http://www.e-stat.go.jp/SG1/estat/GL08020103.do?\\_pdfDownload\\_&fileId=000005141210&releaseCount=1](http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_pdfDownload_&fileId=000005141210&releaseCount=1), 2012. 3., 2012. 10. 13. アクセス.
- 10) 一般財団法人 日本建設情報総合センター：コリンズ項目定義書 XML 定義書, <http://ct.jacic.or.jp/corporation/know/xml/file/c02x01.xls>, 2012. 10. 13. アクセス.
- 11) 一般財団法人 日本建設情報総合センター：コリンズ・テクリスマニュアル(コリンズ編), [http://ct.jacic.or.jp/corporation/know/manual/pdf/cor\\_manual\\_02.pdf](http://ct.jacic.or.jp/corporation/know/manual/pdf/cor_manual_02.pdf), 第2章, pp.2\_12-2\_68, 2012. 9. 5. アクセス.
- 12) 一般財団法人 日本建設情報総合センター：コリンズ選択項目詳細-工種・工法, 型式/担当工事内容, [http://ct.jacic.or.jp/corporation/know/xml/file/c02x01\\_16.xls](http://ct.jacic.or.jp/corporation/know/xml/file/c02x01_16.xls), 2012. 7. 5. アクセス.
- 13) 例えば, Ian H. Witten and Eibe Frank：DATA MINING Practical Machine Learning Tools and Techniques, ISBN 978-0120884070, 2nd ed, p.289, 103, ELSEVIER, 2005.
- 14) 石井一郎・元田良孝：道路工学, 第8章道路土工～第15章維持補修, 鹿島出版会, 1993. 1.
- 15) 国土交通省国土技術政策総合研究所：1. 2 体系階層(レベル)の定義, <http://www.nilim.go.jp/lab/pbg/theme/theme2/sekisan/tree/t1-2.pdf>, 2012. 10. 13. アクセス.
- 16) 国土交通省国土技術政策総合研究所：工事工種体系ツリー(平成23年度改訂版), <http://www.nilim.go.jp/lab/pbg/theme/theme2/sekisan/tree/t1-4.pdf>, 2012. 10. 13. アクセス.
- 17) 例えば, 国土交通省中部地方整備局：道路設計要領(設計編), <http://www.cbr.mlit.go.jp/road/sekkeiyouryou/index.htm>, 2012. 7. 23. アクセス.
- 18) 国土交通省国土政策局：位置参照情報ダウンロードサービス, <http://nlftp.mlit.go.jp/isj/index.html>, 2012. 8. 16. アクセス.
- 19) 日本郵便株式会社：郵便番号データダウンロード, <http://www.post.japanpost.jp/zipcode/download.html>, 2012. 8. 16. アクセス.
- 20) 南佳孝, 関本義秀, 中條覚, 柴崎亮介：路線情報を加味した道路関連情報の位置特定に関する研究, 土木学会論文集 F3(土木情報), Vol. 67, No. 1, pp.7-17, 2011.
- 21) Haibo He and Eduardo A. Garcia: Learning from Imbalanced Data, *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, Vol. 21, No. 9, pp. 1263-1284, 2009. 9.
- 22) 埼玉県：埼玉県土木工事共通仕様書, <http://www.pref.saitama.lg.jp/uploaded/attachment/437216.pdf>, 第1編第1章第1節, pp.25-26, 2012. 10. 7. アクセス.
- 23) 埼玉県道路街路課：道路開通情報, <http://www.pref.saitama.lg.jp/site/kaitసుzyouhou/>, 2012. 10. 7. アクセス.
- 24) 埼玉県道路政策課：道路開通情報(国道), <http://www.pref.saitama.lg.jp/site/dousei1006/905-20091224-51.html>, 2010. 3. 19. 更新, 2010. 8. 15. アクセス.
- 25) 埼玉県道路政策課：道路開通情報(県道), <http://www.pref.saitama.lg.jp/site/dousei1006/dousei035.html>, 2012. 2. 3. 更新, 2012. 8. 15. アクセス.
- 26) 埼玉県道路政策課：道路開通情報(その他), <http://www.pref.saitama.lg.jp/site/dousei1006/dousei059.html>, 2012. 1. 17. 更新, 2012. 8. 15. アクセス.

(2012. 10. 19 受付)



## PREDICTION FOR LOCATIONS OF CHANGES IN ROADS FROM CONSTRUCTION-RECORDS

Wataru KOBAYASHI, Ryosuke SHIBASAKI and Yoshihide SEKIMOTO

To keep road maps and map data in car navigation system current gives great benefits to citizens. However, a lot of time and effort is needed to gather information which follows roads change. To improve the situation, this paper presents construction-records which are collected to control the quality of public works all over Japan can predict for locations where roads change. First, construction-records within a certain period of time are grouped by location. Next, the construction method/ type information is converted into explanatory variables (independent variable). The variables aggregated by each location are calculated by a supervised machine learning. An experimental evaluation showed that the method outputted decision tree models which reflected process of road construction, and the average recall ratio was approximately 95%.