

# データマイニング機能を有する 道路維持管理業務支援システムの提案

寺田 守正・中西 正樹・金田 重郎

## あらまし

自治体における道路維持管理業務は、住民・職員からの「ガードレールが破損している」などの通報・報告に対して、迅速に対応してゆくことを目的としている。しかし、通報・報告を受けてから動き出す受動的対応では、迅速対応には限度があり、その業務効率の改善が望まれる。そこで、本論文では、予防保全型の管理業務形態の実現を目指し、データマイニング手法の適用を提案する。具体的には、案件データが持っている緯度経度情報と案件種別に基づくクラスタリングを採用した。提案手法を、著者らが開発・運用している道路維持管理支援システム「京都道守くん」に実装した結果、繰り返し発生している案件を抽出できるなど、一定の有効性を確認できた。しかし、本システムのような「本来は望ましくない案件」からのデータマイニングでは、案件データの属性として「原因」が記入されていないと、根本的な解決策の提示は困難である。また、案件の種別選択の曖昧性がデータマイニング上の障害となることも示された。種別選択のあいまい性を回避し、「原因」項目を確実に入力してもらうためには、今後は、ナレッジコンシェルジュの導入が必要と考えられる。

## 1 はじめに

自治体が管理している道路における、施設破損や不法投棄などに対応する道路維持管理業務（以下「維持管理業務」と言う）は、自治体の土木事務所（以下「土木事務所」と言う）が担当す

る住民サービスの中でも重要なものの1つである。一方、近年、地方自治体においても、住民満足度の向上と、業務の効率化が強く望まれている。道路に関する維持管理業務も例外ではなく、事務処理の迅速化と効率化が強く望まれている。著者らは、維持管理業務の効率化を目的として、道路維持管理業務支援システム「京都道守くん」（以下「京都道守くん」と言う）の開発・運用をこれまで行ってきた。そして、社会実験により、業務の効率化効果を確認している[1]。なお、類似の機能を持つ道路維持管理業務支援システムは、その後、愛知県などでも開発事例が報告されている[2][3]。

しかし、現状の維持管理業務では、近年の住民ニーズや価値観の多様化から、住民通報も多種多様となっている。一方、地方自治体の財政状況には厳しいものがあり、予算や人員上の制約などから、それらに十分に対応出来ていないとは言いがたい。その結果、(1)案件が発生して通報・報告がなされた後に対応する受動的対応、(2)案件への抜本的対応（一般的に時間を要する）でない対症療法的対応、などにならざるを得ない。これらは、案件の再発による件数増加をもたらす可能性があり、その結果、更に、受動的・対症療法的対応とならざるを得ない。

この現状を打破し、案件発生を事前に予測できれば、緊急度などに応じた補修計画立案が容易となる。このような予防的な保全を行うことで、計画的な予算執行や人員体制が可能となり、結果として、道路行政に対する住民満足度の更なる向上が可能となる。

そこで本論文では、予防保全を支援する維持管理業務の実現を目指し、「京都道守くん」に

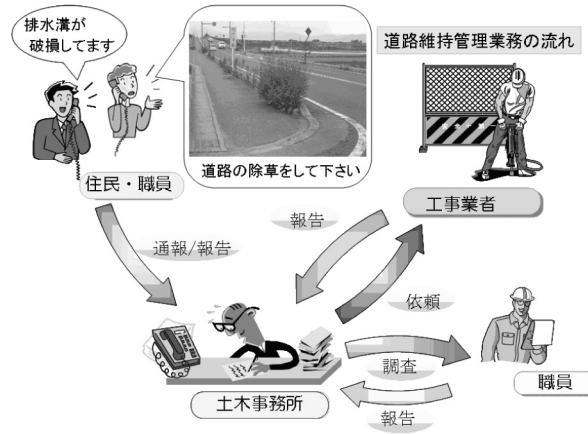


図 1：道路維持管理業務の流れ

蓄積された案件データから、緯度経度と案件種別に基づいて、繰り返し生じている類似案件を抽出するデータマイニング手法を提案する。A 市内における過去 5 年分、約 800 件の案件データに対して提案手法を適用した結果、案件が繰り返し発生している場所を抽出できた。その結果を受けて、データマイニング機能を持つ「近隣案件画面」を「京都道守くん」に導入した。社会実験を行った結果、予防保全的な対応を実現するには、案件データにおける「原因」の記述が重要であること、入力データの正確性を担保するために、ナレッジコンシェルジュの設置が望ましいことが導かれた。

以下、第 2 章で、著者らが開発して現場に導入している「京都道守くん」について紹介する。第 3 章では、データマイニングの可能性について論じ、第 4 章では、実際のマイニングアルゴリズムを示す。第 5 章では、提案されたマイニング手法を実際にシステムに組み込む。第 6 章では、社会実験の結果を受けて、ナレッジコンシェルジュの必要性について論じる。第 7 章は結びである。

## 2 道路維持管理支援システム「京都道守くん」

### 2.1 道路の維持管理業務

道路維持管理業務とは、ガードレールの破損、あるいは、道路上の放置物等、道路の正常な利

用に支障となる「案件」の情報を集約して、工事等の手配を行う業務である。業務内容のイメージを図 1 に示す。具体的には、まず、住民通報や職員報告という形で、道路利用上の障害についての何らかの第 1 報が土木事務所に寄せられる。たとえば、「道路に雑草が発生して、歩きにくい。」と言ったものである。通報・報告を受けた土木事務所では現地を調査する。その結果を受けて、対処方法を判断し、修繕工事などの対応が必要な場合には、工事の施工業者へ対応を依頼する。施工業者は、工事などの何らかの対応を遂行したのち、対応の完了を土木事務所に報告する。土木事務所が対応の完了を確認した後、施工業者へ、対応に対する代金支払いが行われる。

ここで、通報・報告された案件をどのように進捗管理するかが課題となる。著者らの土木事務所でも、以前には、紙の管理台帳で進捗管理していた。しかし、これでは、通報・報告された案件について、それが過去に通報・報告されている内容かどうか、誰が担当していたかも簡単には分からない。また、緊急案件に対して土木事務所以外で対応しようとしても、詳細を、逐一、土木事務所に電話して確認する必要がある。さらに、施工業者に案件の位置を知らせようとしても、地図をコピーしてファックス等で送ることは、著作権上の問題が生じるため、難しかった。このような状況の中で、いかに情報システムを活用して、道路維持管理業務を支援するかが、ひとつの課題となっていた。



図2:「京都道守くん」の基本的な役割

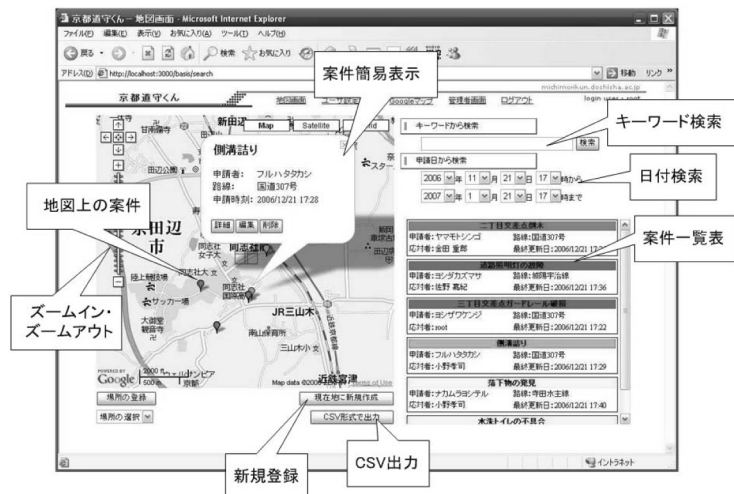


図3:「京都道守くん」のトップ画面

## 2.2 道路維持管理システム「京都道守くん」

以上の状況に鑑みて、著者らは、道路維持管理業務支援システムのプロトタイプを開発した[1]。その全体的な位置づけを図2に示す。本システムは、以下のような機能を持った、一種のマップシステムである。

- Google Map APIをマップシステムとして活用し、それぞれの通報案件を地図上に一括表示
- 住民・職員からの通報・報告案件対処の進捗状況管理
- 幹部報告書類の自動生成
- 工事担当者への連絡用E-mailの作成支援(合わせて、Google Map上の通報・報告案件の位置を緯度経度座標で通知)

図3は、「京都道守くん」のトップ画面を示す。Google Map上に案件を登録できるため、案件の情報を視覚的に管理可能となり、全体像の把握が容易となる。Google MAP上に配置さ

れた案件は、画面右の一覧にも簡易詳細とともに表示される。Google MAP上のマーカー、または一覧をクリックすると、案件に対する「詳細表示」、「編集」、「削除」が可能である。案件を入力項目のひとつである緊急度の値によって色分けしている。更に、本システムでは、キーワードや日付を用いた案件の検索機能も備えており、GIS上に表示する案件を絞り込むことができる。この機能を生かして、特定の種類の案件のみを絞り込み、その種類の案件の特性を調べることもできる。

## 3 維持管理業務形態へのデータマイニングの適用

### 3.1 案件データの蓄積

著者らは「京都道守くん」を現場に導入して、

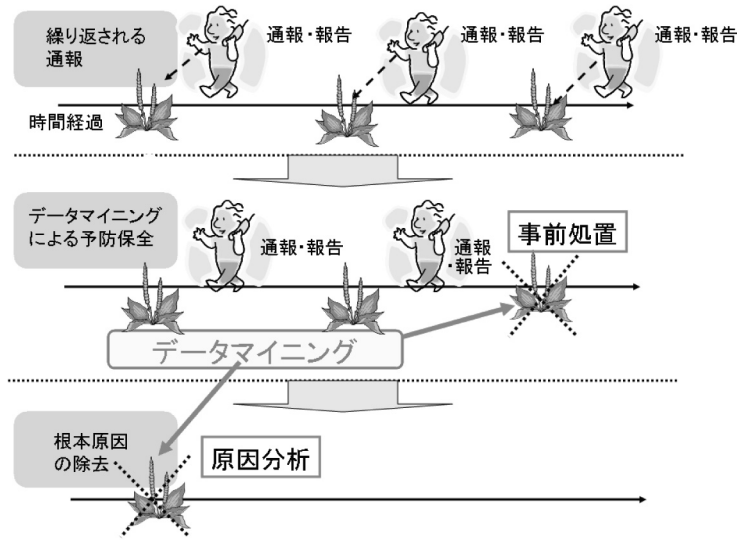


図4：データマイニングの適用イメージ

実証実験を行ってきた[1]。その結果、案件に対して従来よりもより迅速に対応できることを確認している。一方、「京都道守くん」については、これまでの長期の利用によって、案件に関するデータが蓄積されつつある。これらの蓄積データに対してデータマイニング手法を適用することで、業務全体をより効果的・効率的に行うための知見を得ることができれば、従来の対症療法的な案件対応を改善し、計画的な案件対応が実現する可能性がある。そこで、本論文では、そのような予防保全的機能を有する「京都道守くん」にその機能を強化することを試みた[4][5]。

### 3.2 データマイニングの適用

実際に想定される、案件データへのデータマイニング適用のイメージを図4に示す。図4上段のように、毎年、同一案件（例えば、草が伸びる）が繰り返されているとする。これに対して、同図4中段のように、同一箇所で同一種別の案件が繰り返されていることを、過去の案件データからのデータマイニングで発見できれば（例えば、伸びる前に草を刈ってしまうなどの）事前対応を行って、通報・報告に至る前に案件に対応できる。結果的に、住民の満足度も向上する。更に一步進んで、図4下段のように、繰り返

返される案件の根本的な原因が、データマイニング結果から分かれば（例えば、雑草が伸びないようにコンクリートで舗装したり、別の植物を植えるなどの）根本的な対策が可能となるはずである。

更に、図4中段のように、計画的に工事などの事前処置が実施されると、例えば、ある地域に同様の工事がある場合には、同一の施工業者に一日で済ませてもらうように依頼して、工事コストを下げることが可能となる。施工業者からすれば、工事担当者のローテーションに余裕が生まれるなどのメリットも期待できる。

ただし、図4からは、道路の維持管理業務へのデータマイニング適用の課題もいくつか見えてくる。第一に、例えば同一内容で10回も住民からの通報・報告を受けてからやっと事前の対応ができるようでは困る。データマイニングは、極めて少ない個数の、同一種別の案件データから、確実に案件発生の繰り返しを導き出す必要がある。抽出される類似案件に漏れがあってはならない。第二に、データマイニング結果から、その繰り返されている案件の根本的な原因と対応策が導きだされる必要がある。後者については、実際にデータマイニングを行った結果から考察を加える（後述）。いずれにせよ、本論文で必要とされるデータマイニングは、「同じ場所で同じような種別の案件が繰り返されている」現象を取り出すことと定義する。データマイニ

ング手法としては、緯度経度のロケーション情報を用いたクラスタリング処理となる。

また、上記において、同一種別の案件であることを如何にして判別するかが課題となる。ひとつの方法は、案件の種別を事前に職員が入力することである。著者らの道路維持管理業務の支援システム「京都道守くん」では、開発当初からこれを見越して、案件の種別を職員が入力するように設計されていた。しかし、後述するように、この案件種別の入力値には曖昧性があり、それがデータマイニング上で問題となった。また、同一種別の案件を取り出す、他のもう一つの方法は、自然言語処理技術によって、案件内容の記述から、共通的な単語を取り出すことである。本論文のクラスタリングでは、この2つのアプローチを共に採用している。以下、詳細に述べる。

## 4 緯度経度に基づくデータマイニング手法

「京都道守くん」にデータマイニング機能を組み込むため、まず、過去に対処した案件データからのデータマイニング適用を試みた。まず、最初にデータマイニング・アルゴリズムを示し、更に、実験結果について述べる。

### 4.1 データマイニング・アルゴリズム

#### STEP1：形態素解析

維持管理データのタイトルや案件内容は自然言語で書かれているため、これを形態素解析し、各案件に記録された単語を求める。また、形態素解析には「茶筌」[6]を用いることにする。形態素解析の対象は、人手で記録された案件のタイトル、処置欄（テキスト）である。発生箇所住所は誤分類の原因となるため除外する。

#### STEP2：単語発生頻度計算

形態素解析によって得られた単語の発生頻度をカウントし、中程度の発生頻度の単語を取り出す。頻度の多すぎる単語は一般単語に過ぎず、一方、少な

すぎる単語は特殊と見なす。ただし、以下の評価実験では、案件数が約800件と少ないため、発生頻度が多い単語も評価対象に含めている。

#### STEP3：緯度経度によるクラスタリング

上記のそれぞれの単語を持つ維持管理データに対して、緯度経度を用いてクラスタリングする。今回は、3件以上の案件が含まれたクラスタのみを取り出している。最終的には、Web-GIS上で案件を表示し、目視で内容確認した。

なお、上記のSTEP2において、形態素解析によって得られた中程度の頻度の単語（名詞）をクラスタリングの条件に選んでいる。しかし、これは、本来であれば、「案件の種別」などを用いるべきである。類似案件は、まず、「種別」が同一ということで見れると思われるからである。しかし、上記アルゴリズムでは、「種別」は用いていない。これは、当初、「種別」によるクラスタリングを試みたが、本来は同一の種別の案件と思われるものが、異なる種別として入力されていることが多発して、良好なクラスタリング結果が得られなかったためである。種別がきちんと入力されていないことについては、後ほど、分析する。

### 4.2 クラスタリング実験結果

前節のアルゴリズムをA市の過去における約800件の案件データに対して適用した。その結果、一定の周期で特定の箇所において類似案件が発生しているクラスタを確認できた。マイニング結果を表1に示す。表1に示された案件中には、ほぼ毎年のように、特定の箇所に全く同一の案件が発生しているクラスタが多数存在する。図5はその一例である。ある観光地の橋の近くで、害虫に関する案件が繰り返して発生していることが、マイニングの結果、示されている。本提案手法のマイニング・アルゴリズムによって、過去の案件データから、同じ現象が毎年のように発生しているケースを抽出できることが確認された。

しかし、確かに、繰り返して発生していることは検出できたが、この害虫に関するクラスタ図5のみからでは、「なぜ、害虫が発生したか」

表 1：緯度・経度によるクラスタリング結果の一部

抽出単語	抽出単語を含む案件数 (件)	クラスタの数 (個)
がたつき	57	7
閉塞	37	5
落下	25	1
雑草	22	3
損傷	19	2
隙間	14	1
予定	13	3
毎年	12	2
午前	11	2
防止	9	1
...	...	...

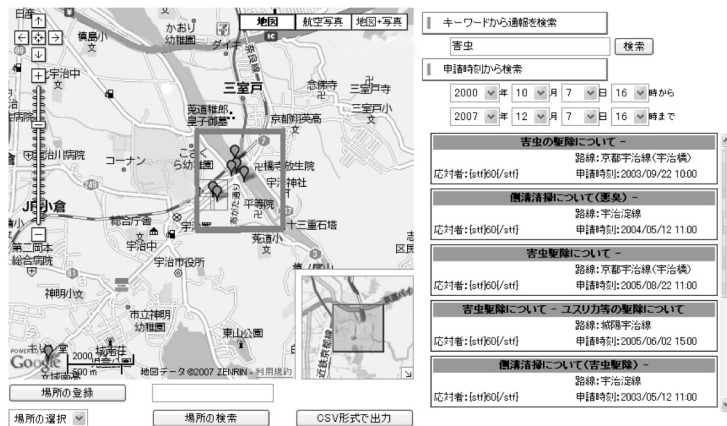


図 5：「害虫」の頻発箇所の特定

が（少なくとも、実際に案件対応を体験していない者には）推測することができなかった。このため、上記の案件群における情報を、実際に案件への対応を実行した職員に確認した。観光地をイメージアップするために植えていた街路樹を、害虫が好んでいたことが分かり、街路樹を取り除く処置を行っていた。

このような現象は、他のクラスタリング結果でも頻繁した。上記の事例は、深刻な問題を提起する。一般的に言って、過去の案件データを次なる業務に生かすためには、以下の2つの知識が必要である。

1) クラスタの存在検出：特定の場所に類似の案件が頻発していることの検知。例え

ば、毎年のように、害虫駆除の依頼がある。

2) 原因に関する知識：その案件を生じさせている「原因」が何であるかを理解できるだけの知識。例えば、害虫の好む樹が街路樹として植えてあること。

1) は予防保全を行うために、「どこに保全しなければならない案件が生じるのか」という保全対象の場所を特定するための知識である。これは、今回提案のデータマイニングによって抽出できることを確認できた。2) は、1) から得た知識を基に「どのような対処法を行ったらよいのか」を考えるためにどうしても必要な知識である。

後者の知識は、データマイニングからは導出できない。データマイニング[7][8]は機械学習の一種(教師無し学習)である。機械学習では、案件データが持っている属性の中から、一部の属性のみを取りだして、ターゲット概念を表現する。即ち、「案件データが持たない全く新規の属性を使って初めて表現可能となる概念」は獲得できない<sup>1</sup>。その意味でも、「原因」が書かれていないと、「なぜそれが生じたか」をマイニング結果から読み取ることは本質的に困難である。

したがって、案件データの記録を見て、読み取るしか方法がない。しかし、今回のデータマイニングの実験では、この「原因」がほとんど記述されて居なかった。この状況では、当該案件と関連した過去の案件に対応をしていない職員は、1)の知識を得ても、2)の知識を抽出することはできない。職員は数年で異動するのが通常である。また、土木事務所には複数人の職員がいる。当該案件の連絡を受けた職員が、過去に発生していた同一箇所の案件に携わっていなかったということは、組織の中では起こりうることである。

維持管理業務では、2章でも述べたが、通報・報告を受けた個々の案件に対応するため、一連の処置をこなすことが主な職務である。つまり、「通報・報告受信・記録」→「現場調査・報告書作成」→「施工業者などへの対応依頼」までの一連の処置を迅速に行うことは要求されるが、案件対応時に「原因」を記述することは必須事項ではない。結果として、「原因」情報をドキュメント化してこなかったと考えられる。

## 4.3 実験結果の分析

### 4.3.1 「望ましくない状態の知識」の問題

上記の「原因」情報が必要となる背景として、

<sup>1</sup> 上記のように「機械学習」は、それまで思いもつかなかった全く新しい知見を帰納的に獲得する能力は持たない。機械学習における「汎化能力」とは、(事例を記述している全属性中の)一部属性のみで構成される条件(ルール)が得られるので、汎用性があり、あらたに生じるであろう事例に対して、何らかの判断を実行できるとの意味に過ぎない。

<sup>2</sup> 例えば、「紙おむつを買う男性はビールを買う」という条件からは、母親に代わって、紙おむつを買いに来た、若い父親の姿が思い浮かぶ。

<sup>3</sup> トヨタの看板方式を提案した大野耐一[11]は、現場でトラブルが起こった際に、5段階まで、深く「原因」を考察することを求めた。この場合も、工場で生じた「望ましくないデータのクラスタ」を解決するためには、表面的な発生条件のみを追究しても意味がなく、一歩進んで、「原因」を追究する必要があることを示唆している。本論文のマイニングでも状況は同じであり、単に案件が集まっている条件(属性から構成される式)を見出したのみでは、本当の意味での問題解決は難しい。

本システムが対象とする業務そのもの持っている特性があるように思われる。データマイニングの適用先として、よく知られたものに、ショッピングバスケット問題がある。ショッピングバスケット問題では、「望ましいデータのクラスタ」が抽出される。したがって、得られたターゲット概念(汎化結果)は、そのまま「一般的に使える知識」である。「商品をたくさん買うお客様の条件」が見つければ、それをそのまま適用して、顧客目簿から顧客を選択し、そこにダイレクトメールを送ればよい。すくなくとも、「たくさん買うお客様の条件」が分かれば、どんな年齢層でどんな職業かなど、「買ってくださる原因」を推論することは、必須ではない。

これに対して、本マイニングでは、「望ましくないデータのクラスタ」が抽出される。望ましくないことが起こる条件は、そのままでは「使えない知識」である。「商品を買わないお客様の条件」を得られても、そのような条件を満たすお客様にダイレクトメールを送送することに意味はない。観光地のある橋で、害虫駆除の要請が頻発しているという「条件」を聞かされても、それから、何が起こっているかは簡単には想起できない。「それはなぜだろう」を考えて、原因を探り、対応をしない限り、「役に立つ知識の獲得」とはならない。その「なぜ」を探る手がかりが、過去の案件には書かれていないことが多かったのである<sup>3</sup>。

### 4.3.2 「案件の種別」の曖昧性の問題

上述の案件の種別に「ばらつき」がある問題も重要である。本システムでは、できるだけ少ない事例からマイニングを実現することが要求される。数個の過去の案件データから関連性を抽出する。「種別」が曖昧であると、本当に必要な近隣案件を見逃す恐れがある。

この曖昧性について、担当職員の意識や勤務

態度に原因を求めることには無理があると思われる。

パタン認識分野では良く知られた定理に、「醜いアヒルの子の定理」がある。この定理は「もの」と「もの」が似ているかどうか、すなわち同一のカテゴリーに属するかどうかは、一般的には規定できないとするものである[9]。このことから、類似性を規定する何らかの「価値観」が与えられない限り、「もの」と「もの」との判定ができない。案件種別の選択は、無意識の内に何らかの「価値観」に基づいて行われるため、職員は必ずしも的確に種別選択を行えない。つまり、ひとつの価値観の下に種別選択を行うのは、それぞれが異なる社会経験と価値観を有する複数人の職員がいる場合には、本質的に困難と思われるからである。

## 5 「京都道守くん」におけるデータマイニング機能

### 5.1 新規追加機能

以上の検討結果を受けて、「京都道守くん」に、データマイニング機能を追加した。ひとつは、(これまでは入力を要求して来なかった)「原因」をデータ属性としての追加である。図6には、追加された「原因」欄を示す。「なぜ、その案件が生じたのか」という文脈化した原因情報をドキュメントとして組織内で共有することができれば、過去に当該案件を経験していない職員でも、根本的な原因を探り当てる一助になると考えたからである。また、案件データの入力インターフェースの「原因」欄新設以外でも、システムからPDF出力される報告書に「原因項目欄」を設けた。この報告書とは、上司に対して施工業者の依頼などの許可をもらうための申請書である。上司への報告書に「原因項目欄」を設けることにより、上司が「原因」の入力が必要だと感じた案件であるにも関わらず、原因が入力されていない場合は、職員に対して入力を行うように指示することが可能である。

更に、新たに案件を受け付けるごとに、当該

新規追加案件の近隣で発生した過去の類似案件を表示する「近隣案件画面」(図7)を設けた。データマイニング結果を、日常の業務の中に生かそうとするものである。これによって、新規の案件が生じた際に、職員が過去の案件データをより有効に利用できることを期待した。このデータマイニングでは、クラスタリングには、緯度経度とともに、「案件の種別」を用いた。図8は、「京都道守くん」における、案件種別の入力フィールドである。実際の案件種別は2段階構成になっている。

第1段階の「対象物<sup>4</sup>」を設定し、対象物で起こりうる第2段階の「状態<sup>5</sup>」が設定される。この「近隣案件画面」では、種別の第1段階のみが一致した場合には「類似度：低」とし、2段階とも一致した場合は「類似度：高」として表示している(図7)。

この「近隣案件画面」は、通報・報告を受けてその内容を保存した際に、自動的に表示される。類似した案件が、近隣で頻発していることを、職員に気付かせることを狙いとしている。なお、「近隣」は、半径500メートル以内と設定した。過去に実施したシステム評価において、類似原因で頻発する案件の発生範囲が500メートル以内であることが多いことを確認していたためである<sup>6</sup>。

### 5.2 実運用による評価

提案システムを2009年2月初旬～同年4月下

図6：「原因」属性の入力欄

<sup>4</sup> 例えば、「トンネル」、「道路の側溝」、「路面、路肩」など。

<sup>5</sup> 案件の特徴のこと。例えば、「雑草」、「側溝つまり」など。

<sup>6</sup> 例えば、「側溝の詰まり」に関する案件では、同様の案件が頻発するのには、ある程度の距離範囲で発生する可能性がある。





図7：「近隣案件画面」の例



図8：案件の種別選択の例

旬までの約3ヵ月、運用を行なった。評価のための稼働時間は十分ではないが、近隣案件は3～4件に対して1回程度の出現頻度で表示されている。ただし、案件の種別の第1段階のみが一致した、類似度が低い過去案件が大半であった。例えば、図9の例では、新規入力した「側溝蓋の危険箇所」案件に類似した過去案件として、排水処理に関する案件が表示されている。種別の第1段階である「排水施設」が一致したことで、表示されたと考えられる。

現状では、蓄積された案件が少ないため、「近隣案件画面」で表示される案件数は限定されている。結果的に、当該過去の案件を担当していない職員には一定の参考になることが確認されている。しかし、この状況では、案件データが蓄積されるにつれて、表示される案件が増加する。案件が大量に表示されてしまうと、その中に新規案件に参考となる案件が（本当は）含まれていたとしても、多数の案件をチェックしないと見つからなくなる。このような状況になると、この画面を利用しない習慣が組織内で浸透する恐れがある。この状況を回避するためには、案件候補の絞り込みが必要である。上記の運用結果を見ても分かるように、案件の種別だ

けに頼ったマイニングではなく、より多くの属性を用いて、適確な過去案件を抽出する必要がある。

また、「原因」欄については、どの案件にも原因が記入されているとはならなかった。「原因」欄が必ずしも記入されないのは、職員が怠慢なのではなく、止むを得ない側面がある。どんな業務システムであっても、システムを利用する職員は、業務に直接関係のないデータ項目に関しては、入力インセンティブを保ち難い。言い換えると、当面の案件対応には必然性が感じられない「原因」欄の入力を職員に強制することは、余分な稼働を現場の職員に課する恐れがある。また、このような状況では、たとえ、入力されても、十分な内容を期待できない。

以上、案件データにおける緯度経度情報と案件種別を用いたクラスタリング手法の、道路の維持管理業務への適用可能性を検討した。この種のデータマイニングを効果的なものとするには、以下の条件が満たされる必要がある。

- ・システムの運用を続けると近隣案件の個数が増大し、近隣案件を業務支援に生かしくくなる。更なる近隣案件の絞り込みを実



図9：側溝蓋の傷みに関する近隣案件

現するため、現状の「案件種別」のみではなく、他の属性を、クラスタリングアルゴリズムに追加すべき。

- 本システムのデータマイニングは、「望ましくない案件の『塊』」を出力する。その結果、そのままでは社会に適用できない。根本的な問題解決を実現するには、案件データに「原因」が書かれていなければならない。
- 以上の「案件種別」「原因」そして、新規に追加する属性は、各案件データ入力時に、正確に入力されていなければクラスタリング結果に影響を与える。

## 6 あるべきシステムの姿

### 6.1 漸増的な属性追加の必要性

前述の3つの要件の中で、まず、データマイニングへの新しい属性の追加について考察する。データマイニングが行われると、何らかの知識を得る。しかし、機械学習で得られる「知識」は、事例の統計的な偏りを、すでに、存在している属性値の塊として表現するのみであり、そ

の事実を生じさせた「文脈」や「原因」を得ることはできない。

新たな知見を得るには、種々の属性を事前にシステムに取り込んでおかなければならない。上記のような案件発生の「原因」を本来、システムには記録すべきであったというようなことも、その一例である。しかし、何が知見として得られるかは、事前にはわからない。このように考えると、案件として記録しておくべき属性がどんどん増えて收拾がつかなくなる。ここで、人工知能における「フレーム問題」を想起しておく必要がある。フレーム問題とは、1969年に、マッカーシー[10]が提示した人工知能の最大の難問とされるものである<sup>7</sup>。あらゆるケースを想定して、あらかじめ将来必要となる「属性」を案件データに埋め込んでおくことはできない。これから生じるどのような文脈的状况にも対応可能するためには、対象世界に対するあらゆる属性情報をあらかじめ準備する必要がある。実際には、それは不可能である。あらかじめ情報システムが扱う状況を限定して、有限の空間の中で判断するしかない。たとえば、最初は施工業者名はデータには無かったとする。しかし、ある数社に工事後のトラブルが集中的に発生するような事態が仮に生じていれば、「施工業者名」も、マイニングの対象とすべできある。この場合、「施工業者名が多くの場合、トラブル

<sup>7</sup> 「フレーム問題」[10]は、人工知能における重要な難問の一つで、有限の情報処理能力しかないロボット（本章では、情報システム）には、現実起こりうる問題すべてに対処することができない。人工知能では、何らかの問題解決を自動的に実施したいわけであるが、その作用が現実社会に関わっている場合、「何が関係あって、何が関係ないか」を調べるために、無限の計算が必要になって人工知能が止まってしまうことをいう。つまり、どんな問題でも対応できる人工知能、すなわち、コンピュータシステムは本質的にできない。

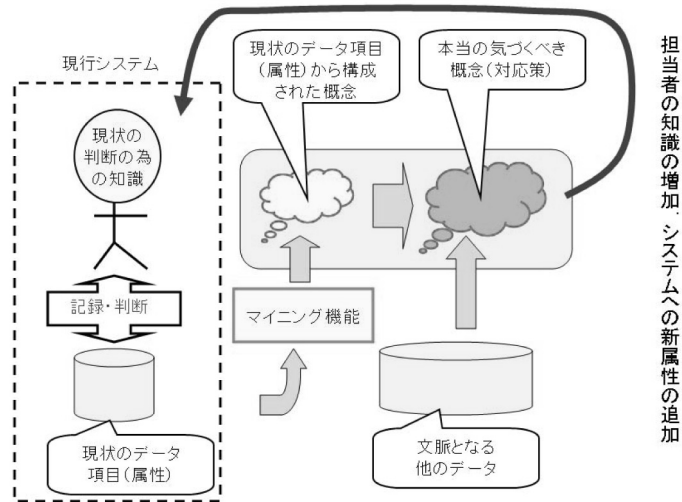


図10：あるべきデータマイニングの姿

の原因分析に重要である」と確信されてはじめて、「では、施工業者名も案件のデータに入れておこう」となるのであって、事前の価値観・視点なしに、マイニング結果により（結果的に）採用される属性を、漏れなく事前に取り込んでおくことは本質的に不可能である。

更に、データマイニングに利用する属性数をむやみに増やすことは別の問題を引き起こす。第1に、業務担当者は多数の属性値を入力せねばならなくなる。この問題は、既に存在しているデータ、たとえば、降雨量のデータや交通量のデータをもってくることである程度対応できる。しかし、他のところから既存データをもってきても、別の深刻な問題が生じる。機械学習では、属性数が多くなると、意味のある概念を取り出すためには、より多くの事例（案件）を必要とする。機械学習は統計処理手法であるため、学習結果として選ばれる候補となる属性数が増加すると、学習用データが多数無いと、統計的推定が正確にはできない。したがって、本システムのように、事例数（案件数）を抑え込みたいアプリケーションでは、属性数を増やすことは危険である。結果として、現実のシステムに取り込まれている属性情報は、すでに業務上で必要なものとして価値観が明確なもののみでなければならない。

以上から、本論文における分析が示す道路の維持管理支援システムとは、固定されたものではなくて、そのシステムが利用される環境状況

が変化すること、あるいは、そのシステムを利用している職員の知識が増えるとともに、案件として記録された属性が取捨選択されて成長してゆくシステムである。ただし、どのような属性がマイニングに効果的となるかを事前に絞り込むことは難しい。だからと言って、不必要に多くの属性をシステムに取り組みと、マイニングに必要な案件数が増大化して、データマイニングが不可能となる。図10に模式化したように、どのような属性が本当の意味で、問題の根本的解決に有効な関連案件の絞込みに生かせるのかの見極めを含めて、他システムで入力されているデジタルデータを、臨機応変に取り込むサービスバスのデータ連携のあり方こそが、データマイニングを真に生かす環境となる。

## 6.2 ナレッジコンシェルジュの必要性

更に、「近隣案件画面」を効果的に活用するためには、種別の選択に正確性が要求される。これは、「案件種別」が特別な属性であるからではなく、データマイニングに利用する属性であるからである。言い換えると、前節の議論のように、近隣案件を絞り込むために、「案件種別」以外の属性を追加する場合には、その追加された属性も含めて、入力値に曖昧性が存在しないことが望まれる。

しかし、「醜いアヒルの子」の定理から、曖昧

昧性が生じるのは、複数の職員が関与する実際の現場では、本質的に回避できない。この問題は、職員、個々の努力では、回避できない。ひとつのアプローチとして、深い業務知識を持った一人の職員が、まずこの案件の種別の統一を図るようなアプローチが考えられる。「原因」欄についても、同様である。専任の職員が、当該欄の記入を確認して、必要に応じて、システムを利用している職員に警告等を発してゆくことが望まれる。このような職員を本論文では「ナレッジコンシェルジュ」と呼ぶ。ナレッジコンシェルジュは、多数のユーザのそれぞれに対して、案件への「原因」欄の記入状況をチェックし、案件の種別選択の妥当性をチェックする。ただし、入力の高難性の解決のすべてをナレッジコンシェルジュにゆだねることには疑問がある。そもそも、「担当職員がそのデータ項目をきちんと入力することに情熱が湧かない。必然性が感じられない。」状態こそが問題であるからである。すなわち、当該職員の担当業務の内容と、情報システムが実現しようとしている機能・役割に乖離があるからこそ、そのようなデータ項目が出現していると考えられるからである。「京都道守くん」が担っている業務は、住民・職員から寄せられた案件について、その案件をいかにすばやく対応するかはその業務組織の目的が設定されている。いわば、「フロー」処理の業務である。一方で、ある場所の側溝なら側溝の設計・施工・管理・修繕といった一連の流れはひとつのひとつの「資産」着目し、そのライフサイクル全体の流れをより効率的にしようとする業務である。いわば、「ストック」の管理についての業務である。そのような現実の担当者の業務内容と運用している情報システムとの整合性が本当にあるのかどうかは、ナレッジコンシェルジュ導入とは別に吟味されなければならない。ただし、そのような吟味がされたとしても、「醜いアヒルの子」の定理や「フレーム問題」のくびきから逃れることはできない。

## 7 おわりに

従来の維持管理業務では、通報・報告を受けた当面の案件への対応に迫られ、蓄積された案件データを有効活用できていなかった側面があ

る。そうしたフロー型の維持管理業務を一步進めて、過去に蓄積された案件データを有益に利用することで、業務の改善を行えないかと考えた。具体的には、予防保全型の業務形態の実現のために、データマイニングを導入し、マイニング結果を表示する「近隣案件画面」をシステムに追加した。しかし、既存の案件データでは、類似案件をマイニングで取り出せても、それが業務上の知識にはなかなか得なかった。大きな原因は、案件データに、「なぜ、そのような案件が起きたか」という「原因」が記述されていることが少なかったからである。

ただし、マイニング結果から知識を作り出すプロセスで観測された問題点は、原因属性の問題のみではない。もうひとつの大きな問題は、マイニングに必要な「案件の種別」が曖昧な点である。この曖昧性によって、貴重な過去の案件データを取りこぼしている恐れがある。案件種別選択の曖昧性は、職員の責任とするのは酷である。「醜いアヒルの子」が示すように、種別の選択は、職員個々の価値観に依存する。本質的に、多数の職員の主観を揃えることは不可能である。この問題は、システムの設計や、職員の意識向上では解決できない。第三者が価値観の統一を行い、「原因欄」の記入漏れや記入の不十分さが無いか否かをチェックすることが望ましい。この目標は、「ナレッジコンシェルジュ」という業務知識に精通した役職を置くことで解決できると考える。

データマイニングは機械学習（教師なし学習）の一種である。機械学習は、基本的にシステムが持っている属性の部分集合を組み合わせで新しい「概念」を生成することしかできない。したがって、未知のどのような状況にも対応するためには、無限の種類の属性をあらかじめ準備する必要がある。もとより、そのようなことはできないし、仮に膨大な数の属性を準備してしまうと、機械学習に必要な事例数が膨大となり、データマイニングの実用性がなくなる。人工知能で言う、「フレーム問題」である。このため、データマイニングの実現には、例えば、施工業者名や工法といった道路管理に関係した属性を、職員の判断で自由にマイニングに追加できるようなシステム環境の構築が必要である。近年、電子調達・電子納品など、道路関係工事においても、施工業者と自治体との間をデジタルデー

タでやりとりするような動きになっている。このような既存のデジタルデータを、システムと連携させ、分析に役立てるようなサービスパス的なシステム環境を構築することが、今後のひとつの課題と考える。

本論文は、平成19年度 京都府・政策ベンチャー事業「府民と職員の満足度向上を目指した道路維持管理について」の議論を踏まえています。関係者に深謝いたします。なお、本論文で述べられた道路維持管理業務に関する見解は、京都府としての見解を示すものではないことをお断りします。

## 参考文献

- 年。
- [10] McCarthy, J., Hayes, P. J., 1969, Some Philosophical Problems from the Standpoint of Artificial Intelligence, *Machine Intelligence Vol.4*, pp.463~502, 1969.
- [11] 大野耐一、「トヨタ生産方式—脱規模の経営をめざして—」、ダイヤモンド社、1978.
- [1] 吉澤憲治、古畑貴志、小野孝司、寺田守正、吉田和正、矢野高一、中村喜輝、佐野嘉紀、井上明、金田重郎：「Web-GIS を用いた道路管理業務支援システム“京都道守くん”の開発」、情報処理学会、SIG-IS、情報処理学会研究報告・2007-IS-99、pp.39-44、2007年3月
- [2] 愛知県建設部道路維持課、「自治体の取り組み～愛知県統合道路管理システムについて」、建設マネジメント技術、Vol.368、pp.33-39、経済調査会、2009年1月
- [3] 窪田諭、小澤田貴泰、加藤誠、小田島直樹、阿部昭博：「道路維持管理システムの長期運用によるデータの整理と分析」、情報処理学会・情報システムと社会環境研究会、情報処理学会研究報告・IS-107-11、pp.71-76、2009年3月
- [4] 藤澤裕樹、古畑貴志、吉田和正、吉澤憲治、寺田守正、中西正樹、金田重郎：「道路管理情報にもとづくデータマイニングの可能性について」、情報処理学会・第70回全国大会、情報処理学会研究報告・6ZF-2、pp.4-637~4-638、2008年3月
- [5] 一瀬邦継、中西正樹、寺田守正、藤澤裕樹、吉澤憲治、金田重郎：「道路維持管理業務支援システムにおけるデータマイニングとその評価」、情報処理学会・第71回全国大会、情報処理学会研究報告・4ZA-6、pp.4-529~4-530、2009年3月
- [6] 茶筌: <http://chasen.naist.jp/hiki/ChaSen/>
- [7] 元田浩、津本周平、山口高平、沼尾正行、「データマイニングの基礎」、オーム社、1996年12月
- [8] 藤井美和、李政元、小杉考司、「福祉・心理・看護のテキストマイニング入門」、中央法規出版、2005年7月
- [9] 渡辺慧、「認識とパターン」、岩波書店（岩波新書）、1978