

道路維持管理業務支援システムにおける データマイニングの適用

一 瀬 邦 継^{†1} 中 西 正 樹^{†2} 寺 田 守 正^{†3}
藤 澤 裕 樹^{†1} 吉 澤 憲 治^{†1} 金 田 重 郎^{†1}

自治体の道路維持管理業務では、住民等からの通報・報告に迅速に対応する必要があり、著者らは、道路維持管理支援システムの開発・導入を図ってきた。しかし、現状では、通報・報告を受けてから動き出す受動的対応である。そこで、予防保全型の管理業務形態の実現を目指し、蓄積された過去の業務データからのデータマイニングを試みた。その結果、案件の属性として「原因」項目の追加をする必要があり、案件の種別選択にも曖昧性があることが明らかになった。これらの問題を解決するには、ナレッジコンシェルジュの設置が望ましい。

Data-Mining Approach to Road Maintenance Support System

KUNITSUGU ICHINOSE,^{†1} MASAKI NAKANISHI,^{†2}
MORIMASA TERADA,^{†3} YUKI FUJISAWA,^{†1}
KENJI YOSHIZAWA^{†1} and SHIGEO KANEDA^{†1}

The authors developed a road maintenance support system named “Kyoto Michimori-kun” to tackle the damage, accident and other troubles on the roads. Currently, the road maintenance operation tends to be passive handling and inefficient because the operation starts after the requests occurrence. To resolve this situation, this paper proposes a new data-mining approach to achieve a active road maintenance scheme. Our the approach adopts a clustering algorithm based on the location and type of the each request. The pilot evaluation shows that there are two serious problems: the ambiguity of the demand type selection and lack of causal analysis in the real operation. This paper recommends the introducing of a knowledge concierge to resolve these problems.

1. はじめに

近年、業務システムに蓄積された業務データからの知識（ナレッジ）の収集とその活用に注目が集まっている。さまざまな種類の情報システムが企業・自治体の業務を支えており、膨大な業務データが蓄積されている。その蓄積された膨大な量の業務データから、データマイニングによって、業務改善につながる知識を生成することができれば、業務の質を高めてくれることが期待される。

著者らは、自治体の道路維持管理業務（以下、「維持管理業務」と呼ぶ）を支援する、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」と呼ぶ）の開発・運用をこれまで行ってきた¹⁾。本稿での維持管理業務とは、住民通報や職員の道路巡視で見発見された、道路における施設破損や不法投棄など（以下「案件」と呼ぶ）に対応する業務を指し、自治体が行う重要な住民サービスのひとつである。

現状の維持管理業務は、近年の住民ニーズや価値観の多様化から住民通報が多様多岐となりやすい。一方、予算や人員上の制約などから、それらに十分対応出来ない状況にある。その結果（1）案件が発生してから対応する受動的対応（2）案件に対して一般的に時間がかかる抜本的対応でない対症療法的対応などに為らざるを得ない状況である。これらは案件の再発と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・対症療法的対応と為らざるを得ない状況になると考えられる。

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業務全体の最適化が図れ、結果として、道路行政に対する住民満足度のさらなる向上が図れると考えられる。

そこで本稿では、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”に蓄積された案件情報から、データマイニングを試みた。対象とするデータとして、A市内における過去5年分、約800件の案件情報に対して緯度経度と案件種別によるクラスタリング処理を行った。その結果、案件が繰り返し発生している場所を抽出できた。また、案件に

†1 同志社大学大学院工学研究科
Graduated School of Engineering, Doshisha University

†2 京都府山城広域振興局
Yamashiro Koiki Shinkoukyoku, Kyoto prefecture

†3 京都府流域下水道事務所
Ryuiki Gesuido-jimusho, Kyoto prefecture

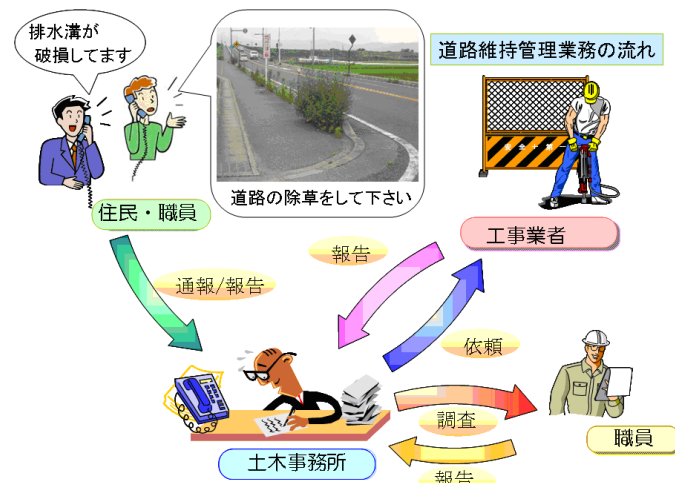


図1 道路維持管理業務の流れ

においては「原因」の記述が重要であり、また、案件種別の統一的な選択には、ナレッジコンシェルジュの設置が望ましいと示された。

以下、第2章で、詳細な維持管理業務の形態について述べる。第3章では、データマイニングの適用を行い、周期的に多発している案件群を抽出する。第4章では、3章の分析結果から「京都道守くん」に機能追加を行った結果を示す。第5章では、ナレッジコンシェルジュの必要性を論じる。第6章はまとめである。

2. 維持管理業務形態とデータマイニングの適用

維持管理業務では、道路上で発生している道路利用上支障となる状況を住民通報や道路巡視業務職員からの報告を受けてからそれらの状況に対応することが多い(図1)。そこでは、いかに迅速に案件の対応をして道路管理上支障となる状況を取り除くかが第一義的な使命である。言い換えれば、その職務は、個々の案件に対応する一連のフロー^{*1}の実現にある。

著者らは「京都道守くん」を開発・導入し、実証実験を行ってきた¹⁾。その結果、案件に

*1 本来は、経済用語として用いられるが、本稿では「案件対応」のことを指すことにする

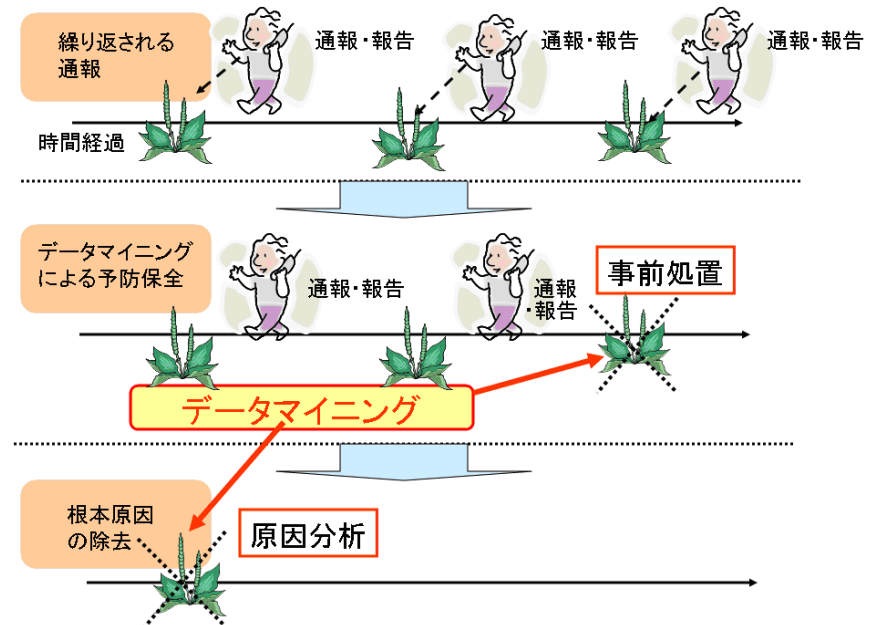


図2 データマイニングを適用した場合のイメージ

対してより迅速な対応が実現できることが確認出来た。しかし、「京都道守くん」は個々の案件への対応フローを支援する狙いから設計したシステムであるため、その効果は限定的と為らざるを得ない。一方、これまでの「京都道守くん」の利用によって、案件に関するデータが蓄積されている。従来のフローの支援だけでなく、蓄積されたこれらのデータを利用することで、業務全体をより効果的・効率的に行える可能性がある。

図2には、道路の維持管理業務へのデータマイニングの適用イメージを示す。図2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、図2中段のように、同一箇所でも同一種別の案件が繰り返されていることを、過去の案件情報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)事前の対応を行って、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。更に一歩進んで、図2下段のように、データマイニング結果から、繰り返される案件の根本的な原因が分かってくれば(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2の中段のように、計画的に工事が実施されると、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。逆に、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2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、同一種別の案件情報から、確実に案件発生を繰り返すを導き出す必要がある。マイニングアルゴリズムは、同一内容の案件を、漏れなく、確実に取り出して利用できなくてはならない。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。この点については、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。いずれにせよ、ここ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。本稿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。

3. 緯度経度に基づくデータマイニング手法

2章で述べた予防保全型の維持管理業務の実現に向けて、過去に実施した案件情報からデータマイニングの適用を試みた。以下に、データマイニングの処理内容と結果を示す。そして、データマイニング結果を踏まえ、システムに反映させる必要がある機能を列挙する。

3.1 データマイニングの処理

本研究では、A市内で京都府が管理する道路で発生した過去数年分の案件情報、約800件に対してデータマイニングを行った。具体的に、以下の手順を踏んだ。

3.1.1 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³⁾」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。

3.1.2 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。

3.1.3 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ものを表1に示す。

3.2 結果

3.1節での手順を踏み、目視により一定の周期で特定の箇所において、まとまった案件として頻発している案件群が確認できた。このようなクラスタは複数見つかり、一部の案件には一定の周期で特定の箇所に頻発することが確認できた。また、図3の結果例のように、いくつかの周期性のある案件群を発見することができた。しかし、この結果から「害虫」が頻発している原因について、実際に案件対応を体験していない者は、推測することができなかった。そこで、上記の案件群における情報を、実際に案件対応を体験した職員に確認した

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3 「害虫」の頻発箇所の特定

ところ、観光地をイメージアップするためと、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。

上記の例では、根本的な対策に至るプロセスにおいて、以下の2つの知識が必要であった。

- 1) 周期性のある案件群が発生していること
- 2) 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、過去の案件情報からデータマイニングによって抽出できる。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」という原因を特定する上で必要となる知識である。後者の知識は、データマイニングからは導出できない。データマイニングが出力した過去の案件に対して、職員が過去に案件対応をしていない場合には、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居る。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。しかし、「害虫」案件に対応した際に、職員が「それがなぜ生じたのか」、すなわち「原因」について分析して、案件情報に記録してい

ば、上記の2) の知識は比較的容易に得られる可能性がある。しかし、この「原因」に関する記述は、実際の案件情報には書かれていないことが多かったのである。

3.3 データマイニングの結果を生かすには

データマイニングの結果を生かすためには、当該案件が生じた原因「害虫の好む樹が街路樹として植えてあること」が、自治体という組織で共有される必要がある。そこで「なぜこれまで組織内で案件が生じた原因を共有してこなかったのか」について考える必要がある。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った。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。そのため「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。

「原因」情報が必要となる背景として、データマイニング自体が持っている性質も関係している。データマイニングは機械学習の一種（教師無し学習）である。機械学習では、案件情報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。逆に言えば、すでに案件情報が持っている属性以外の属性を用いない限り表現できないターゲット概念は獲得できない。

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は、そのまま「使える知識」である。これに対して、本稿が扱う問題は、「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。一步、突っ込んで「それはなぜだろう」を考えて、その「原因」を探し、その原因への対応をしない限り「役に立つ知識の獲得」とはならない。

トヨタの看板方式を提案した大野耐一⁶⁾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、工場が生じた「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、一步進んで「原因」を追究する必要があることを示唆している。本稿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。

ただし、前述のクラスタリングアルゴリズムでは、あらかじめシステムに入力しておいた「案件の種類」をクラスタリングには利用していない。当初、クラスタリングには緯度経度

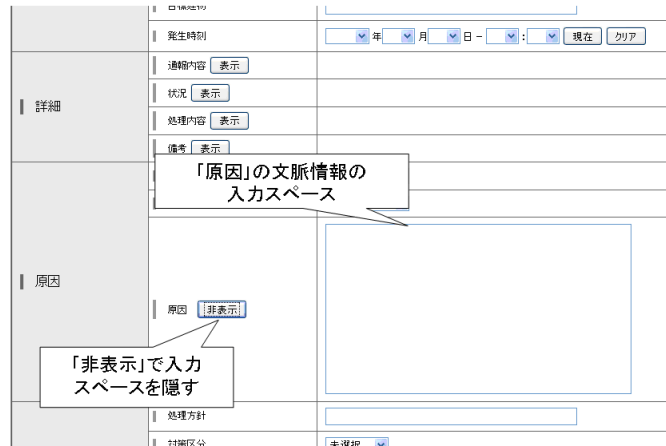


図4 「原因」属性の入力欄

と案件の種別を利用することを試みた。しかし、十分な結果を得ることができなかった。その理由は、「案件の種別に曖昧性があり、同一事象と思われる案件でも、異なる種別が指定されていた」ことによる。本稿のシステムでは、きわめて少ない案件事例から、確実に同一種別の案件を抽出することが要求されている。案件の種別入力値に「ばらつき」があると、本当に必要な近隣案件を見逃す恐れがある。

上記の種別選択の不正確性は、「醜いアヒルの子の定理」によって説明できる。「醜いアヒルの子の定理」は、パタン認識分野では良く知られた定理である。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⁵⁾。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ことを示している。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、複数人の職員がいる場合には、本質的に困難である。

以上の分析結果から、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追加することとした。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(図4)。



図5 「近隣案件画面」の例

また、新たに住民や職員からの案件を受け付けるごとに、近隣で発生した過去の類似案件を表示できれば、過去の案件に蓄積された情報をフィードバックしやすくなると考えた。このことから、通報・報告を受けた案件の近隣で類似した案件が過去に発生しているならば、同じ地図上に表示する「近隣案件画面(図5)」を設置した^{*1}。この画面は、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、職員に気付かせるためである。なお、「近隣」は、半径500メートル以内と設定している。なぜなら、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からである。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。

*1 前述のクラスタリングアルゴリズムでは、ロケーションと単語に基づいてクラスタリングを行った。しかし、この近隣案件画面では、あらかじめシステムがデータとして持っている案件の種別とロケーションによってクラスタリングを行っている。これは、あくまで、新しく通報された案件と類似の案件を検索することを目的とするためである。

申請	申請内容1 (対象物)*	路面、路肩	種別選択リスト
	申請内容2 (状態)*	選択して下さい 選択して下さい	
件名*	雑草 倒木 水たまり 車道部との段差 穴・決壊 危険箇所 その他		
天気			
路線名*			
市町村名*			

図 6 案件の種別選択の例

4. 適用事例

3章では、データマイニングの適用とそこから得られた課題をどのようにしたら解決に向けることができるかを述べた。そして、以下にそのシステムへの適用と評価について述べる。

4.1 システムへの適用

著者らは、維持管理業務を支援するシステム（“京都道守くん”）を開発・導入してきた。このシステムに「原因属性の追加」、「近隣案件画面」の機能を追加した。具体的には、「原因属性の追加」に関して、データベースの概念スキーマに「原因」カラムを付加した。インターフェース上では、案件情報の入力欄以外に、報告書の作成欄に「原因項目欄」を設けた。上記の報告書とは、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。このことから、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できる。

一方、新設された「近隣案件画面」では、案件の種別が一致した案件を類似案件として表示させている。案件の種別は2段階構成になっており、第1段階の「対象物^{*1}」を設定し、対象物で起こりうる第2段階の「状態^{*2}」が設定される（図6）。そして、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」と設定した。このような表示形式で、案件の種別が同じで半径500m範囲の過去の案件を表示するようにした。そして、これらの機能を追加したシステムを実際に運用した。

*1 例えば、「トンネル」、「道路の側溝」、「路面、路肩」など

*2 案件の特徴のこと。例えば、「雑草」、「側溝つまり」など

The screenshot shows a web application interface for managing drainage tank issues. At the top, there's a search filter for '側溝蓋のガタンキについて' (About side drain cover rattling) with details: 応対者:中西正樹, 申請時刻:2009/03/25 15:00, 申請内容:排水施設 - 側溝蓋、樹蓋 補修, 原因: -. Below the map, there are buttons for '地図', '航空写真', and '地図+写真'. A red note says '※近隣の類似した苦情を表示しています' (We display similar complaints from nearby areas). On the right, a '類似案件一覧' (List of similar cases) is shown with two entries: one for '側溝蓋の修繕について' (About side drain cover repair) and another for '道路側溝の新設要望' (Request for new road side drain).

図 7 側溝蓋の傷みに関する近隣案件

4.2 “京都道守” くんにおける評価

運用期間は、2009年2月初旬～同年4月下旬までの約3ヵ月とし、同時にシステム評価を行った。評価対象は「近隣案件画面」である。「原因」属性については、データマイニングの結果から過去の案件を見直す場合に必要になる。即ち、「近隣案件画面」を有効に使えるという評価であれば、「原因」情報を有効に利用していることと等価である。以下に、システムの評価結果を示す。

過去の近隣案件は3、4件に対して1回ほどの出現頻度であった。次に、どのような近隣案件が表示されていたのかについて評価を行った。結果として、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7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^{*3}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。

*3 案件種別は、2段階のメニューになっており、「対象物」、「状態」の順番で選択する。例えば、「対象物」で選択したものによって、「状態」が影響を受ける。

5. 今後の課題

本章では、今後の課題について述べる。「原因属性の追加」によって、原因が持つ情報をドキュメント化することにより、データマイニング結果を案件の対応策に生かすことができると考えた。これは、当該案件と同一箇所での案件対応を経験していない職員でも「原因」を文脈情報として受け取ることによって、根本的な対処方法を検討可能と考えたからである。

しかし、実際に運用してみると、なかなか、すべての案件に原因が記入されているとはなかった。「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が保ち難い。この問題は、職員の意識も問題という見方も出来るかもしれない。しかし、どのような業務でも、「なれ」は生じる。そして、何より、案件対応に必然性が無い「原因」入力は職員に負担となる。仮に、入力されても、次の類似案件に対応するために十分な内容を期待できない恐れがある。別途、「原因」欄が入力されているか否かの何らかのチェックを行う職員が必要である。

また、4章で述べた「近隣案件画面」を効果的に活用するためには、種別選択の正確性を確保しなければならない。しかし、「醜いアヒルの子」の定理から、曖昧性が生じるのは、職員が多数居る場合には、本質的に回避できない。したがって、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効果的と思われる。

この業務知識を持った職員は、「原因」をはじめとする案件に関する記録内容の記入状況、次の案件に対して役立つような内容かどうかをチェックする責務を担う。また、種別の曖昧性を回避するためには、複数の価値観が混在してはならないため、一人で種別選択をチェックする必要がある。本稿では、このような人物のことを「ナレッジコンシェルジュ」(図8)と呼ぶことにする。図8のナレッジコンシェルジュは「原因」欄の記入状況をチェックし、案件の種別選択の管理を行う。この図における案件の種別選択では、「トンネル内でトンネルの壁が崩れ落ちている」案件に対して、「対象物」の選択を行っている。この「ナレッジコンシェルジュ」の設置により、本来の目的である予防的な保全を行う維持管理業務を実現できると考えられる。それ以外にも、「ナレッジコンシェルジュ」は組織内および情報システムに有益な知識を提供することができる。また、過去に生じなかった案件に対しても、新たな知識を提供し、システムの質を向上させてくれる。職員に対しては、過去に起きた案件から効率のよい対処方法やシステムに蓄積されていない案件に対して、過去の知識をどのように生かして対応したらよいのかというノウハウなどを提供する。

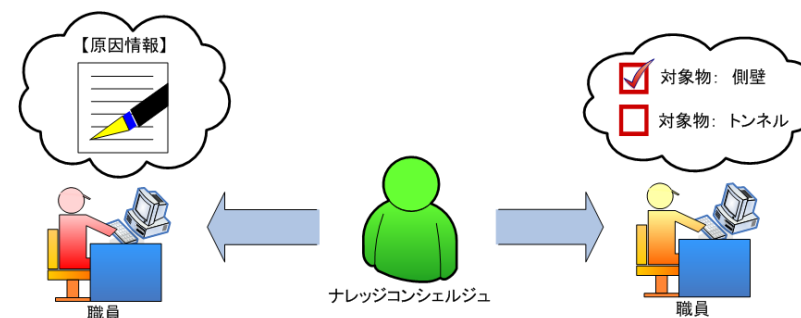


図8 ナレッジコンシェルジュが行う主な業務

更に、運用する中で、将来の課題も抽出された。現状の「近隣案件画面(図7)」では、案件の種別である「対象物」と「状態」で一致した案件^{*1}が上位にリストアップされ、「対象物」のみ一致した案件^{*2}が下位にある。しかし、システムを長期的に運用した場合、案件が膨大に蓄積され「類似度:高」と表記された案件が大量に表示される可能性がある。案件が大量に表示されても、対応するために十分な内容を持つ過去案件があるとは限らない。何より、多数の案件をチェックする負担が大きくなる。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、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。例えば、対象物件の工法から「近隣案件画面」に表示する過去案件を絞り込むことが考えられる。工法に問題があれば、新たに通報・報告を受けた案件が、同じ箇所でも過去に処置した案件と同じ工法を用いているかも知れないからである。工事そのものは道路管理担当とは異なる部署(例えば、建築部門)が担当しているような場合には、維持管理部門と建築部門で工法に関する情報を交流することが必要となる。

近年、調達の公平性を担保するために、電子調達・電子納品といった、施工業者と自治体との間の連絡を、デジタルデータ化する動きも進んでいる。このようなデジタル化された調達・納品データの中に、施工業者名や工法名が記録されており、それをマイニングにも利用

*1 「類似度:高」と表記された案件

*2 「類似度:低」と表記された案件

できるようになれば、更に、マイニング結果の範囲を絞り込むことができる。

しかし、人工知能で言う「フレーム問題」⁴⁾からは逃れることは本質的にできない。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。さりとて、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。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. 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追われ、蓄積された案件情報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情報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、既存の案件情報では、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、案件情報に「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、貴重な過去の案件情報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な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て新しい「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關係した属性を、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バスの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。

謝辞

最後に、本研究に関してご協力いただいた自治体職員各位に、深く感謝いたします。なお、本稿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最後にお断りします。

参考文献

- 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
- 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) 茶筌: <http://chasen.naist.jp/hiki/ChaSen/>
- 4) McCarthy, J., Hayes, P. J., 1969, Some Philosophical Problems from the Standpoint of Artificial Intelligence, Machine Intelligence Vol.4, pp.463~502, 1969.
- 5) 渡辺慧,「認識とパタン」, 岩波書店(岩波新書), 1978.
- 6) 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.
- 7) 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
- 8) 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
- 9) 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