

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

藤澤裕樹^{*1} 古畑貴志^{*2} 吉田和正^{*2} 吉澤憲治^{*2} 寺田守正^{*3} 中西正樹^{*4} 金田重郎^{*5}

同志社大学工学部^{*1} 同志社大学大学院工学研究科^{*2} 京都府中丹広域振興局建築部^{*3}

京都府山城広域振興局建設部^{*4} 同志社大学大学院・総合政策科学研究科/工学研究科^{*5}

1. はじめに

自治体における道路管理業務では、除草や側溝蓋の破損などの苦情を住民から受けてから動く対症療法的業務となる場合が多い。しかし、これでは、自治体職員・工事担当者双方にとって非計画的となり、業務効率があがらないだけでなく、苦情そのものも減少しない。住民からの苦情を事前に予測できれば、業務計画を立てやすく、住民満足度向上にもつながる。

そこで、本稿では、過去の維持管理データの苦情の詳細や発生箇所、時期を用いて、繰り返し発生している苦情を抽出する手法を提案する。具体的には、管理データ中の苦情を特徴つける単語出現頻度と発生箇所を併用したマイニングを行い、繰り返しの苦情の抽出を試みた。

京都府が管理する道路の宇治市内分の維持管理データに本提案手法を適用した。この管理データには過去 5 年分、約 800 件の苦情データが含まれる。本提案手法を適用した結果、苦情が繰り返し発生している場所が抽出された。一方、工法や発生原因などクラスタリングに必要な情報を残しておくには、当該情報を残すことが業務上不可欠となる枠組みが必要であることも明らかとなった。

2. 現状の課題

2.1 現在の道路維持管理業務

現在の道路維持管理業務では自ら巡視も行って、住民からの通報があつて初めて苦情を処理する対処保全型維持管理[1]とでも言うべき方法がとられていることも多い。

しかしこの方法では次の様な問題が発生する。

- 1) 住民からの通報の対応に時間がとられるため、本来の業務を行うための時間の確保が困難
- 2) 業者への発注が通報単位になるため効率的に工事が行われず工事費用が増大

この状況を打破するためには、新たな道路維持管理業務の方法が必要となる。

2.2 予防保全型維持管理

対処保全型維持管理によって起こる課題を解決するため、苦情の発生箇所と発生時期を事前に予測し、住民からの通報以前に対応する予防保全型維持管理手法[1]とでも言うべき方法が提案されている。この方法によれば、次の様な効果が期待できる。

- 1) 予測により苦情が出る前に対処すれば、住民からの苦情数が減少するため、苦情への対応時間が減少し、本来の業務管理を行う時間が確保できる。
- 2) 隣接した工事については、住民への通報単位ではなく、業者へ一括発注でき、コスト削減につながる。

但し、予防保全型維持管理の実現のためには、苦情の予測が必要となる。

2.3 データマイニング適用の可能性

住民からの苦情は内容によって起こる季節や時期、場所等それぞれ特徴がある。そこで、維持管理データにデータマイニング技法を適用し、苦情の発生を予測できれば、予防保全型維持管理の実現が可能となる。

図 1 には「害虫」に関する苦情の発生場所の例を示す。苦情の種類によっては、図 1 のように、限られた場所に集中して発生しているケースがある。そこで、同一種別の苦情を、緯度経度によってクラスタリングすれば、繰り返し発生している苦情を抽出できると考えた。

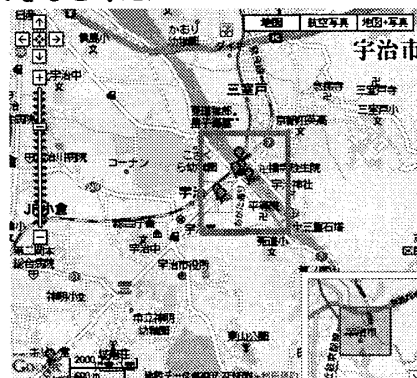


図 1: 「害虫」に関する苦情の発生場所

Data Mining Approach from Road Management Records

^{*1} Faculty of Engineering, Doshisha University

^{*2} Graduate School of Engineering, Doshisha University

^{*3,4} Kyoto Prefecture

^{*5} Graduate School of Policy and Management, Doshisha University

3. 緯度経度を用いたクラスタリング手法

以下に苦情種別と発生箇所を併用したデータマイニング手法を提案する。

【STEP1】形態素解析

維持管理データのタイトルや苦情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て、各苦情に記録された単語を求めた。形態素解析には「茶筌」を用いた。形態素解析対象は、人手で記録された苦情のタイトル、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した。

【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て、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した。ただし、以下の評価実験では、苦情数が約 800 件とすくないため、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 件以上の苦情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 上で案件を表示して、目視で内容確認する。

4. 評価実験

京都府が管理する道路の宇治市内分の過去 5 年間の維持管理データに対して、すべて緯度経度情報を付加して、提案手法を適用した。その結果、形態素解析で得られた 6,120 個の単語のうち、171 個の単語でクラスタリングが成功した。いくつかのクラスタで繰り返し発生している案件を発見した

1) 結果例

「側溝」での結果例を図 2 に示した。

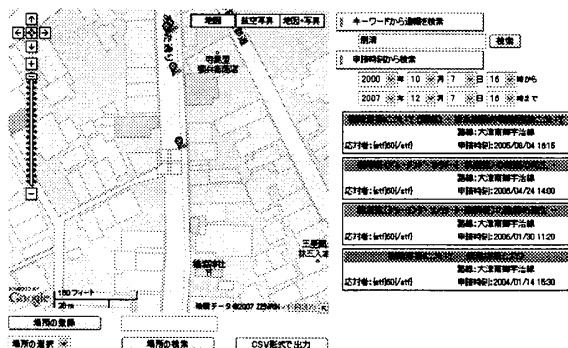


図 2: 「側溝」に関する苦情での結果例

「側溝」に関する苦情は 103 件あり、クラスタは 7 個できた。しかし、業務多忙などのために、現場対応が優先となり、結果的に手間となる原因や処理のデータ化がされていない。その為、どのクラスタでも、繰り返し発生する苦情の原因を特定することができなかった。3 つの苦情が 2004 年 1 月 14 日、2005 年 8 月 4 日、2006 年 4 月 24 日とほぼ一定の期間で繰り返されている。泥が溜まるなどの原因がある可能性があり、対応が必要と考えられる。

また、「除草」に関する苦情は 39 件あり、クラスタは 3 個できた。しかし、これも同様に繰り返し発生する苦情の原因を特定することができなかった。

2) 考察

提案手法により、いくつかの事例で周期性を発見できた。しかし、苦情の原因や処理がデータ化されていないため、なぜ繰り返し苦情が発生しているのかがわからなかった。また、実際の苦情の多くは、その原因や苦情に対する処理によって、次に苦情が発生する時期が変わる。例えば、側溝蓋の破損の場合、何が原因で側溝蓋が破損したのか、どんな種類の側溝蓋を修理に使うかによって、次に破損する時期が変化する可能性がある。現在の維持管理データには苦情の原因やそれに対する処理が記入されていない。予測精度を上げるためにはこれらの記入が必要である。しかし、現状システム[2]では原因や処理を記入する枠組みがない。破損原因の調査結果や補修に使用した材料工法などのデータ記入といったいわゆる電子納品や支払い等の全体の業務を一体としてシステム化し、それらの中で、工事の種類、担当者、工事日や、トラブルが起きたときの原因を報告させる枠組みが必要である。

5. おわりに

維持管理データを形態素解析で分類項目を決めて分類し、緯度経度によってクラスタリングを行った。提案手法により繰り返し発生する苦情を一定量は抽出可能である。しかし、精度を上げるためには、苦情の原因やそれに対する対処の記述が必要となる。そのためには、電子納品・支払い等の全体の業務を一体としてシステム化し、業務担当者の負担を増加させずに、それらが記入される環境を作ることが重要と考えられる。

参考文献

- [1] 寺田守正, “道路管理を科学する”, 国土交通省近畿地方整備局管内技術研究発表会, 2003
- [2] 吉澤他, 「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.