

(39) structural topic model を援用した 橋梁点検所見の分析

小川 福嗣¹・近田 康夫²・中山 晶一朗³

¹正会員 金沢大学 総合技術部 (〒920-1192 石川県金沢市角間町)
E-mail: ogawa@se.kanazawa-u.ac.jp

²正会員 金沢大学教授 地球社会基盤学系 (〒920-1192 石川県金沢市角間町)
E-mail: chikata@staff.kanazawa-u.ac.jp

²正会員 金沢大学教授 地球社会基盤学系 (〒920-1192 石川県金沢市角間町)
E-mail: nakayama@staff.kanazawa-u.ac.jp

本研究では、橋梁点検データの所見記述に着目し、structural topic modelを援用することにより、専門技術者の総合的な判断を利用した点検結果の分類および損傷の把握を試みた。北陸地域における早期劣化の一因であるASRに関する損傷を生じているなど特徴的な損傷形態をトピックとして抽出されることを示した。

Key Words: structural topic model, bridge management, maintenance

1. はじめに

(1) 背景

橋梁をはじめとした社会インフラの老朽化が進行しており対策が急務となっている。2012年の笹子トンネル天井板落下事故を契機に社会的にもインフラ維持管理の重要性が認知され、定期点検実施体制の整備や長寿命化修繕計画の策定などの取組みが進められてきた。また、点検結果の管理および活用するために橋梁マネジメントシステムが構築・導入され、蓄積された点検データは、現状の分析や劣化予測などの研究でも活用されてきた。既存研究では、主に定期点検要領に定められた各部材ごとの定性的な損傷評価項目や交通量や床版厚さといった属性情報に着目し、橋梁点検結果の特徴や分類が行われてきた¹⁾²⁾³⁾。しかし、橋梁は架設や設計の条件が多様であり、上述の定性的な項目のみでは現状を十分に表現できず、専門技術者の判断も重要であると考えられる。定性的な項目のみでは表現できない事項については専門技術者による点検所見に表れると考えられる。本研究では近年開発がすすめられた自然言語処理技術を用い点検所見から、橋梁点検結果の分析を試みた。

(2) 既存研究

a) 橋梁点検の課題

老朽化が進行する橋梁の適切な維持管理実施の必要性

については十分に認知されているが、財政や人材面において課題がある。工藤・杉本⁴⁾はある地方公共団体に属する市町村について、橋梁及び財政状況に関する情報を利用して橋梁維持管理のシミュレーションを行ったところ、十分な補修が実施できない状況に陥る可能性があり、安全確保に支障をきたす可能性があるため、漏水や滞水といった対応を行ない中長期的な維持管理の抑制を図ることが重要であるとしている。美濃・森川⁵⁾は橋梁の維持管理体制の実態について、専門技術者として長寿命化修繕計画策定にかかわった事例を分析し、特に市町村においては人材及び管理橋梁数が少なく、土木全般に幅広い知識が求められ、維持管理に関する専門的な知見が蓄積されない状況を明らかにしている。

b) 橋梁点検結果の活用

得られた点検結果を活用して詳細に分析や評価を行う研究は多数行われている。本城ら¹⁾は注視すべき橋梁を簡易にスクリーニングできるように、橋梁諸元、環境条件やひび割れや漏水の有無といった多項目の点検結果を有効に活用して変状の特徴に応じて橋梁を層別する手法を提案している。玉越ら²⁾³⁾は国土交通省の各地方整備局で得られた橋梁点検結果を用いて、コンクリート橋及び鋼橋の劣化の特徴について詳細な分析を行っている。鋼橋の主桁の腐食損傷に焦点を当てた重回帰分析では、考慮可能な属性情報を説明変数に加えた場合でも決定係数は0.28程度であり、点検結果に表れる橋梁の不確実性

が大きいことが示されている。橋梁の健全度評価は離散値であることから、このようなデータの性質に合わせた劣化予測が行なえるようにマルコフ劣化ハザードモデルが開発され⁶⁾、拡張モデルの開発⁷⁾も進められてきた。

c) 記述データの分析

Bleiら⁸⁾によって文章における単語の出現のモデル化 LDA(Latent Dirichlet Allocation) が提案され、その後改良モデルが多数提案されてきた。これらは総称して潜在トピックモデルと呼ばれ、単語の出現に限らず共起情報が大量にあるデータに応用できることから、画像処理や情報推薦等の様々な分野において応用されている。土木分野においても既にいくつかの応用例がある。塚井・塚野⁹⁾は地理情報であるメッシュデータに対してトピックモデルを適用し、因子分析と比較考察を踏まえて解析結果の解釈や集計の容易さからトピックモデルの有効性を示している。塚井・椎野¹⁰⁾は比較的語数の少ない地域公共交通会議の討議録に対してトピックモデルを適用し、討議の内容に応じたトピックが抽出できることおよび討議の進行を反映した内容の推移が得られることを明らかにし、比較的簡便な手続きによって討議の要約が行える可能性について検討を行っている。

(3) 研究目的

橋梁は多種多様であるため、橋梁の健全性に関しては専門技術者の総合的な判断が重要と考えられる。点検所見には橋梁全体に関する判断事項や定性的な評価項目で表現しきれない事項が記載されると考えられる。これらは点検項目やそのカテゴリを増やすことにより診断の精緻化ができると考えられるが、入力や管理の手間が増大する。橋梁点検所見から専門技術者の知見の抽出、損傷の特徴や分類を行えることができれば、点検項目を不必要に増加させる必要がなく、管理の省力化や効率化につながると思われる。本研究では、蓄積された点検所見に着目し、自然言語処理技術の一つである **structural topic model**¹¹⁾を援用して橋梁点検結果の分析を試み、橋梁評価の特徴を明らかにするとともに橋梁の分類に利用できることを示す。

2. 方法

(1) 利用するデータ

本研究では、富山県の橋梁点検データを対象とする。富山県橋梁点検マニュアルによると、すべての部材に発生した損傷を詳細に把握することを目的とし、橋梁各部に触れる程度まで近接して目視を行う詳細点検、代表径間を対象として地上から目視できる範囲で行う概略点検がある。本研究では所見欄の書込み量が充実していた詳

細点検結果を解析対象とする。富山県の橋梁点検データのうち、詳細点検が行われ所見記述が記載された 760 橋梁を解析対象とした。

点検所見の解析を行うにあたり、文章から情報を抽出するためには意味を持つ最小の記号・文字列の単位である形態素に分割する必要がある。文を形態素に分割し、品詞や記号の名称等を付与することを形態素解析といい、既に複数の形態素解析器の開発が行われている。本研究では辞書追加や後の解析などの利便性の点から MeCab¹²⁾を用いる。標準の形態素解析では、例えば「床版」といった専門用語は「床」と「版」という単語に分類されてしまう。そのため、必要に応じて専門用語等の単語を辞書に追加する必要がある。このような専門用語の形態素解析の精度向上に取り組む研究もある¹³⁾が、本研究で対象としたテキストデータは語彙数が限られていることから形態素解析結果をもとに適当に用語が抽出されるように手動で辞書の追加登録を行なった。得られた形態素解析結果から橋梁の損傷状態等を表すと考えられる品詞として名詞を解析対象とした。また、品詞の中でも細分類されているが、損傷状態等とあまり関係がないと思われた接尾、代名詞、数に細分類された単語は解析対象から除外した。抽出された形態素の中には、“◇”などの記号も含まれるため、これらはストップワードとして任意に設定し除外した。総出現単語数の 90%を約 360 の語彙が占めており、比較的少数の語彙で橋梁点検所見が記載されていると考えられる。この中でも頻出度の高い単語としては、「腐食」、「補修」、「損傷」、「ひび割れ」が総単語数の約 10%を占めており、頻出度の非常に高い単語といえる。腐食やひび割れは国土交通省の定める橋梁定期点検要領の損傷の種類として設けられている単語・項目であり、これらに関する損傷状態を中心に点検及び診断が行われていると考えることができる。

単語の出現に影響を与えるその他の属性情報として本研究では、橋長、上部工および下部工の健全度 (Health Index) を利用する。健全度は維持管理システムにより自動的に算出され、算出方法は損傷の程度と損傷割合を基に、損傷がなく健全な状態を 100 とし、0~100 の範囲で算出される。損傷評価点 (Damage Index) は、上部工は床版、主構、床版・主構以外の部材別に、下部工は躯体及び基礎の部材別に評価評価の重みづけ平均によって評価される。また各部材別評価は定期点検要領に示される損傷の標準的な項目の重みづけ平均で評価される。なお、このときの重みづけの設定は、橋梁への重大な損傷を考慮し、例えば、鋼床版では「亀裂」に関する損傷評価点が 100、つまり健全度が 0 となるように重みづけが行われている。

健全度 (Health Index)

$$= 100 - \sum \text{損傷評価点(Damage Index)} \quad (1)$$

(2) 分析方法

点検所見においては、橋梁の健全度が低い場合には、断面欠損などの重大な損傷がみられるため補修が必要といった記述、健全度が高い場合には、損傷はほとんど見られないといった記述が想定される。つまり、健全度に応じて記述内容が異なると考えられるため、記述文章以外の健全度などの補助情報の活用により、より適切な単語生成のモデル化を行うことが可能と考えられる。本研究では補助情報を考慮可能な structural topic model を用いる。structural topic model では、式(2)~(5)による単語の生成モデルを仮定している。 θ_d は文章 d におけるトピック分布、 β_d^k は文章 d 、トピック k における単語分布をそれぞれ式で表現されるとし、文章 d における単語 $w_{d,i}$ は潜在トピック $z_{d,i}$ を通して生成されると仮定する。なお、それぞれ多項分布であるとする。

$$\theta_d \sim \text{LogisticNormal}(\mu, \Sigma)(\theta_d = \theta_{d,1}, \dots, \theta_{d,K}) \quad (2)$$

$$\beta_d^k \propto \exp(m + \kappa^k + \kappa^y + \kappa^{k,y}) (\beta_d^k = \beta_{d,1}^k, \dots, \beta_{d,v}^k) \quad (3)$$

$$z_{d,i} \sim \text{Multi}(\theta_d) \quad (4)$$

$$w_{d,i} \sim \text{Multi}(\beta_d^{k=z_{d,i}}) \quad (5)$$

μ を K 次元正規分布の平均ベクトル、 Σ は $K \times K$ の共分散行列を表し、これによりトピック間の相関を考慮できる。また、 m はベースとなる単語の出現確率、 κ_v^k はトピックにおける単語分布、 κ_v^y は補助情報による単語分布、 $\kappa_v^{k,y}$ はそれらの相互作用を表しており、これにより共変量に応じた単語の出現を表現することができる。

モデルの適用に当たっては、事前にトピック数を設定する必要がある。トピック数の選択指標はいくつか提案されているが、Roberts ら¹¹⁾は Semantic coherence と Exclusivity の2つの指標を用いて、トピック数を決定することを推奨している。Semantic Coherence は意味の近い単語がトピックに集約されているかを表す指標であり、文章中に単語 v と v' が同時に出現する回数を $D(v, v')$ 、トピック k における最頻出単語を M とすると(6)式で表現される。

$$C_k = \sum_{i=2}^M \sum_{j=1}^{i-1} \log \left(\frac{D(v_i, v_j) + 1}{D(v_j)} \right) \quad (6)$$

Exclusivity は FREX という指標を用いて(7)式で表現される。

$$\text{FREX}_{k,v} = \left(\frac{\omega}{\text{ECDF}(\beta_{k,v} / \sum_{j=1}^K \beta_{j,v})} + \frac{1 - \omega}{\text{ECDF}(\beta_{k,v})} \right) \quad (7)$$

ここで ECDF は経験分布関数(Emprirical CDF)であり、 ω は重み係数でありここでは 0.7 を用いる。これらの指標を参考に本研究ではトピック数は 30 とした。

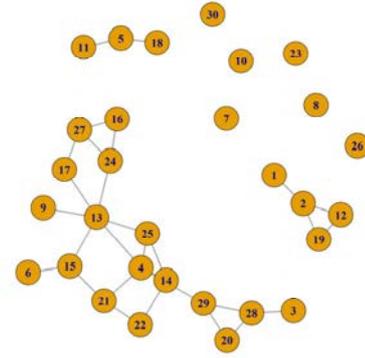


図-1 トピック間の関係

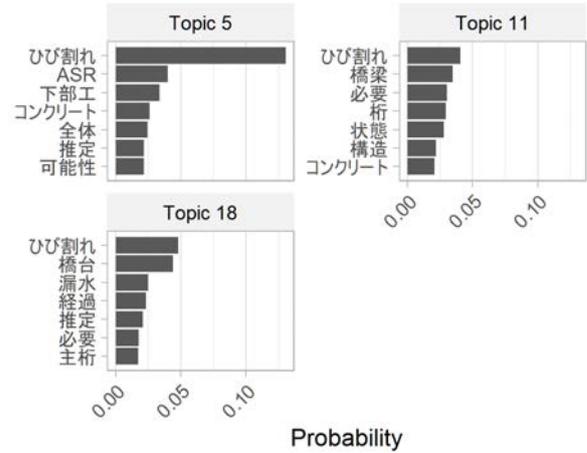


図-2 トピックにおける出現確率上位の単語

3. 結果

推定されたトピック間の関係を図-1に示す。トピック間の相関が 0.01 以上の場合リンクが張られている。図からトピックは一つの大きなグループとそのほかのグループに分類されていることが分かる。大きなトピックグループについて、各トピックの上位頻出単語には、一般的な損傷に関する単語が多く確認されたことから、管理橋梁全般に観察される一般的な損傷トピック群であると考えられる。ここでは、その他の特徴的な損傷が表れると考えられるトピックグループに着目する。トピック 5, 11, 18 のグループに着目する。これらのトピックの出現確率上位の単語を図-2に示す。上位頻出単語には「ひび割れ」、「橋台」や「コンクリート」と言った単語が含まれており、コンクリート部材に関する損傷に関するトピックと考えられる。また、トピック 5 には他のトピックの上位出現単語には表れない「ASR」が確認できる。ここで、各文章ごとに推定されたトピック分布から最も確率が高いトピックをその文章のメイントピックとした場合、トピック 5 が主要トピックと推定された橋梁は 27 橋中 19 橋、トピック 11 が主要トピックと推定された橋梁は 11 橋中 5 橋において、アルカリ骨材反応や

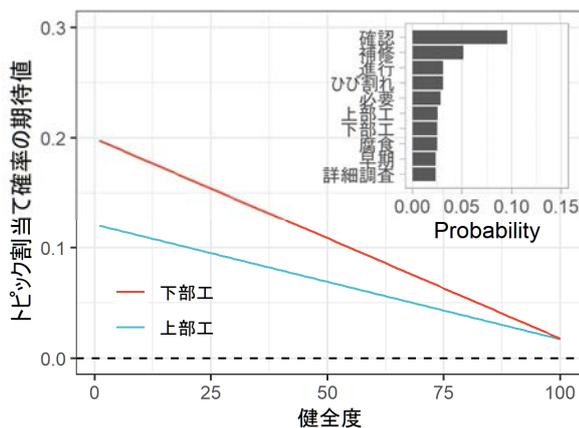


図-3 トピック 10 と健全度の関係

アル骨, ASR といった単語が含まれていることが確認され, ASR に関連するトピックであるといえる. 一方で, トピック 18 が主要トピックと推定された橋梁は 14 橋中 2 橋でしか ASR 関連の単語が確認されなかったが, 広範囲にひび割れが見られる等の記述が多く確認された. このことから, これらの 3 つのトピックは広範囲にわたるひび割れが主要な記述内容であるトピックとしてグループを形成していると考えられる. その他のトピックグループについてみると, 主要トピックが 8 と推定された所見記述には「コンクリートの剥離が生じており, 鉄筋露出がみられる」といった比較的重大な損傷が生じている記述が多く確認されるなど, 特徴的な損傷形態に関するトピックとして形成されていることが確認された.

次に上部工および下部工の健全度とトピックの割り当ての関係について確認する. 本稿では特に健全度の関係が明確に確認されたトピック 10 について, トピック出現確率と健全度の関係について図-3 に示す. 図中右上にはトピック 10 の上位頻出単語を示す. 図からトピック 10 は上部工及び下部工の健全度が低いほどトピックの出現確率が大きくなっている. つまり, トピック 10 の出現確率が高い橋梁は健全度が低い傾向にあることが想定される. また, 主要トピックがトピック 10 と推定された橋梁の健全度を確認すると他のトピックを主要トピックとして推定された橋梁に比べて健全度が低い傾向にあることが確認された. トピック 10 の出現確率上位の単語を確認すると, 「早期」, 「補修」, 「必要」, 「進行」があり, このことからトピック 10 は損傷が進行しており補修が必要な段階にある橋梁を表しているといえる. モデルにその他の属性情報を取り込むことによってより明確に橋梁の損傷形態別のトピックを抽出することが可能となったといえる.

4. おわりに

本研究では, 専門技術者による点検所見に着目し, structural topic model を用いて点検結果の特徴分析を行った. 橋梁の北陸地域において早期劣化の要因となっている ASR に関する損傷 (橋梁) をトピックとして抽出され, 損傷形態に関するトピックが抽出できることを示した. また, トピックに健全度などの情報を考慮することにより, 損傷程度も考慮したより詳細な損傷形態に関するトピックが抽出可能であることを示した. 点検所見の分析により橋梁や損傷の分類を行えることで, 点検項目やそのカテゴリの追加とは別の形により橋梁管理システムの高度化が期待できる.

参考文献

- 1) 本城勇介, 大竹雄, 佐藤敦, 流石堯, 小林孝一, 宗宮裕雄: 岐阜県橋梁点検データベースの統計解析に基づく簡易橋梁健全度評価指標, 構造工学論文集 A, Vol. 60A, pp. 462-474, 2014.
- 2) 玉越隆史, 横井芳輝, 石尾真理: 全国規模の実測データによる道路橋の劣化特性とその定量的評価, 土木学会論文集 F4 (建設マネジメント), Vol.70, No.4, pp. I_61-I_72, 2014.
- 3) 玉越隆史, 横井芳輝, 石尾真理: 全国規模の点検データに基づく道路橋のコンクリート部材の劣化の特徴, コンクリート工学論文集, Vol.25, pp.167-180, 2014.
- 4) 工藤正行, 杉本博之: ある地方公共団体に属する市町村の橋梁維持管理体制と課題, 構造工学論文集 A, Vol.61A, pp.486-494, 2015.
- 5) 美濃智広, 森川英典: 自治体道路橋梁維持管理体制の現状と課題, 土木学会論文集 F4 (建設マネジメント), Vol.71, No. 3, pp.182-189, 2015.
- 6) 津田尚胤, 貝戸清之, 青木一也, 小林潔司: 橋梁劣化予測のためのマルコフ推移確率の推定, 土木学会論文集, Vol.2005, No. 801, pp. 801_69-801_82, 2005.
- 7) 貝戸清之, 小林潔司, 青木一也, 松岡弘大: 混合マルコフ劣化ハザードモデルの階層ベイズ推計, 土木学会論文集 D3 (土木計画学), Vol.68, No.4, pp.255-271, 2012.
- 8) D. Blei, A. Ng, and M. Jordan: Latent dirichlet allocation, *Journal of Machine Learning Research*, Vol.3, pp.993-1022, 2003.
- 9) 塚井誠人, 塚野裕太: トピックモデルによる詳細地理情報分析, 土木学会論文集 D3 (土木計画学), Vol. 74, No.2, pp.111-124, 2018.
- 10) 塚井誠人, 椎野創介: 討議録に対するトピックモデルの適用, 土木学会論文集 D3 (土木計画学), Vol. 72, No.5, pp. I_341-I_352, 2016.
- 11) M. Roberts, B. Stewart, and D. Tingley: stm: R package for structural topic models, *Journal of Statistical Software*, Vol.91, pp.1-40, 2019.
- 12) T.Kubo., Mecab: Yet Another Part-of-Speech and Morphological Analyzer, <http://taku910.github.io/mecab/>. (アクセス 2020.6.1)
- 13) 木浪孝治, 池田哲夫, 村田嘉利, 高山毅, 武田利明: 看護学分野の専門用語抽出方法の研究, 自然言語処理, Vol.15, No.3, pp.3-20, 2008.