

テキストマイニングによる建設現場における労働災害の要因分析

Factor Analysis of Occupational Accidents at Construction Site by Text Mining

○都築 彩音^{*1}, 斎藤 寛彰^{*2}
Ayane Tsuzuki^{*1}, Hiroaki Saito^{*2}

*1 戸田建設株式会社総合企画部 学士(情報)

General Planning and Policy Department, Toda Corporation, B.IS.

*2 戸田建設株式会社イノベーション推進センター 主任 修士(工学)

Chief, Innovation Center, Toda Corporation, M.Eng.

Summary: Creating safe environment without occupational accidents, control of working long hours and innovation in work style for making work life balance happen are hot topic. Construction industry is recognized as industry that accidents happen frequently by Ministry of health Labor and Welfare. This study clarifies factor which result in occupational accidents to aim at prevent from that in construction industry. The main analysis is carried out focusing on accident situation which is written by natural language. Among those situations, three factors (falling from high position, get caught in something, falling) which have either high number of accidents and high probability of serious accidents occurrence are analyzed by morpheme analysis, TF-IDF analysis and co-occurrence analysis. These are analyzed by KHCoder3 and serious factors for occupational accidents are detected. One is caused by not following equipment usage rules, second is when carry heavy goods and last is when doing shortcut. All three factors caused by human mistake and should be considered for decreasing serious accidents. But for considering countermeasures, it is not enough to analyze data from only one company, and data from the entire construction industry should be available.

キーワード: 労働災害; テキストマイニング; 建設業; 自然言語処理

Keywords: Industrial accident; Text mining; Construction industry; Natural language processing.

1. はじめに

1.1 研究背景

労働災害のない安全・安心な職場づくり、長時間労働の抑制やワーク・ライフ・バランスの実現に向けた働き方改革が国家主導で進められている。中でも、建設業は厚生労働省の労働災害重点業種と認定されており、労働災害の多い業種とされている。また、建設業就労者の高齢化が進み、若い労働力が不足している中で「建設業＝危険」というイメージで認識され、建設業が若者から敬遠される要因の一つとなっている可能性もある。建設業の将来を考えるにあたって、労働災害の防止・抑制が喫緊の課題である。

1.2 本研究の目的と既往の研究

本研究では、建設業における労働災害の防止・抑制を目的に、テキストマイニングによって労働災害につながる要因を労働災害の発生状況の視点から明らかにする。

構造化データによる労働災害要因の分析は多くの事例があり、例えば土木工事における斜面工事での墜落を扱った伊藤の研究¹⁾や、自然災害の復旧・復興工事の労働災害の発生状況を分析した伊藤らの研究²⁾がある。一方で、労働災害の発生状況やその要因の詳細は自然言語で

記述される部分が多く、これらは情報量が多いものの分析が困難であったため積極的に活用されてこなかった。しかし、近年は自然言語処理を労働災害の分析に用いた研究も現れ始めてきており³⁾、従来では困難であった詳細な分析が可能となっている。

そこで、本研究では、主に自然言語で記述された労働災害発生状況に着目し、労働災害に繋がる要因を明らかにすることを試みる。

2. 利用データ

2009年から2019年の過去10年間に当社の建築工事・土木工事の施工中に発生した労働災害データを使用する。全3,156件のうち熱中症、私病、通勤災害、交通事故、その他、分類不能、といった管理上防止が困難な災害を除く2,413件を対象とする。なお、建築工事に関する労働災害が78.1%、土木工事に関するものが21.6%である。データは大きく3つに分類され、職種や事故の型などの構造化データと年齢や進捗率などの数値データ、労働災害発生状況や人的要因といった自然言語データが存在する(Table1)。

本研究では、従来積極的に活用されてこなかった自然

言語データを対象として、構造化データのみでは詳細な分析が困難であった作業内容や起因物、シーン、行動・意識などの視点から分析を行う。

なお、利用するデータには、就業した作業員データのうち労働災害が発生しなかったデータは収集されていないという制約がある。

Table1 Data set

データ項目	データ内容	データ構造
職種	土工など49種	構造化データ
年齢	実年齢	数値データ
経験年数	実経験年数	数値データ
作業種別	型枠組立など63種	構造化データ
休業区分	不休など5種	構造化データ
事故の型	転倒など18種	構造化データ
進捗率	工事進捗状況 (%)	数値データ
災害発生状況	-	自然言語データ
人的要因	-	自然言語データ
物的要因	-	自然言語データ

3. 分析プロセス

3.1 分析対象の選定

本研究で分析対象とする「事故の型」として、労働災害発生件数 200 件以上かつ重大労働災害発生率 15% 以上であり、特に何等かの対策が必要と考えられる「墜落・転落」「はさまれ・巻き込まれ」「転倒」の 3 種類に着目して分析を進める。「墜落・転落」に分類される労働災害は、例えば足場や脚立といった高所からの転落を指す。「はさまれ・巻き込まれ」は資材に指を挟むや工具に手を巻き込まれるなどが分類され、「転倒」は何かにかみつき転倒するや階段で滑って転ぶなどの労働災害が分類される。重大労働災害発生率とは全災害の中で休業区分が休業 4 日以上またはそれより重篤な状態となった割合のことを指す。

3.2 手法

まず、災害発生状況を記述したテキストに対して形態素解析を行い、頻出単語を抽出する。また、特徴語を抽出するため、形態素解析で抽出した上位 200 語を「墜落・転落」「はさまれ・巻き込まれ」「転倒」の 3 種類に分けて TF-IDF 値を算出する。その後、単語の出現頻度と特徴語を踏まえて、語と語の関係性の強さを図る共起分析を行う。共起分析では、労働災害の特徴を表すために、語と語の関係性を図る分析と職種を外部変数とした職種別特徴分析の 2 種類の分析を行う。一連の分析には KHcoder3 を用いる。

3.2.1 形態素解析

形態素解析の前処理として、建設業特有の用語で未知

語に分類される「クランプ」「法面」「親綱」などの単語を強制抽出対象として登録し、逆に事故状況の抽出の妨げとなる「安衛」「前」「他」などの単語は除外対象とした。また、「作業」のような汎用性の高い単語は「作業床」「作業員」「作業台」「高所作業車」のように頻出の複合語として強制抽出対象とし、「可搬式作業台」「立馬」「立ち馬」「カッターナイフ」「カッター」のような表記ゆれが観測された語を統一する処理を行った。以上の処理を行ったうえで、形態素解析を実行した。

なお、以降の分析で使用される語の品詞は、考察する上で有益な知識を提供すると考えられる、「名詞」「サ変名詞」「固有名詞」「副詞可能」「未知語」「タグ」「動詞」「名詞 C」を採用した。

3.2.2 TF-IDF

先ほど行った形態素解析の結果から、単語の出現数が特に多い上位 200 語を対象に TF-IDF 値を算出した。TF-IDF とは文書に出現する単語に重み付けを行い、特徴語を抽出する分析手法である。出現頻度を表す TF(Term Frequency) と希少性を表す IDF(Inverse Document Frequency) をかけ合わせた値である。この値が大きい単語は文書の特徴語であると言える。TF-IDF 値は次式で算出する。

$$tfidf = tf \cdot idf$$

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

$$idf_i = \log \frac{|D|}{|\{d: d \ni t_i\}|}$$

$\sum_k n_{k,j}$: 文書 d_j におけるすべての単語の出現回数の和

$|D|$: 文書の総数

$\{d: d \ni t_i\}$: 単語 t_i を含む文書数

3.2.3 共起分析

労働災害の状況を特定するため、単語の出現頻度や特徴語だけでなく語と語の関係性を分析する必要がある。語と語の関係性を確率的に表す Jaccard 係数を用いて分析を行う。Jaccard 係数は次式で算出する。

$$jaccard\ index(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

4. 分析結果

4.1 形態素解析

形態素解析の結果を、筆者が事故の型ごとに5W1Hに沿って作業内容(HOW)、起因物(WHAT)、シーン(WHERE)、行動・意識(WHY)の4つに分類し上位5件を抽出した。括弧内の数字は件数を示す。

「墜落・転落」の起因物として可搬式作業台、安全帯が多く関与していることがわかる(Table2)。また、降りるや昇降といった移動する際や外す、解体といった解体作業時に労働災害が発生している。

「はさまれ・巻き込まれ」(Table3)ではシーンに分類される単語が抽出されなかった。これは、「はさまれ・巻き込まれ」は、他の労働災害と比べて発生した場所よりも起因物や作業内容の方が直接的な要因となるため、簡潔

Table2 Frequent words (falling from high position)

	HOW作業内容	WHAT起因物
1	降りる(43)	可搬式作業台(65)
2	昇降(36)	安全帯(44)
3	移動(24)	脚立(24)
4	解体(17)	荷(21)
5	外す(16)	養生(18)

	WHERE シーン	WHY 意識・行動
1	足場(57)	滑る(16)
2	階段(21)	慣れる(13)
3	開口部(20)	怠る(11)
4	作業床(19)	踏み外す(11)
5	梁(19)	崩す(9)

Table3 Frequent words (get caught in something)

	HOW作業内容	WHAT起因物
1	吊る(47)	荷(53)
2	入れる(39)	ワイヤー(31)
3	持つ(34)	台車(24)
4	移動(32)	フック(20)
5	操作(26)	パイプ(19)

	WHERE シーン	WHY 意識・行動
1	-	挟む(38)
2	-	認識(20)
3	-	経験(16)
4	-	意識(13)
5	-	怠る(13)

Table4 Frequent words (falling)

	HOW作業内容	WHAT起因物
1	降りる(21)	資材(18)
2	持つ(20)	可搬式作業台(11)
3	歩く(16)	隙間(10)
4	跨ぐ(11)	デッキ(9)
5	掛ける(8)	雨(9)

	WHERE シーン	WHY 意識・行動
1	段差(23)	滑る(22)
2	階段(18)	怠る(14)
3	足場(17)	慣れる(12)
4	通路(12)	近道行動(11)
5	梁(12)	考える(7)

な記載が求められる状況の記述において特筆されることが少ないことが影響していると考えられる。

起因物を見ると、荷やワイヤーが関連する労働災害が、持つ、移動、操作など単語同士にあまり関連が見られなかった。

「転倒」(Table4)の起因物としては資材や可搬式作業台が抽出されたが、他の事故の型と比較して件数が少ないことがわかる。ここから「転倒」は直接的な起因物が少ないと言える。また、行動要因としては近道行動が挙げられた。「転倒」が起こるシーンは「墜落・転落」が起こるシーンと類似していることがわかった。

4.2 TF-IDF 値

4.1.で行った形態素解析から抽出した頻出語上 200 件の TF-IDF 値を算出した。「墜落・転落」「はさまれ・巻き込まれ」「転倒」の TF-IDF 値をそれぞれ上位 15 件ずつ示した。「はさまれ・巻き込まれ」は他の 2 つの労働災害より比較的 TF-IDF 値が低い。ここから「はさまれ・巻き込まれ」は他の 2 つの事故と比べて特徴が薄い労働災害であることがわかる。

「墜落・転落」(Table5)では可搬式作業台や足場、安全帯、梁など高所作業特有の単語が多く抽出されている。「墜落・転落」にのみ抽出されている単語は可搬式作業台、安全帯、昇降、高い、設置の 5 語である。可搬式作業台、安全帯は高所作業に関連する単語であり、昇降や設置は仮設の昇降設備の設置に関する語であると考えられる。「はさまれ・巻き込まれ」(Table6)では荷、重量、吊る、クレーン、持つなど吊り荷作業特有の単語が抽出されている。「はさまれ・巻き込まれ」にのみ抽出されている単語は、移動、置く、持つ以外の 12 単語であった。ここから「はさまれ・巻き込まれ」は他の 2 種類の労働災害とは異

Table5 Characteristic frequent words (falling from high position)

	抽出語	TF-IDF値
1	バランス	0.0048
2	崩す	0.0047
3	可搬式作業台	0.004
4	降りる	0.0037
5	足場	0.0032
6	踏み外す	0.0031
7	安全帯	0.003
8	移動	0.0023
9	昇降	0.0023
10	床	0.0022
11	滑る	0.0022
12	高い	0.0022
13	梁	0.0022
14	設置	0.002
15	足元	0.002

Table6 Characteristic frequent words (get caught in something)

	抽出語	TF-IDF値
1	移動	0.0021
2	荷	0.002
3	位置	0.002
4	入れる	0.002
5	置く	0.0019
6	重量	0.0019
7	吊る	0.0019
8	合図	0.0019
9	クレーン	0.0018
10	添える	0.0018
11	持つ	0.0016
12	持ち上げる	0.0013
13	掛ける	0.0013
14	回転	0.0012
15	フック	0.0012

なる特徴があると言える。「はさまれ・巻き込まれ」は持つ、持ち上げる、添える、入れるという手を使う動作が多く抽出されているのに対し、他の2種類は滑る、歩行、降りる、踏み外すといった足を使う動作が抽出されてお

Table7 Characteristic frequent words(falling)

	抽出語	TF-IDF値
1	足元	0.0045
2	滑る	0.0038
3	バランス	0.0029
4	歩行	0.0028
5	移動	0.0026
6	崩す	0.0025
7	降りる	0.0025
8	持つ	0.0022
9	踏み外す	0.002
10	床	0.002
11	段差	0.0018
12	置く	0.0017
13	通路	0.0017
14	スラブ	0.0017
15	足場	0.0017

り、傷病部位にも差があると考えられる。

「転倒」(Table7)では特有の作業と思われる単語は抽出されなかったが、移動や歩行、通路、床など移動中を表す単語が多く抽出された。「転倒」にのみ抽出されている単語は歩行、段差、通路、スラブであり、床面に関する単語が特徴語として表れている。

4.3 共起分析

共起分析では、語と語の関係性を図る労働災害特徴(Fig.1,3,5)と職種を変数とした職種と語の関係性を図る職種別特徴(Fig.2,4,6)の2種類の分析を行った。Jaccard 係数の大きさによって線の太さが、出現頻度の高さによってバブルの大きさが変化し、共起グループによって異なった配色となる。

労働災害の特徴を表す共起分析(Fig.1,3,5)では、描画する共起関係は上位 60 件バブルの大きさ 100%に設定した。職種別特徴を表す共起分析(Fig.2,4,6) では職種と語の共起関係を分析するため、外部変数として職種を設定した。職種は全 49 種と多岐にわたるため、事故の型ごとに頻出する職種を 6 種から 8 種程度を特定し外部変数として加えた。外部変数である職種は四角、語は丸で描画され職種との共起線の数により語の配色が異なる。また、描画する共起関係は上位 60 件バブルの大きさは 100%に設定した。

「墜落・転落」の共起ネットワーク図(Fig.1)から「移動-足元-怠る」「不安定-姿勢-無理」「可搬式作業台-降りる-

ル-禁止」「持つ-階段-昇降」「梁-上-通路-歩行」に強い

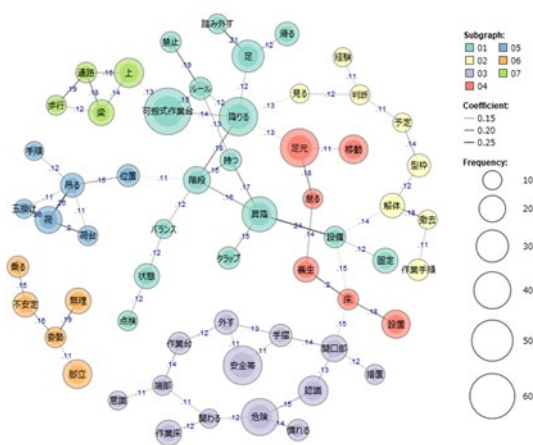


Fig.1 Cooccurrence network diagram (falling from high position)

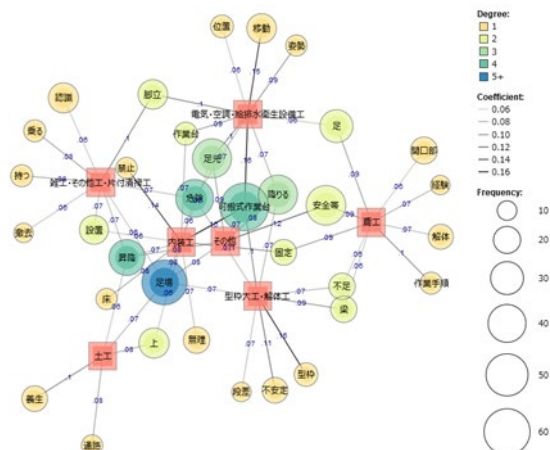


Fig.2 Cooccurrence network diagram by job type (falling from high position)

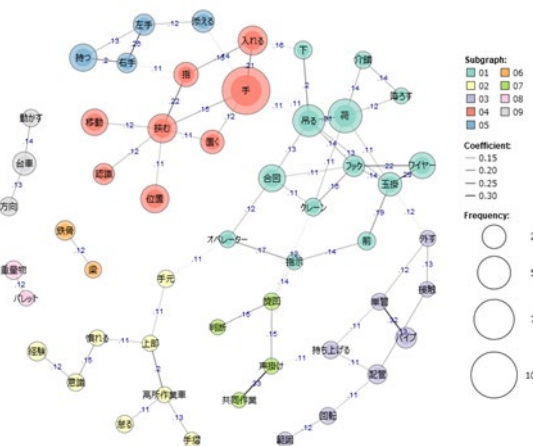


Fig.3 Cooccurrence network diagram (get caught in something)

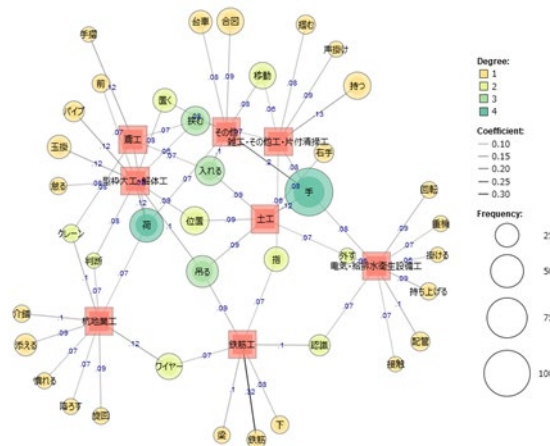


Fig.4 Cooccurrence network diagram by job type (get caught in something)

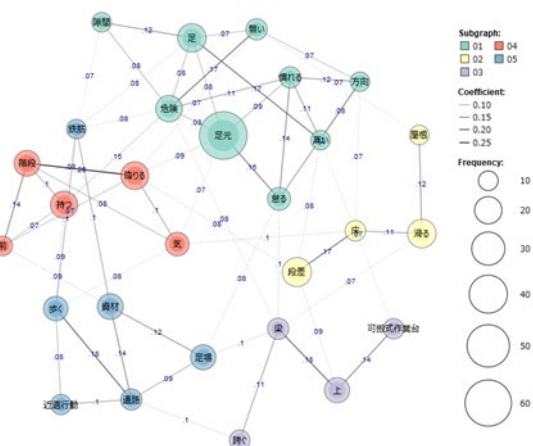


Fig.5 Cooccurrence network diagram (falling)

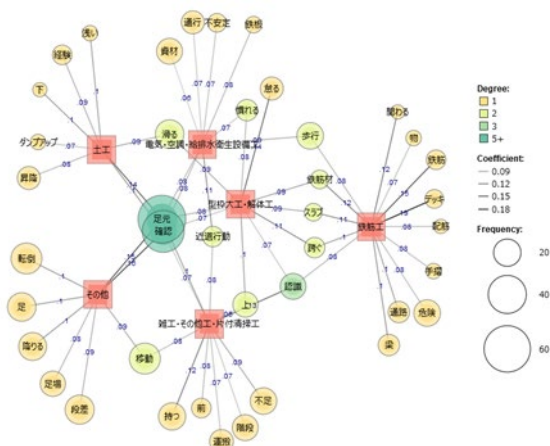


Fig.6 Cooccurrence network diagram by job type (falling)

ルール-禁止」「持つ-階段-昇降」「梁-上-通路-歩行」に強い共起関係が確認された。ここから移動時に足元の確認を怠る、不安定・無理な姿勢での作業、可搬式作業台を降りる際の禁止ルールの違反、何かを持った状態での階段の昇降、梁上の通路を歩行するなどの状況が抽出された。可搬式作業台を降りる際の禁止ルール違反という状況から可搬式作業台に背を向けて昇降しないというルールが徹底されていないことがうかがえる。また、職種別特徴を表す共起ネットワーク図(Fig.2)では内装工、設備工、型枠大工・解体工、雑工など多くの職種で可搬式作業台と共起しており可搬式作業台が起因した墜落・転落災害が多く多くの職種で発生していることがわかる。特に内装工は可搬式作業台の禁止ルール違反による労働災害が多いと示唆される。内装工、雑工、土工は昇降時が墜落・転落の要因となっている。

「はさまれ・巻き込まれ」の共起ネットワーク図(Fig.3)から「手-入れる-指-挟む」「荷-吊る-フック-玉掛」「介錯-荷-降ろす」「台車-動かす」「共同作業-声掛け」に強い共起関係が確認された。ここから何かに手や指を入れたとき、吊り荷・吊りフックを用いた玉掛作業、荷を降ろすとき、台車を動かすとき、共同作業時の声掛け不足などの状況が抽出された。職種別特徴を表す共起ネットワーク図(Fig.4)では、型枠大工・解体工、鳶工、杭地業工、鉄筋工は荷と共に共起されており、荷にはさまれる労働災害が多いことが示唆される。型枠大工・解体工、土工、鉄筋工は吊ると共起されており、物を吊っている際に労働災害に遭遇していることがわかる。杭地業工と鉄筋工はワイヤーが起因した労働災害が多く発生している。

「転倒」の共起ネットワーク図(Fig.5)から「足元-怠る」「通路-近道行動-歩く」「鉄筋-資材-通路-歩く」「床-段差-滑る」「持つ-前-階段-降りる」に強い共起関係が確認された。ここから足元確認を怠ったとき、近道行動をするとき、鉄筋などの資材が通路に放置された場所を歩くととき、床に段差があるとき、何かを持って階段を降りるとときという状況が抽出された。また、職種別特徴を表す共起ネットワーク図(Fig.6)では土工、雑工、設備工、型枠大工・解体工は足元・確認と共に共起されており足元の確認不足が要因として挙げられた。設備工、雑工は近道行動が、土工、型枠大工・解体工は足元が悪い状態で滑ることが要因としてそれぞれ共起されており、躯体工事での作業環境が転倒リスクを高めていると考えられる。

5. 結語

本研究では社内の過去 10 年分の労働災害発生データを用いてテキストマイニングを行った。労働災害の要因として構造化データから特定困難であった作業内容、起因物、シーン、行動・意識の特定を試みた。また、特定した上記の 4 種類の共起関係から、労働災害の要因とし

て頻繁に挙げられた状況を特定した。建設業の労働災害を対象としたテキストマイニングによる研究はこれまで行われてこなかったが、本研究により労働災害防止・抑制に活用可能な分析が可能であることが示唆された。

分析の結果の概観としては、労働災害の大きな要因として、「墜落・転落」では可搬式作業台の使用ルールの不徹底、手に何か持っている状態での昇降が挙げられた。「はさまれ・巻き込まれ」では、吊り荷や吊りフックを使用する玉掛け作業時の声掛け不足が挙げられた。「転倒」では、近道行動をした時や資機材の放置、雨天で床が滑ることが挙げられた。今後、これらの特定できた要因について適切な対策を施していく必要がある。

今回の分析では、テキストデータで収集・蓄積されている労働災害発生状況データを使用した。自由記述式のテキストデータのため表記揺れが多くデータ加工に時間を要した。しかし、労働災害の防止・抑制を目的とした分析は今後も継続して取り組んでいく必要があり、できるだけ容易に分析が行える仕組みが必要である。そのためには、データの収集方法を自由記述式ではなく、頻出する要因については構造化データとして蓄積していくことが有効であると考えられる。5W1H に沿って選択式でデータを蓄積することで、誤字脱字や表記揺れの防止が可能となり、データ分析のリテラシーを問わず容易に分析を継続することができるようになる。また、特定の状況で発生した労働災害件数をモニタリングすることが可能になり、施した対策の検証を行うことで有効な対策のみを積み重ねることができる。

最後に、本研究の限界として、当社が蓄積した過去 10 年分の労働災害データを用いて分析を行ったが、1 社で蓄積されているデータだけでは分析に必要な十分なサンプル数を確保できない事例も多い。建設業全体として労働災害を防止、抑制するためには業界としてデータを収集、蓄積、利活用していくことが不可欠であると考えられる。建設業の将来を考えるにあたり、業界を挙げて労働災害の要因を分析し、適切な対策を講じていく必要があると示唆される。

【参考文献】

- 1) 伊藤和也：建設業における斜面工事での墜落による労働災害の調査・分析と墜落災害防止対策の検討,土木学術論文集 F6(安全問題),Vol.71,No.1,13-24,2015.
- 2) 伊藤和也,高梨成次,堀智仁,日野泰道,吉川直孝,高橋弘樹,大幡勝利,玉手聡,豊澤康男：東日本大震災の復旧・復興工事における労働災害の発生状況に関する調査分析,土木学会論文集 F6 (安全問題), Vol.69, No.1, 32-45, 2013.
- 3) 北川広大,永崎孝之,中野聡太,肥田光正,岡将吾,和田親宗：保健衛生業における労働災害事例のテキストマイニング解析-動作の反動・無理な動作による事例を対象として-,労働安全衛生研究,2020.