

機械学習手法を用いた建設現場における 労働災害予測および事例提示方法の検討

林 祐光*¹・佐藤 貢一*¹・小島 歩*²・内田 大介*²

Keywords : work-related accidents, human factor, human error, unsafe actions, safety management, logistic regression
労働災害, ヒューマンファクター, ヒューマンエラー, 不安全行動, 安全管理, ロジスティック回帰

1. はじめに

建設現場における労働災害は 1972 年の労働安全衛生法制定以来減少傾向を示してきたが、近年は下げ止まりの様相を見せている。その理由の一つとして、人間の行動特性である錯覚・不注意・近道行為・省略行為等が作業者に内在化することで、事故災害に繋がるヒューマンエラーが生じることが挙げられる。日ごとの作業環境や作業者の熟練度が異なる建設現場において、これらの行動特性を完全に排除することは困難である。こうした人間の特性に起因する事故災害を減らしていくためには、作業員一人ひとりが作業上のリスクを理解し、効果的に注意しながら行動することが重要である。そのために元請会社による事故災害予防の啓発や、職長を中心とした危険予知 (KY) 活動等のリスクアセスメント活動が継続的に行われている。それらの活動の中でリスクアセスメントの根拠となる過去の事故災害事例情報は重要な役割を持つと考えられる。

当社では過去に社内外で発生した事故災害事例のデータベースを構築し、事例の検索システムを公開することで上記の活動をサポートしている (図-1)。事例検索システムは自然言語処理による「あいまい検索」「同義語検索」「固有名詞のマスキング」等の機能を備えており、認証基盤である X-grab^{注1)}と連携することによって当社社員だけではなく、グループ会社や協力業者も利用できるようになっている。これらのデータベースを更に活用するためには自発的に事例の検索を行うことに加え、作業員属性や作業環境、作業内容といった



図-1 事例検索画面の構成

Fig.1 Structure of case search screen

日ごとに変わる作業の状況からどのようなリスクが存在するかを過去の事故災害事例から把握し、その結果を作業員に対してプッシュ型 (あらかじめ決められたタイミングで情報を提供) で提示することが有効であると考えられるが、その方法は確立されていない。

本稿では機械学習手法を用いて、個々の作業員の属性や作業環境、作業内容から起こりやすい労働災害を予測し、さらに具体的な事故災害事例を提示する方法の検討を行った結果を報告する。

*1 技術センター 先進技術開発部 AI連携技術開発室

*2 安全本部 安全部

2. 分析方法および結果

表-1 偏回帰係数の分析結果（一部抜粋）

Table 1 Results of partial regression coefficient analysis

2.1 分析方法

特定の事故災害発生の有無や種別といった 2 値または多値の目的変数を機械学習で予測する場合、ランダムフォレストや Gradient-Boost Decision-Tree (GBDT) など決定木をベースとしたアルゴリズムが広く用いられているが、ここでは説明変数となる各要因が労働災害の発生にそれぞれどの程度影響しているかの解釈が容易な多重ロジスティック回帰分析を採用した。多重ロジスティック回帰の式は以下のように表される。

$$p_i = \frac{1}{1 + e^{-y_i}} \quad (1)$$

$$y_i = b_{i0} + b_{i1}x_1 + \dots + b_{ij}x_j + \dots + b_{in}x_n \quad (2)$$

ここで、 p_i は特定の作業者の状況において労働災害が発生した場合に事故型（労働災害の種別）が*i*である確率、 x_j と b_{ij} はそれぞれ*j*番目の説明変数と偏回帰係数である。説明変数は表-1 の左列に例示される作業者属性、作業環境、作業内容に関わる変数とする。ここでは事故型*i*の集合Cを{墜落・転落、転倒、激突、飛来・落下、崩壊・倒壊、激突され、挟まれ・まきこまれ、切れ・こすれ、動作の反動・無理な作業、その他（熱中症含）}とし、事故型の確率モデルは式(1)、式(2)に基づいて事故型*i*毎に定義される。過去の労働災害データの学習から判定の尤もらしさを示す尤度関数を最大化する全ての b_{ij} を求めることで、未知の状況における p_i を推定することができる。

2.2 データの準備

当社が保有する 2015 年～2022 年に発生した労働災害に関するデータから各変数で欠損値のないもの 7,517 件を抽出して分析した。質的な説明変数においてカテゴリ数が多い場合、各カテゴリのサンプルサイズが小さくなり有意な分析結果を得ることが難しくなる。そのため、カテゴリ数の多い質的な変数に関してはカテゴリの分布状況や違いの出やすさを考慮して値を二つのグループに分類した。例えば危険作業はカテゴリ数が多く各サンプルサイズが小さいため、「危険作業あり」と「危険作業なし」に分類し、天候は地面や路面が濡れているかを考慮して「晴・曇」と「雨・雪・雷」に分類した。二つの分類が難しい変数については、サンプルサイズが一定以上確保できるカテゴリ（例：職種における普通作業員、鳶工、型枠工など）のみダミー変数化した。

事故型 説明変数	墜落・転落 (Ref.その他)	転倒 (Ref.その他)	切れ・こすれ (Ref.その他)
年齢	① 0.0172 ** (0.0184)	① 0.0239 ** (0.0163)	-0.0098 ** (-0.0078)
曜日 月火水木 (Ref.金土日)	③ -0.1596 * (-0.0055)		③ -0.1864 * (-0.0048)
天候 晴曇 (Ref.雨雪雷)	-0.2754 * (-0.0056)	-0.3947 ** (-0.0052)	
建土区分 建築・住宅 (Ref.土木)	0.2367 * (0.0071)		0.7617 ** (0.0170)
作業場所 構築物床 (Ref.その他)	-0.9918 ** (-0.0337)	0.2848 ** (0.0062)	0.2445 ** (0.0062)
地域 東日本 (Ref.西日本)	④ -0.3931 ** (-0.0131)		
季節 夏期 (Ref.その他)	-0.6459 ** (-0.0236)	-0.8819 ** (-0.0206)	
季節 中間期 (Ref.その他)			0.3063 ** (0.0084)
職種 普通作業員 (Ref.その他)	-0.6755 ** (-0.0198)		
職種 鳶工 (Ref.その他)		-0.7223 ** (-0.0103)	-0.9663 ** (-0.0161)
作業内容 機械類使用 (Ref.その他)	-1.1695 ** (-0.0357)	-1.4757 ** (-0.0288)	② 1.7668 ** (0.04018)
作業内容 危険作業 (Ref.その他)	0.3134 ** (0.0102)	-0.3837 ** (-0.0079)	-1.3895 ** (-0.0336)
作業内容 技能的作業 (Ref.その他)		-1.0925 ** (-0.0196)	0.2368 * (0.0049)
モデル有意確率	< 0.001	< 0.001	< 0.001
正解率 (%)	85.5	90.6	88.7

※1 括弧内の数値は標準偏回帰係数

※2 偏回帰係数の有意水準 ** p<0.01, * p<0.05

2.3 事故型の決定に影響する変数の分析結果

事故型の集合Cのうち、「墜落・転落」および「転倒」、
「切れ・こすれ」の分析結果を表-1に示す。各事故型
モデルにおいて尤度比を利用したステップワイズ法で
変数選択し、影響が認められた変数について偏回帰係
数と括弧内に標準偏回帰係数^{注2)}を記載している。なお
標準偏回帰係数は Menard¹⁾による方法で計算した。対
象となる事故型に対して p_i を高くする正の影響が認め
られるものを赤色に、 p_i を低くする負の影響が認めら
れるものを青色で示している。この結果から、例えば
年齢が高くなることが「墜落・転落」や「転倒」に正
に影響する(表中①)ことや、機械類を利用する作業
が「切れ・こすれ」に正に影響する(表中②)ことなど
経験と整合できる要因を抽出できる一方、「墜落・転
落」や「切れ・こすれ」に対して曜日の影響が認めら
れる(表中③)ことや「墜落・転落」に地域性が影響
する(表中④)ことなど、経験的に意識することが難
しい影響要因をあわせて抽出できることが分かる。

表-2 作業予定者 A の状況
Table 2 Propaties of worker A

変数名	値	変数名	値
地域	東京	年齢	49 歳
建土区分	土木	経験年数	10 年以上
曜日	金曜日	入場後日数	8 日~30 日
天候	曇	機械類使用	あり
作業時間帯	午後	作業場所	ダム土捨場
時期	秋	作業種類	土工事
職種	特殊作業員	危険作業	あり

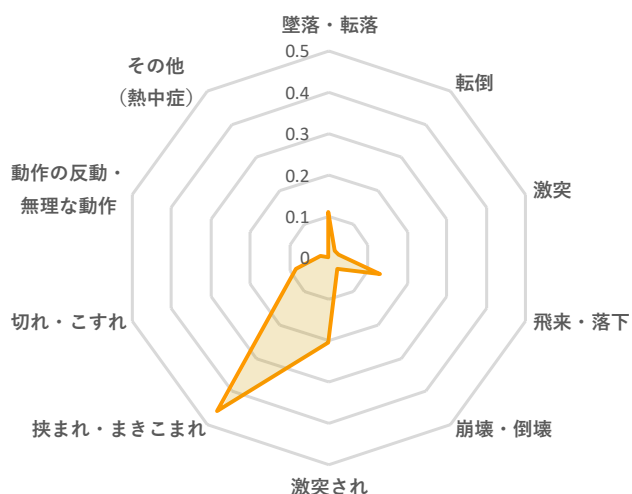


図-2 作業予定者 A に対する事故型の発生しやすさ
Fig.2 Accident types likely to occur for worker A

3. 事故型の予測および事例提示方法の検討

3.1 発生しやすい事故型の予測

各事故型に対する説明変数の偏回帰係数と、各作業
者の状況を示す説明変数の実現値を式(1)、式(2)に代入
することにより、各事故型別に定義されたモデルにお
いて労働災害が発生する場合の事故型が*i*である確率 p_i
を得ることができる。これを基に特定の状況において
起こりやすい事故型を提示することで、作業者の個別
の状況に応じた注意喚起が可能となる。例えば表-2 お
よび表-3 のような状況の作業予定者 A および B 対し
ては、それぞれの状況を示す各変数の値から、図-2、
図-3 のような各事故型の発生しやすさの分布を得る
ことができる。作業予定者 A は「挟まれ・まきこまれ」
の数値が高いことから、重機周辺での作業に特に注意
すること、作業予定者 B は「墜落・転落」に加え「挟
まれ・まきこまれ」「その他(熱中症)」の数値が高い
ことから、高所における作業のほかに重機周辺の作業

表-3 作業予定者 B の状況
Table 3 Propaties of worker B

変数名	値	変数名	値
地域	千葉	年齢	23 歳
建土区分	建築	経験年数	1 年~3 年未満
曜日	金曜日	入場後日数	31 日~90 日
天候	雨	機械類使用	なし
作業時間帯	午後	作業場所	高所作業車
時期	夏	作業種類	高所作業車使用
職種	電工	危険作業	あり

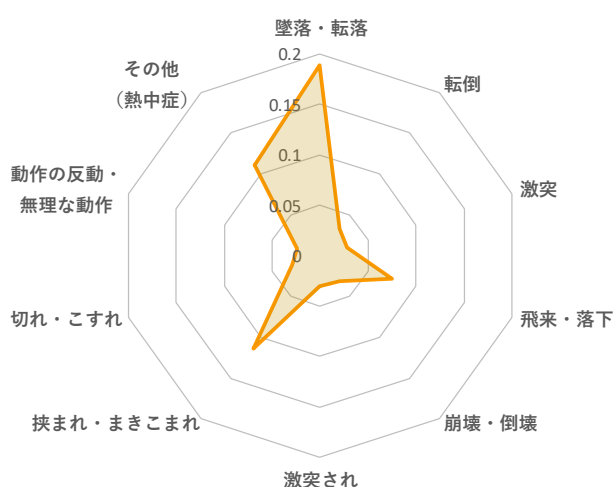


図-3 作業予定者 B に対する事故型の発生しやすさ
Fig.3 Accident types likely to occur for worker B

や水分補給等に特に注意するなど、具体的な行動に移すことができる。

3.2 過去の具体事例の提示

ある作業員に対して、上記のように起こりやすい事故型を提示することに加えて、さらに的確な事故災害事例を示すことができれば、安全対策を具体的にイメージしやすくなる。そこで過去の事故災害事例から作業員の状況で発生しやすい事例を抽出して提示する方法について検討した。各事故型モデルにおける標準偏回帰係数はその絶対値が事故型を決定するうえでの相対的な影響度を示していると言える。つまり、作業予定者の状況と過去の事例の比較において影響度の大きい説明変数が一致しているほど、その事例に近い労働災害が発生しやすい状況であると考えられる。そこで、標準偏回帰係数を利用して過去の労働災害事例を抽出して提示する方法を考案した(図-4)。まず「発生しやすい事例」として事例種別を選択し、次に抽出したい事故型を特定する。特定された事故型モデルにおいて対象となる作業予定者の説明変数の値と全ての過去事例の説明変数の値を比較し、値が一致した変数の標準偏回帰係数の絶対値の合計を「発生スコア」として事例毎に算出し、発生スコアが最も高い事例を抽出する。ここで発生スコアが最も高い事例が複数ある場合は、その中からランダムに一つの事例を抽出する。抽出された事例を作業予定者の状況において発生しやすい事例として提示する。あわせて、「意外性のある事例」を提示する方法を考案した。例えば、年齢が「墜落・転落」や「転倒」に影響することは加齢に伴う体力やバランス感覚の衰えを考慮すると自然であり意外性は小さいと考えられる。しかし、「墜落・転落」に地域性が影響することは理由が分からず意外と言える。そのような意外性のある影響変数を安全管理担当や現場管理者が事前に抽出しておく。これらの意外性のある影響変数のうち、一致した変数の標準偏回帰係数の絶対値の合計を「意外性スコア」と定義し、意外性スコアが最も高い事例を、意外性のある事例として提示する。

以上のアルゴリズムに従い事例を抽出すると、表-2の作業予定者Aの「挟まれ・まきこまれ」事故においては、発生しやすい事例として機械に指を巻き込まれた事例が、意外性のある事例としてブロックを運搬中に指を挟んだ事例が提示された。表-3の作業予定者Bの「墜落・転落」事故においては、発生しやすい事例として点検歩廊の手すりの未設置部からの墜落の事例が、意外性のある事例として雨で濡れたトラックの荷台からの転落の事例が提示された。

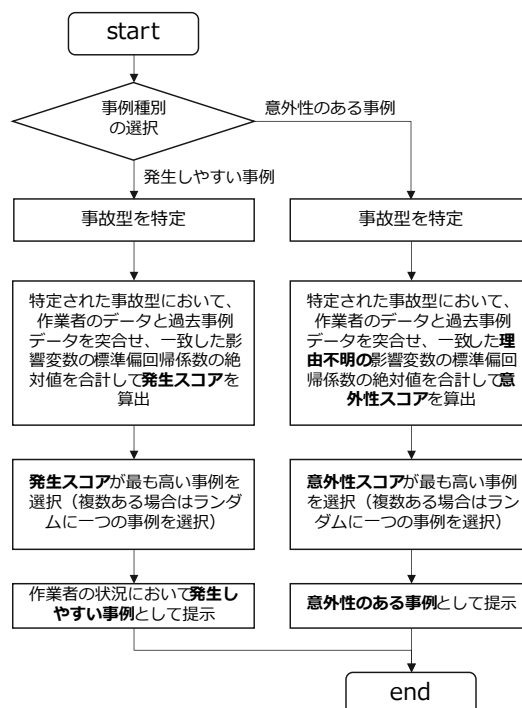


図-4 事例提示フロー

Fig.4 Flow of presenting a case study

4. まとめ

本稿では作業員属性や作業環境、作業内容といった特定の作業員の状況において、どのようなリスクが存在するかを過去の事故災害事例から把握し、その結果を作業員に対してプッシュ型で提示する手法について検討した。その結果、多重ロジスティック回帰分析を用いた発生しやすい労働災害の予測と、「発生しやすい事例」および「意外性のある事例」を多くの過去事例から抽出して提示する方法を開発できた。

今後はこれらのリスクを効果的に提示するタイミングやインターフェースの検討を行い、当社作業所への展開を図っていく予定である。

注

- 注1) X-grab : 当社が開発・運用している認証基盤、データ基盤、制御基板を備えるデータプラットフォーム
 注2) 標準偏回帰係数 : あるモデルにおいて目的変数に対する説明変数の影響度を標準化し説明変数間で比較可能にしたもの

参考文献

- 1) Menard, S.: Applied Logistic Regression Analysis (2nd ed.). SAGE Publications, Inc., pp.50-56, 2002