

## 汎用言語モデルを用いた労働災害に対するリスクアセスメント支援システムの開発

大阪大学大学院工学研究科 学生員 ○石川 大智

大阪大学大学院工学研究科 正会員 貝戸 清之

HAMADA 建設マネジメント研究所 正会員 浜田 成一

### 1. はじめに

国内の建設労働災害の発生状況は、長期的には減少傾向で推移してきていたが、近年は下げ止まり傾向あるいは若干の増加傾向が見られている。建設業の安全性の確保のためには、下げ止まり傾向の原因の分析とその改善策の提案が急務である。このような建設業の状況下において、労災が多発した時代を経験した熟練技術者が定年を迎えることにより、安全技術の継承や安全教育の実施が不十分となりうるものが労災抑制に対する課題となっている。この課題に対する一対策として、国内外問わず4M分析と呼ばれるリスクアセスメントがなされている。4M分析では、労災事例における要因分析を人的要因 (Man)、機械要因 (Machine)、環境要因 (Media)、管理要因 (Management) の4つの視点で行うことにより、労災抑制に対する本質的対策を行うことが目的とされる。しかしながら、4M分析を行うには熟練技術者の経験や暗黙知が不可欠であり、そのような経験や暗黙知を伝承していかなければ4M分析を行っていくことは難しくなる。

本研究では上記の課題に対する一対策として、汎用言語モデルであるBERT<sup>1)</sup>を用いて、過去に4M分析を行った類似事故事例の推薦を行うとともにAIによる参考値の出力を実装することにより、4M分析支援システムを構築し、知識の伝承を行うことを提案する。

### 2. 本研究の基本的な考え方

#### (1) 本研究における4M分析

4M分析は、従来より製造業における品質管理、労働における安全管理など多岐にわたる分野で行われており、様々な定義のものが存在する。本研究における4Mとは、安全工学で用いられるMan, Machine, Media, Managementを指す。また、4M分析の方法としては、一つの事故事例に対し各要因に該当する場合は✓を付すという最も単純な方法で行うものとした。

#### (2) BERT

キーワード：労働災害、4M分析、深層学習、リスクアセスメント、BERT

連絡先：〒565-0871 大阪府吹田市山田丘2-1 工学研究科AR棟605号室 TEL：06-6879-7622

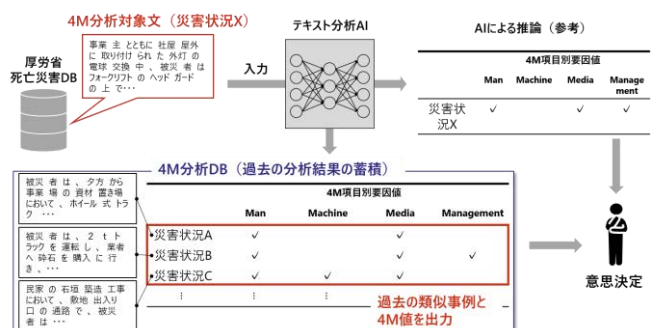


図-1 4M分析支援システムの概要

BERTは、TransformerをベースにMulti-Head Attentionを導入することにより、RNNやCNNを用いずとも文中の長距離の依存関係をとらえることが可能なニューラルネットワークである。大規模なコーパスを用いて事前学習を行い、各タスクに応じてfine-tuningを行うことにより、比較的小規模のデータ・計算資源で高い精度を出すことが可能である。BERTでは、文を入力することで、対応した文のベクトル表現が出力され、それらを用いて文章・単語単位の様々なタスクが行われる。

### 3. 提案手法

図-1に、本研究で構築する4M分析支援システムの概要を示す。国内の労災事例は、厚生労働省の死亡災害DBに毎年集計されている。そこから建設技術者が新たな事例の4M分析を行う場合に、まず対象となる事例をBERTに入力することを想定する。その結果、出力として、過去に熟練技術者が4M分析を行った類似事例を推薦するとともに、対象の事例に対してAIによる推論結果(参考値)を出力するようなシステムを考える。これにより、建設技術者はこれらの結果を参照しつつ、新たな事例に対する4M分析を行うことができる。以下、BERTによるシステム全体の処理を述べる。

はじめに、入力文章を形態素解析器であるMeCabを用いて分割し、単語列を得る。これに対し、BERTの入力に必要な”[CLS]”トークンを文頭に、”[SEP]”ト

クンを文尾に付加する。作成した単語列を事前学習済み BERT に入力することで得られる出力値の平均を計算し、これを文のベクトル表現とする。推薦システムではこのベクトルを用いて、コサイン類似度を計算することで対象事例に類似した過去の事例を抽出する。

次に、AI による参考値出力の過程を述べる。本研究では、事前学習済み BERT の最終層に、活性化関数をシグモイド関数とした全結合層を 1 層追加することでマルチラベル分類に対応させ、fine-tuning を行う。これにより、上記で述べた文のベクトル表現を各 4M ラベルが ✓ である確率を要素にもつ 4 次元ベクトルに変換できる。日本語の事前学習済み BERT として、本研究では東北大版 BERT を用いた。東北大版 BERT では、日本語 Wikipedia をコーパスとして事前学習が行われている。また、tokenizer としては McCab 及び NEologd を使用しており、WordPiece アルゴリズムによりサブワード化を行っている。損失関数には binary cross entropy loss を用い、評価指標は分類問題で使用される precision, recall, f1-score に対しそれぞれ micro 平均を用いた。

#### 4. 実証分析

提案手法を実際の建設労災事例に適用した。データについては、厚労省死亡災害 DB から抽出した 2015 年から 2020 年までの建設業の労災文書を用いた。これらに対し、建設業における勤務経験のある熟練技術者により、各事例に対し 4M 項目別のラベル付けを行ってもらい、それらを用いて提案手法の検討を行った。全事例のうち、文章から事故の状況がはっきりと読み取れないものは除外し、前処理として、句点の付加や括弧記号・空白の削除、数字等の不要な情報の除去を行った。上記の処理を行った結果、事例数は計 567 個となった。その後、全データを訓練データとテストデータに 9:1 で分割を行った。学習におけるエポック数は 60、バッチサイズは 64 とした。重みの最適化手法には Adam を用い、ハイパーパラメータの値は学習率  $lr = 0.001$ ,  $\beta_1 = 0.999$ ,  $\beta_2 = 0.9$  を用いた。モデルの学習時、BERT の fine-tuning を最終層のみに行った。

図-2、図-3 に、それぞれ得られた文書ベクトル空間、推薦システムによる結果の例を示す。毎年頻発している溝内の土砂崩落に対して、類似した事例が出力されていることが確認される。また、参考値出力システムの精度として、テストデータに対する precision, recall, f1-

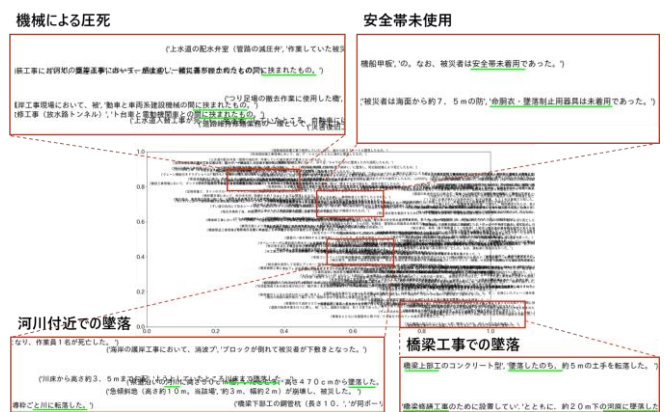


図-2 文書ベクトル空間 (t-SNE により次元圧縮)

テストデータ中の対象文 (モデルへの入力) : 溝内の土砂崩落事故 (2018年事故事例)

市発注の下水道工事において、民家へ引き込むための取出し管周辺の掘削を行っていたところ、手掘りをするため掘削溝内に入った際、突然、**砕石・土砂が崩壊**し、被災者の首付近に直撃した。崩壊した砕石・土砂は、**砕石**であった。

※ 類似度はコサイン類似度で計測

モデルからの出力 (入力文に対し、訓練データ中の類似度top3)

1st.: 2017年事故事例, コサイン類似度 0.970

$$\text{sim}(v_a, v_b) = \frac{v_a \cdot v_b}{\|v_a\| \|v_b\|}$$

宅地造成工事現場において、マンホール・管の設置作業中、型枠にコンクリートを流し入れたところ、型枠が開き、壊れる危険が生じた。そのため、被災者の副責任者の2名で型枠の補強を行うべく、**掘削した箇所**に進入したところ、**側面の土砂が崩壊**し被災者に降り掛かり生き埋め状態となり、翌日に死亡した。

2nd.: 2017年事故事例, コサイン類似度 0.968

宅地造成工事現場の下水管敷設工事において、被災者2名の共同作業で、溝掘削内の西側側壁に下水枝管用の横穴をブローカー等で掘っていたところ、**東側側壁が崩壊**し、被災者2名が土砂に埋まった。

3rd.: 2019年事故事例, コサイン類似度 0.967

汚水管を敷設するために、地山の掘削、汚水管の掘え付け、埋設等、一連の作業を3人1組で行っていた。汚水管を2本目まで敷設し、地上のマンホール周辺を地固めしていたところ、被災者が**掘削面に立ち入った瞬間**にボックスカルバート付近の**土砂が弱層崩壊**し、土砂に埋もれ死亡に至った。

図-3 類似事故事例の推薦例

score は0.754, 0.748, 0.751となった。この結果は高い性能であるとはいえないが、以下の事柄に起因すると考えられる。第一に、建設労災文書は、文書ごとの長さや説明の詳細度合いが大きく異なり、表記ゆれ (同一な概念を指す複数の表記) が多分に存在することがあげられる。そのため、精度の改善には事故に無関係な事柄を除去し、表記を統一することが考えられる。第二に、専門用語が多く、事例描写に非文法的かつ断片的な情報が含まれていることがあげられる。これに対しては、土木分野に特化したコーパスで事前学習済み BERT を追加学習することで精度を改善できる可能性がある。

#### 5. おわりに

本研究では、熟練技術者の知識の伝承を目的とした 4M 分析支援システム構築のため、汎用言語モデルである BERT により 4M 分析結果の学習を行い、その精度を検証した。今後は土木分野に特化したコーパスを用いて BERT を追加学習することを検討する必要がある。

#### 【参考文献】

- 1) Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.:BERT: Pre-training of deep bidirectional transformers for language understanding, *arXiv:1810.04805*, 2018.