

令和 4 年 5 月 26 日現在

機関番号：13901

研究種目：研究活動スタート支援

研究期間：2020～2021

課題番号：20K22185

研究課題名（和文）機械学習を活用した化学物質リスクアセスメントのプログラム構築

研究課題名（英文）Building a Program for Chemical Risk Assessment using Machine Learning

研究代表者

原田 敬章（Harada, Takaaki）

名古屋大学・環境安全衛生推進本部・准教授

研究者番号：80888925

交付決定額（研究期間全体）：（直接経費） 1,600,000円

研究成果の概要（和文）：研究教育活動に伴う実験事故を防ぐため、安全教育の一環としてリスクアセスメントが注目されている。リスクアセスメントとは、健康障害や事故を防止するため、取り扱う化学物質や実験操作に伴うリスクを事前に見積もり、そのリスクの低減対策を評価することである。しかし、多様な化学物質や実験室での作業に対しリスクアセスメントを実施することは多くの時間を費やしてしまう。本研究では、深層学習を用いて迅速にリスクアセスメントを行うプログラムの構築を目的とした。過去の実験事故のデータを解析してモデルを作成し、入力した実験室での作業に伴うリスクを予測することが可能となった。

研究成果の学術的意義や社会的意義

本研究で構築したプログラムに実験室での作業を入力するだけで、予測されるリスクを知ることができるため、リスク低減対策の実施や実験室での作業の見直しに時間を注ぐことができ、事故防止につながると考えられる。また本研究では大学等で発生した実験事故データを使用した。化学工業等ほかの分野の事故データを同じ深層学習のモデルで解析することで、その分野専用のリスクアセスメントへの活用も期待できる。英語でプログラムを構築したため、入出力時にのみ翻訳すれば、国内外でも幅広く活用される可能性がある。

研究成果の概要（英文）：Risk assessment is to estimate the risks associated with handling chemicals and laboratory work in advance then evaluate the measures to reduce the risks in order to prevent health problems and accidents. Risk assessment has drawn much attention as part of safety training to prevent laboratory accidents during research and educational activities. However, risk assessment on various chemicals and laboratory work is a time-consuming process. This study aims to generate a program for quick risk assessment using deep learning. A model has been generated by analyzing past laboratory accidents, which has the ability to predict risks in the input of laboratory work.

研究分野：環境安全学

キーワード：安全教育 リスクアセスメント 化学物質管理 機械学習 深層学習 事故分析 実験安全

1. 研究開始当初の背景

大学等の実験室では、引火性をもつ危険物や人体に有害な毒物・劇物を取り扱うだけでなく、危険有害性が明らかでない新規化合物の合成も行っているため、潜在的な危険が高く、化学物質による火災や薬傷などの事故が多発している[山本, 2015]。化学物質による事故を防止するため、安全教育の1つとして化学物質のリスクアセスメントが注目されている。化学物質のリスクアセスメントとは、化学物質のもつ危険有害性をすべて特定し、それを取り扱う作業の危険(リスク)を見積もり、そのリスクの低減対策を評価することである。しかし、既存のリスクアセスメントの手法では、大学等の実験室で取り扱われる多様な化学物質のリスクをすべて特定できず、またその実施には非常に多くの時間を費やしてしまう問題がある。実験室にある多様な化学物質それぞれの危険有害性の調査だけでも時間を要する。実験で合成した新規化合物の危険有害性は明らかではないため、そのリスクを正確に特定できない。加えて、実験室では化学物質の種類や量など条件を頻繁に変える合成等の実験作業だけでなく、器具の洗浄や廃棄物処理でも化学物質を使うため、実験中以外の作業でも事故が発生している[山本, 2015]。

このように実験室で行う各作業について化学物質のリスクアセスメントを実施することは多くの時間が必要となり、本来の研究教育活動を阻害してしまう。よって、非定常作業で多様な化学物質が取り扱われている実験室においては、潜在的なリスクは高いものの、従来のリスクアセスメントの手法では、すべてのリスクを特定できず、また多くの時間を費やすため、網羅的かつ効率的な手法が強く求められている。

2. 研究の目的

本研究では、深層学習を用いて迅速に化学物質のリスクアセスメントを行うプログラムを構築することを計画した。深層学習は近年非常に注目されている機械学習の一種であり、これまでに病院内の患者の転倒・転落事故のリスクアセスメントにも使われている[難波, 2019]。本研究では、実験室での作業に伴うリスクの予測と化学物質の危険有害性の予測を出力するモデルの構築を目的とした。

3. 研究の方法

実験室での作業に伴うリスクの予測には、大学で実際に発生した実験事故の概要を用いた。国立七大学(北海道大学、東北大学、東京大学、名古屋大学、京都大学、大阪大学、九州大学)では、これまでに1万件を超える大学で発生した事故情報を共有しており、そこから実験事故の概要約2000件を抽出した。各実験事故の概要は体裁を整え、機械翻訳で英語化した。深層学習の精度向上のために、豊富な英語の言語モデルの使用と過去の研究手法と同様に適切なデータ拡張を行った[Wei & Zou, 2019]。実験事故の概要の文章を同意語に置き換えることで、約5000件相当のデータに拡張した。実験事故の概要から読み取れる原因に対し、火災の危険や切傷注意などの21種類のリスクから、最も適切と思われる1次及び2次リスクを手動でラベリングした。深層学習による予測モデルは、自然言語処理として高く評価されているBERT(Transformerによる双方向のエンコード表現)及びBiLSTM(双方向長短期記憶)を用いた。BERT、BiLSTM、BERT-BiLSTMの3つの深層学習のモデルそれぞれについて、実験事故の概要を学習させ、入力した実験室での作業で予測される1次及び2次リスクを比較した。

化学物質の危険有害性を予測するため、GHS区分と引火点等の性質の既知情報を用いた。独立行政法人製品評価技術基盤機構が提供する「政府によるGHS分類結果」を利用し、化学物質の危険有害性を示すGHS区分の情報を得た。また、アメリカ国立衛生研究所(NIH)の下国立生物工学情報センター(NCBI)によって維持管理されている化学分子データベース「PubChem」から、GHS区分、引火点等の物理的及び化学的性質等の情報を得た。

深層学習では大量のデータを解析するが、一般的なコンピューターでは研究の進捗を妨げるほど多くの時間がかかってしまうため、比較的安価にデータ解析を高速で行うことを可能にするGPU(NVIDIA GeForce RTX 3090等)を搭載したコンピューターを本研究で用いた。深層学習のモデル構築には、MathWorks社のMATLABと必要なツールボックスを用いた。

4. 研究成果

本研究の主な成果は、過去に発生した実際の実験事故の概要を解析して作成した深層学習のモデルを用いて、入力した実験室での作業に伴うリスクを予測することが可能となったことである。抽出した実験事故の概要約 2000 件に頻出する名詞を図 1 に示す。文字が大きいほど、実験事故の概要に頻繁に使われていることを表している。実験中の事故が多く、特にガラス事故が頻発していた。また、学生による事故だけでなく、洗浄など実験に直接関係ない作業中の事故も多く発生していることがわかった。各実験事故の概要にラベリングした 1 次及び 2 次リスクは、実験室での作業に伴うリスクを予測する時の出力となるため、火災の危険などの危険（ハザード）だけでなく、切傷注意などの注意事項も含めることで、学生等でも気づきやすいような表現を用いた。

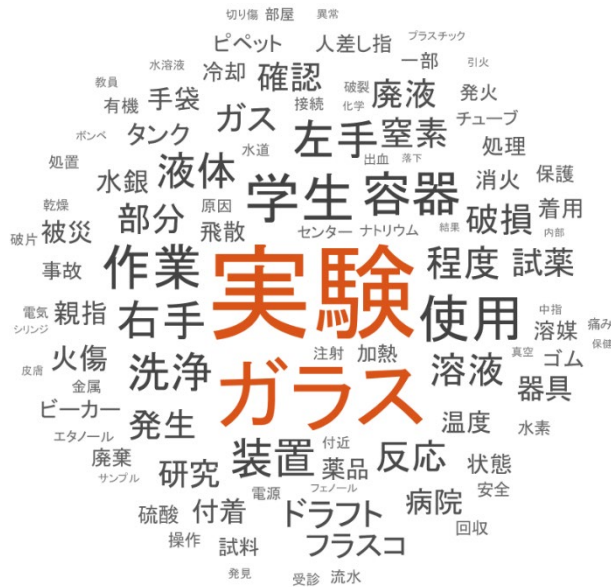


図 1：実験事故約 2000 件に頻出する名詞

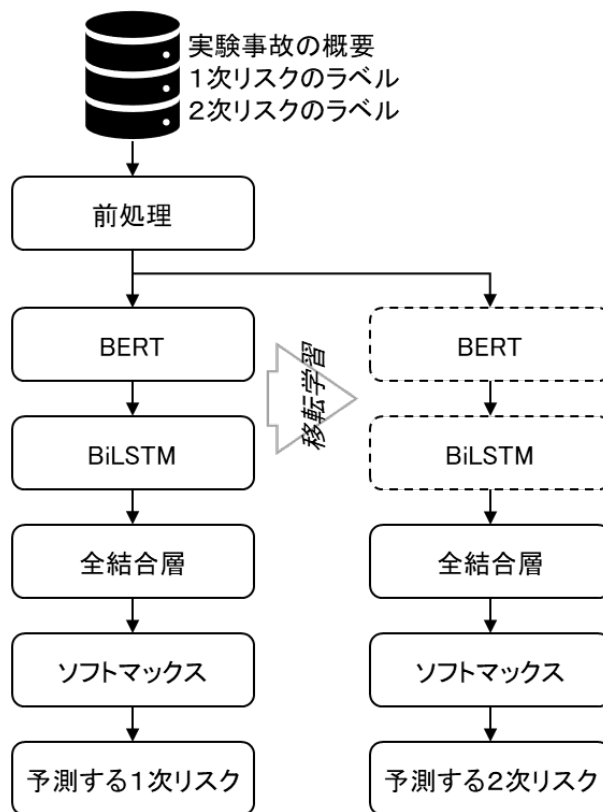


図 2：BERT-BiLSTM モデルのアーキテクチャの概略図

深層学習を用いて過去の実験事故の概要を解析し、実験室での作業に伴うリスクを予測する BERT-BiLSTM モデルを構築した (図 2)。深層学習で解析する前に、実験事故の概要を前処理する必要がある。実験事故の概要は過去形で記載されているため、深層学習で解析する前に、時制の削除等の前処理を行うことで、現在形や未来形で書かれる実験室での作業に伴うリスクを予測できるよう工夫した。次に、1 次及び 2 次リスクでラベリングした実験事故の概要を深層学習で解析した。自然言語処理で高く評価されている言語モデル BERT を用いて、実験事故の概要の特徴を抽出し、BiLSTM によって文脈等を学習することで、1 次及び 2 次リスクを予測する BERT-BiLSTM モデルを構築した。予測モデルの評価指標の 1 つである F 値は、1 次及び 2 次リスクに対しておおよそ 8 割と高い精度となった。比較のため、同様に構築した BERT のみのモデル及び BiLSTM のみのモデルも同程度の精度であった。また、実際に学生等が使用することを想定し、例として実験室での作業を入力した際の出力について考察した。例えば、「組織のスライドを作成する」という作業の入力に対し、BERT-BiLSTM モデルは、メス等の刃物が文中になくとも、「切傷注意」を 1 次リスクと予測し、出力した。他の実験室での作業例について予測したリスクについても、BERT 又は BiLSTM のみのモデルの出力よりも、BERT-BiLSTM モデルの方が適切な 1 次及び 2 次リスクを予測していると考えられた。よって、深層学習で予測される実験室での作業に伴うリスクを学生等が事前に迅速に知ることができるようになり、安全教育としての効果が考えられるだけでなく、リスクアセスメントや作業の見直しを行うことで、実験事故の低減にもつながると考えられる。

実験室での作業に伴うリスクの予測モデルについては、上記の成果をまとめ、国際論文 (査読有) や学会等で発表する。今後は、実験事故の概要を追加することでデータ数を増やし、さらなる精度の向上が期待できる。優れた日本語の言語モデルを用いることができれば、日本語で実験室での作業を入力し、同様の出力を得ることも可能になる。また、本研究で得られた予測モデルはプロトタイプであり、現状 MATLAB 上でしか動作しない。今後は、アプリ化したベータ版の作成とアンケート調査等を検討する。学生や経験の浅いポスドク等にアプリ化した予測モデルを活用してもらい、アンケート調査等でフィードバックを得ることで、予測モデルの修正やリスクの表現方法の改善等を行うことができ、安全教育として効果の高いものに仕上がっていくと考えられる。また、収集した化学物質の危険有害性を示す GHS 区分、物理的及び化学的性質等の情報については、今後も解析を継続し、データ整理、モデルの修正等、予測精度の向上に努め、学術論文等でその成果を発表していく。化学物質の危険有害性や物理的及び化学的性質等の情報も加味することで、実験室での作業に伴うリスクの予測は、より実用的なものに発展していくと考えられる。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

深層学習を用いたモデルによる実験室での作業に伴うリスクの予測について、国際論文（査読あり）として本研究の成果をまとめ、投稿した。現在審査中である。

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------