日常的な行動を表す文からのリスク文の検索と生成

川原　敬史†　　湯本　高行††　　大島　裕明††,†

† 兵庫県立大学 大学院応用情報科学研究科　〒 650–0047 兵庫県神戸市中央区港島南町 7–1–28
†† 兵庫県立大学 大学院情報科学研究科　〒 651–2197 兵庫県神戸市西区学園西町 8-2-1
E-mail: †{aa20m503,ohshima}@ai.u-hyogo.ac.jp, ††yumoto@sis.u-hyogo.ac.jp

あらまし　本研究では，日常的な行動を入力として，そこから考えられるリスクを予測する手法を提案する．入出力として扱う行動とリスクは，その内容が自然言語で記述された 1 文程度のテキストを想定する．これらのテキストのことを，行動文，リスク文と呼ぶ．研究で取り組む問題は，行動文がクエリとして与えられた時に，それに適合したリスク文を出力する問題である．本研究では，上記の問題に対して，検索アプローチと生成アプローチの 2 種類のアプローチを提案する．検索アプローチは，行動文が入力された時に，過去に蓄積された既存のリスク文に対してランキングを行い，入力と適合するリスク文を取得するアプローチである．生成アプローチは，言語生成モデルが生成した複数のリスク文に対して，検索アプローチと同様のランキングを行い，リスク文を取得するアプローチである．本研究では，リスク文のランキングや，生成の際に，BERT や，GPT-2，BERT2BERT といった言語モデルをファインチューニングする手法を提案する．ファインチューニングには，ウェブ上のサービスから収集した大量の事故情報データを使用する．各アプローチによって出力されたリスク文が，入力した行動文の内容から実際に考えられるかどうかを人手により判断し，評価を行う．評価指標 nDCG@5 を用いた評価では，検索アプローチよりも，生成アプローチの方が，入力した行動に関係するリスクを予測できる結果となった．

キーワード　自然言語処理，情報検索，ディープラーニング

1　はじめに

日常生活において何らかの行動をとると，それが原因で事故が発生することがある．例えば,「自転車に乗る」行動からは「不意に段差に乗り上げて転倒する」事故の発生が考えられる．従って，我々の普段の行動には，常に事故に繋がるリスクが含有されているといえる．事故に備えるには，日頃からリスクを認識し，事前に対策することが好ましい．そこで，本研究では，人の行動が入力された時に，そこから考えられるリスクを提示する手法を提案する．図 1 には，入力と出力の例を示す．

本研究における行動やリスクは，その内容が自然言語で記述されたものを想定し，以後はこれらを行動文，リスク文と呼ぶ．行動文は「自転車に乗って出掛けた」や,「体育館でバスケットボールをした」といった日常生活における行動が記述されたテキストである．リスク文は「段差に躓き転倒」や「ボールが破裂し，怪我」といった，行動の結果発生した想定外の事柄や，患った傷病について記述されたテキストである．本研究で取り組む問題は，行動文が入力された時に，行動文から懸念されるリスクが記載された，リスク文を出力する問題である．行動文を q，リスク文を d とすると本研究で取り組む問題は以下のように定義できる．

入力　q

出力　$d_1, d_2, d_3, \cdots, d_k$

行動文 q から考えられるリスクを表すリスク文 $d_i (1 \leqq i \leqq k)$ が出力である．

近年では，日常生活中の事故に関するデータ（事故情報データ）がウェブ上に多く蓄積されている．事故情報データの中には，発生した事故の概要が記されたテキストが含まれている場合が多い．まず，本研究では，ウェブ上のサービスから大量の事故情報データを収集し，研究に使用するデータセットを構築した．次に，データセット中の事故の概要について記載されたテキストから行動文とリスク文を抽出することで，行動文とリスク文のペアデータを作成した．

本研究では，行動文からリスク文を出力する問題に対して，以下に示す 2 種類のアプローチを提案する．

- **検索アプローチ**
- **生成アプローチ**

図 2 には各アプローチの概要を示す．検索アプローチは，データセット内のリスク文に対して，入力との適合度が高い順にランキングを行い，その結果を出力するアプローチである．適合度とは，行動文とリスク文の内容が適合する度合いのことである．生成アプローチでは，言語生成モデルが生成した複数のリスク文に対して，同様のランキングを行い，結果を出力するアプローチである．

リスク文のランキングを行う際には，行動文とリスク文の適合度を算出する手法が必要となる．本研究では，適合度を算出する手法として，汎用言語モデル BERT [2] を用いた手法を提案する．手法では，まず，行動文とリスク文のペアに対して，それらが適合するか否かを 2 値分類するタスクで BERT をファインチューニングする．ファインチューニングされた BERT を用いて，行動文とリスク文の適合度を計算する．

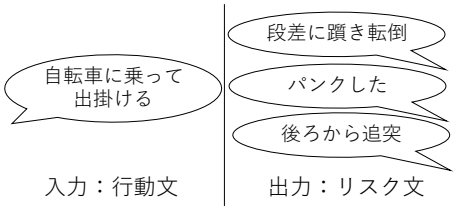

図 1　本研究の取り組み

リスク文を生成する手法としては，言語生成モデルである GPT-2 [5] や BERT2BERT [6] を用いた手法を提案する．手法の概要としては，行動文を入力とし，リスク文を生成するタスクで言語生成モデルをファインチューニングする．ファインチューニングされたモデルを用いて，リスク文の生成を行う．

評価では，出力されたリスク文の内容が，入力した行動文の内容から実際に考えられるかを人手により評価する．評価指標には nDCG を用いる．

続く 2 節では関連研究について記述する．その後，3 節で本研究で使用する事故情報データについて説明する．4 節では提案アプローチの詳細について記述する．

2　関 連 研 究

2.1　過去のリスク分析に関する研究

過去に発生した事故の傾向や，原因の分析は幅広い分野で取り組まれている [10], [11], [13], [14], [17]．事故に対する分析から得られた知見は安全教育や，今後の事故予防を考える時に役に立つことが期待されている．

山本ら [10] は総務省消防庁が管理する「消防ヒヤリハットデータベース」で公開される救急業務に関連する事故事例の分析に取り組んでいる．その中での，事故の背景要因に関する分析結果では「危険情報を把握・予見できなかった」が全体の事故事例の中で多くの割合を示している．このことから，事故に関する知識の不足は事故発生の要因となることが考えられる．

森泉ら [12] の研究では過去の事故事例から見受けられた知見を活かして，今後の安全教育への応用に取り組んでいた．森泉ら [12] は「急ぎ」や「焦り」が事故の心理的要因として考えられることに着目し,「急ぎ」や「焦り」を疑似的に体験できる「エラー体験システム」を用いた安全教育を提案している．その結果「エラー体験システム」の利用者は，日々の安全意識が向上することが示唆された．このことから，過去に発生した事故の状況の疑似的な体験は，人の安全意識を向上に繋がる効果が期待できる．

本研究で提案する，行動からの事故リスクの予測が可能になれば，実際に行動を起こす前に，懸念されるリスクを把握できるようになる．上記の関連研究から，本研究における取り組みは，事故に関する知識の不足を解消し，実際の事故発生を予防する効果が期待できる．

2.2　言語モデルとしての Transformer

Transformer は Vaswani [7] らによって提案された機械翻訳のための Encoder-Decoder モデルである．このモデルは，RNN や CNN といったそれまでの機械翻訳に使われていたモデルよりも良い性能を示した．Transformer の大きな特徴としては，内部の Attention 機構により，テキスト中の文脈を考慮できる点である．Transformer の Attention 機構を機械翻訳以外のタスクにも応用できるように派生したモデルとして，BERT [2] や GPT-2 [5] がある．

BERT は，2018 年に Google の Devlin ら [2] によって提案されたモデルである．これは，Transformer の Encoder 層が複数積まれたようなモデル構造を持つ．これを，Wikipedia などの大規模なコーパスを用いて事前学習することで，汎用的な言語モデルとなる．テキストはトークンレベルに分割されたトークン列として BERT に入力する必要がある．上記のような分割のことを以後はトークン化と呼ぶ．事前学習時のタスクとしては Masked LM（MLM）タスクと Next Sentence Prediction（NSP）タスクが一般的である．MLM タスクでは，入力されるトークン列のうち一部のトークンが [MASK] トークンで隠されている．周囲の文脈から，[MASK] で隠されたトークンを予測するタスクが MLM タスクである．NSP タスクは，2 つのテキストが入力され，それらが連続するテキストかどうかを予測するタスクである．さらに BERT では，個別の言語処理タスクを行う層を追加し，ファインチューニングを行うことで，様々な言語処理タスクに対応させることが可能である．解かせるタスクの例としては，分類タスクや系列ラベリングタスクなどが挙げられる．近年では日本語を事前学習させたモデルが公開され始めている．このような取り組みから，日本語における自然言語処理タスクにおいても BERT を活用する研究が増えてきている [9], [15]．

GPT-2 は Radford ら [5] によって提案されたモデルである．BERT は Transformer の Encoder 層が複数積まれていたのに対して GPT-2 は Decoder 層が複数積まれたモデル構造を持つ．GPT-2 は主に言語生成に特化したモデルである．Wikipedia などの大規模なコーパスで事前学習されることで汎用的な言語生成モデルとなる．こちらも目的に応じたデータを用いてファインチューニングすることで，目的とする返答文の生成が可能になる．

近年では，ファインチューニングモデルの性能をより高めるためには，ファインチューニングの前段階で，追加の事前学習を行うことが有効だと示唆されている [3], [16]．追加の事前学習とは，ファインチューニングで使用するテキストや，それに関連したドメインのテキストを用いて，事前学習と同じタスクで学習を行うことである．本研究では，この追加の事前学習のことを追加学習と呼ぶ．

3　事故情報データと前処理

本節では，まず，本研究で使用する事故情報データの概要について記述する．その後，それらの事故情報データを収集した方法や，前処理の詳細について説明する．

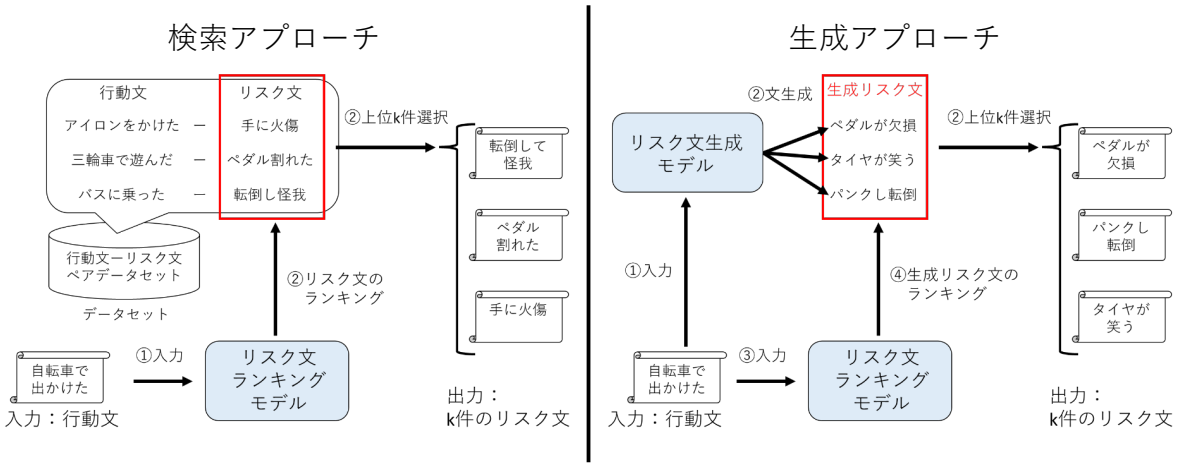

図 2　提案アプローチ

3.1　研究に使用する事故情報データの概要

本研究で使用する事故情報データは 1 件につき複数の要素で構成されている．その基本的な構成を表 1 に示す．事故情報データ中の要素の一つである事故説明文は，事故発生時における，人の行動や，発生した被害の内容について自然言語で記載されたテキストである．本研究では，事故情報データ中の事故説明文を主として取り扱う．

本研究では事故情報データを「事故情報データバンクシステム [1]」(以降，事故情報データバンクと呼称する）と「職場のあんぜんサイト [2]」の 2 つのウェブ上のサービスから取得した．「事故情報データバンク」由来の事故情報データを，本研究では主として使用し,「職場のあんぜんサイト」から取得したものは評価時にのみ使用する．

ウェブ上のサービスから取得した事故情報データには，行動文やリスク文は含まれていない．事故情報データバンクから取得したデータに対しては，3.3 節で後述する前処理によって，事故説明文から行動文とリスク文を抽出した．抽出はルールベースで行ったため，行動文とリスク文が抽出されないケースもあった．抽出できなかった事故情報データは行動リスク未抽出データとして扱う．行動リスク未抽出データは，言語モデルに対する追加学習時にのみ使用する.「職場のあんぜんサイト」から取得した事故情報データに対しては，人手により，行動文とリスク文を抽出した．

上記のようにして取得した事故情報データは，訓練データ，検証データ，テストデータに分割して使用する．実験に用いる各データの分布を表 2 にまとめる．

本研究における提案アプローチでは，取得した事故情報データ内のリスク文に対して検索を行う．検索対象となるリスク文は訓練データ内のリスク文のみとした．従って，訓練データ中のリスク文集合を D とすると，提案アプローチで検索の対象となるリスク文は以下のように定式化できる．

$$D = \{d_i \mid i = 1, 2, \cdots, 26714\}$$

d_i は 1 件あたりのリスク文を表す．$|D|$ は 26,714 である．また，訓練データ中のリスク文には必ずそれに対応した行動文がある．対応する行動文は $action(d_i)$ として定義する．

3.2　ウェブ上のサービスからの事故情報データの取得

「事故情報データバンクシステム」は消費者庁と国民生活センターにより管理，運営されているウェブ上のサービスである．ここから，事故情報データを取得するに際して，まず，そのデータの項目の一つである「傷病の程度」項目に着目する．本研究では「傷病の程度」項目が「医者にかからず」,「治療 1 週間未満」,「1〜2 週間」,「3 週間〜1ヶ月」,「1ヶ月以上」の事例のみを取得の対象とした．収集の結果，91,005 件の事故情報データが取得できた．収集した事故情報データにおいて，研究に使用する項目を表 3 にまとめる．なお，取得した事故情報データは，いずれも，2020 年 9 月までに事故情報データバンクで公開されたものである．

収集した事故情報データのうち,「傷病の程度」項目が「医者にかからず」以外のものを対象に，無作為に 8:1:1 に分割し，それぞれを訓練データ，検証データ，テストデータとした．治療期間ラベルが「医者にかからず」のものは全て追加の訓練データとした．これらのデータの件数は，訓練データが 82,035 件，検証データが 4,485 件，テストデータが 4,485 件であった．

「職場のあんぜんサイト」は厚生労働省が管理，運営しているウェブ上のサービスである．ここでは，日常生活で発生したヒヤリハット事例が紹介されている．本研究では，公開されるヒヤリハット事例を事故情報データとして扱う．同サービスから無作為に 30 件の事故情報データを収集した．本研究では,「ヒヤリハットの状況」項目に記載されたテキストを事故説明文として扱う．

3.3　事故説明文からの行動文とリスク文の抽出

本節では，事故情報データにおける事故説明文から，行動文とリスク文を抽出する手法について記述する．

「事故情報データバンク」から取得した事故説明文に対して

1：事故情報データバンクシステム, http://www.jikojoho.go.jp/
2：職場のあんぜんサイト，https://anzeninfo.mhlw.go.jp/hiyari/anrdh00.htm

表 1　事故情報データを構成する要素

要素名	特徴	記載例
事故説明文	自由記述	自転車に乗って走行していたら、不意に段差に乗り上げてしまい転倒し、骨折してしまった。
行動文	自由記述	自転車に乗って走行していた
リスク文	自由記述	不意に段差に乗り上げてしまい転倒し、骨折してしまった
商品カテゴリ	全 11 種類のカテゴリ	車両・乗り物
傷病内容ラベル	全 23 種類のカテゴリ	擦過傷・挫傷・打撲傷
治療期間ラベル	全 5 種類のカテゴリ	治療 1ヶ月以上

表 2　使用する事故情報データの分布

データの種類	データ数
訓練データ	26,714 件
検証データ	1,708 件
テストデータ（データバンク）	50 件
テストデータ（あんぜんサイト）	30 件
行動リスク未検出データ	55,321 件

表 3 「事故情報データバンクシステム」の項目ごとに記載される要素

項目名	記載される要素
事故の概要	事故説明文
商品など名称	商品名称
商品など分類	商品カテゴリ
傷病内容	傷病内容ラベル
傷病の程度	治療期間ラベル

は，ルールベースで行動文とリスク文を抽出する．抽出の対象となる行動文とリスク文は，事故説明文の最初の 1 文目から抽出されるものとした．まず，収集した全ての事故説明文に対して，MeCab[3] を用いて形態素解析を行う．形態素解析解析の結果，以下に示す 3 つのパターンのうちの，いずれかに該当するパターンを含む，説明文から行動文とリスク文の抽出を行った．

- パターン 1：仮定形の助動詞（た，だ）
- パターン 2：助動詞（た，だ）と副詞可能な名詞が連続
- パターン 3：副詞可能な名詞（中）

詳細としては，文頭からパターンの直前までを行動文として抽出した．パターン以降の名詞，動詞，形容詞のいずれかから，文の終わりまでをリスク文として抽出した．また，事故説明文中に複数のパターンが含まれていた場合は，基本的には，文頭から見て最初に検出されたパターンを参照した．例外として，パターン 3 とそれ以外のパターンが同時に含まれていた場合は，他のパターンを参照した．

「職場のあんぜんサイト」から取得した事故説明文に対しては，全て人手で，行動文とリスク文を抽出した．

抽出の結果，訓練データ及び検証データからは，表 2 に示す件数の行動文とリスク文のペアが抽出された．テストデータからは，1,773 件のペアが抽出された．

3：unidic 辞書, https://pypi.org/project/unidic/

3.4　人手評価のためのテストデータの用意

本研究では，提案アプローチの評価の際に，テストデータを用いて人手による評価を行う．「事故情報データバンク」由来のテストデータは 1,773 件である．テストデータ 1,773 件すべてに対して人手による評価を行うのは作業量が多いため，本研究では一部のデータを選択し，人手評価用のテストデータとする．人手評価用テストデータを選択する時には，事故情報データ中の要素「商品カテゴリ」に着目し，1 つのカテゴリごとに無作為に 5 件ずつ選択した．結果として，55 件の事故情報データが選択された．しかし，うち 5 件は不適切な内容の行動文，リスク文が含まれていたため，テストデータから除外した．

「職場のあんぜんサイト」から取得した 30 件の事故情報データは全て，人手評価用のテストデータとして使用した．

4　行動文からのリスク文の予測

本研究では，行動文が入力された時に，そこから懸念されるリスク内容が記述されたリスク文を予測する問題に取り組む．1 節でも述べたように，任意の行動文を q，リスク文を d とすると，取り組む問題は以下のように定式化出来る．

入力　q

出力　$d_1, d_2, d_3, \cdots, d_k$

出力は k 件のリスク文である．以下では，まず，上記の問題に対する提案アプローチの概要について述べた後，その詳細について説明する．

4.1　検索アプローチ（Ret）と生成アプローチ (Gen)

本研究では，行動からリスクを予測する問題に対して，検索アプローチと生成アプローチの 2 種類の手法を提案する．

検索アプローチでは，任意の行動文 q が与えられた時に，3.1 節で述べた，既存のリスク文集合 D に対して，ランキングを行う．

$$D = \{d_i \mid i = 1, 2, \cdots, 26714\}$$

d_i は 1 件あたりのリスク文を指す．ランキングに必要なランキング関数を f と定義する．検索アプローチは，$d \in D$ に対して，$f(q, d)$ を計算し，その上位の結果を取得するのが目的となる．

生成アプローチでは任意の行動文 q が与えられた時に，言語生成モデルを用いてリスク文を複数生成する．生成されたリスク文集合を $\hat{D}$ とすると言語生成モデルからの出力は以下のように定式化できる．

$$\hat{D} = \{\hat{d}_i \mid i = 1, 2, \cdots, n\}$$

$\hat{d}_i$ は言語生成モデルから生成されたリスク文を指す．n は任意の整数である．生成アプローチは，生成されたリスク文 $\hat{d} \in \hat{D}$ に対して，検索アプローチと同様のランキング関数 f で $f(q, \hat{d})$ を計算し，その上位の結果を取得するのが目的となる．

まとめると，検索アプローチと生成アプローチを実現するためには，ランキング関数を用意する必要がある．さらに，生成アプローチでは，リスク文を生成するための言語生成モデルが必要となる．

4.2 ランキングのための行動文リスク文適合判定モデル

本研究では，異なる 2 文の適合度合いを推論する手法として，汎用言語モデル BERT を用いた手法を提案する．BERT はファインチューニングにより，目的のタスクに特化させる．ファインチューニングで解くタスクは，行動文とリスク文のペアを入力として，それらが「適合する」か「適合しない」かを二値分類するタスクである．以後はこのタスクのことを適合判定タスクと呼ぶ．また，詳細は 4.6 節で述べるが，本研究では追加学習を行った BERT を用いる．

適合判定タスクでのファインチューニングには，表 2 における訓練データと検証データを使用する．しかし，各データにおける行動文とリスク文のペアは，全てが「適合する」ペアである．BERT のファインチューニングには「適合しない」ペアが必要となる．本研究では，行動文に対応したリスク文をランダムで別のリスク文に置換することで「適合しない」ペアを機械的に作成した．従って，ファインチューニング時のデータの量は，実際のデータ量の倍である．

本研究では，BERT のファインチューニングにおいて，以下の 2 手法を試す．括弧内には各条件によって作成されたモデルの略称を示す．

- シングルタスク学習（**STB**）
- マルチタスク学習（**MTB**）

シングルタスク学習は，適合判定タスクのみで BERT をファインチューニングする手法である．マルチタスク学習は，適合判定タスクを含んだ複数タスクで BERT をファインチューニングする手法である．マルチタスク学習の狙いは，異なるタスク間の相互作用による，モデルの性能向上である．マルチタスク学習で解く他のタスクは，表 1 における「傷病内容ラベル」を推定するタスクと,「治療期間ラベル」を推定するタスクとした．従って，本研究におけるマルチタスク学習では，3 つのタスクを同時に解かせる．

マルチタスク学習のネットワーク図を図 3 に示す．ファインチューニングの際には，BERT の最終層の上に，解くタスクの数だけ追加の全結合層を設ける．以後，全結合層のことを FC 層と呼ぶ．シングルタスクの場合，FC 層は適合判定タスクを解く一つのみである．BERT から出力された [CLS] トークンに対応する 768 次元のベクトルを，タスク毎の FC 層に入力し，推論を行う．損失値は，クロスエントロピーロスにより算出した．マルチタスクの場合は，タスク毎の損失値の和を逆伝播した．

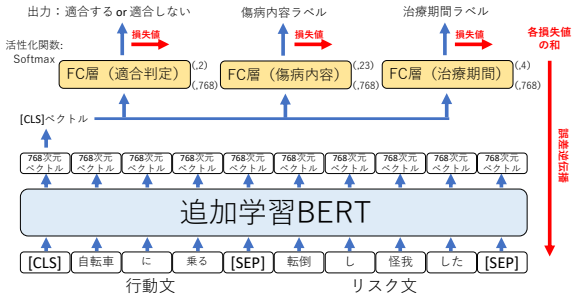

図 3　マルチタスク学習による BERT のファインチューニング

本研究でファインチューニングに使用する,「適合しない」ペアは，機械的に作成したものである．このため,「傷病内容ラベル」,「治療期間ラベル」をもたない．従って，マルチタスク学習において「適合しない」ペアの学習時には，他タスクにおける損失値は計算しないようにした．

適合判定タスクを解く FC 層からは，2 次元のベクトルが出力される．ベクトル内の各要素の値はそれぞれ，入力ペアが「適合する」確率,「適合しない」確率として見なすことが出来る．本研究では,「適合する」確率の値を，ランキング関数 $f(q, d)$ の値として扱う．

4.3 リスク文生成のための言語生成モデル

提案アプローチの一つである生成アプローチでは，入力した行動に対してリスク文の生成を行う．本研究では，リスク文を生成する手法として，以下に示す 2 種類の言語生成モデルを用いた手法を提案する．括弧内には各手法で作成されたモデルの略称を示す．

- GPT-2（**GPT**）
- BERT2BERT（**B2B**）

各モデルをファインチューニングすることで，リスク文の生成に特化させる．また，詳細は 4.6 節で後述するが，本研究では追加学習を行った GPT-2，及び BERT を用いる．以下ではリスク文の生成の詳細について記述する．

4.3.1 GPT-2 を用いたリスク文の生成

言語生成モデル GPT-2 への入力は，トークン化されたテキストである．GPT-2 は，入力されたトークン列の次に続くトークンの予測を繰り返すことで，続きのテキストを生成する．本研究では，行動文を入力した時に，リスク文を生成させるために，GPT-2 をファインチューニングする．

まず，ファインチューニングにおけるネットワーク図を図 4 に示す．ファインチューニング時のモデルへの入力は，行動文とリスク文である．一方で，推論時は行動文のみが入力となる．このため，行動文のみが与えられた時に，続きとしてリスク文を生成するように，生成文を制御する工夫が必要になる．生成文を制御する手法としては，学習時に新たな特殊トークンを導入する手法が過去に提案されている [4], [8]．そこで，本研究では，ファインチューニングの際に，新たな特殊トークン [RISK] を GPT-2 に追加する．特殊トークン [RISK] は，行動

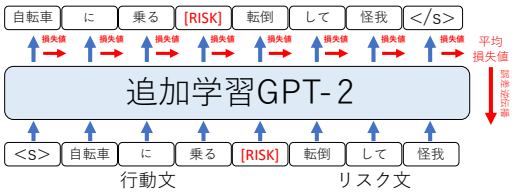

図 4　GPT-2 のファインチューニング

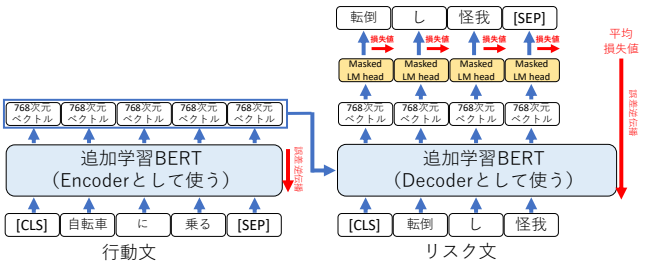

図 5　BERT2BERT のファインチューニング

文を GPT-2 に入力する際には，必ず末尾に置くトークンとして扱う．

ファインチューニングでは，入力した全てのトークンに対して，次に続くトークンを予測するタスクを解かせる．損失値は，クロスエントロピーロスにより算出する．推論時は，行動文の末尾の [RISK] から，次に来るトークンを予測することで，リスク文の生成を行う．

4.3.2　BERT2BERT を用いたリスク文の生成

BERT2BERT は，Rothe ら [6] により提案された Sequence-to-Sequence モデルの 1 種である．モデル構造としては，Encoder と Decoder の双方に BERT を用いた構造となっている．Decoder 側の BERT の Attention 機構は単方向化している．本研究では，追加学習された BERT を用いて，BERT2BERT モデルを構成する．その後，モデルをファインチューニングすることで，リスク文の生成に特化させる．

ネットワーク図を図 5 に示す．ファインチューニングの際には，Encoder 側に行動文を入力し，Decoder 側にリスク文を入力する．図 5 中における，Decoder 上部の「Masked LM head」は，追加学習時に用いたものと同じである．追加学習では，[MASK] トークンで隠されたトークンを予測していたが，ファインチューニングでは，入力トークンの次に続くトークンの予測を行う．

リスク文の生成時には，Encoder 側に行動文を入力し，Decoder に [CLS] トークンのみを入力する．[CLS] トークンに続くトークンを，予測させることで，リスク文を生成する．

4.3.3　生成文を複数取得するための工夫

言語生成モデルを用いた，一般的なテキストの生成手法として，グリーディサーチが挙げられる．言語生成モデルに対してテキストを入力すると，モデルが網羅する語彙ごとの生成確率が出力される．モデルの語彙数が 32,000 であった場合，32,000 個の生成確率が出力される．グリーディサーチでは，各語彙に対応する生成確率の中で，最も高い確率を示す語彙を生成トークンとして選択する．生成トークンを入力テキストの末尾に追加し，再び言語生成モデルに入力する．上記の処理を繰り返すことで，テキストを生成する手法がグリーディサーチである．従って，グリーディサーチでは，1 つの入力に対して，1 つの生成文を得ることができる．

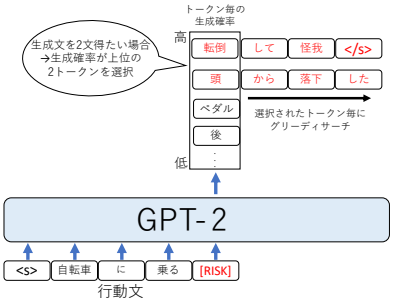

図 6　生成文を複数得るための工夫

本研究では，1 つの入力に対して複数の生成文を得るために，グリーディサーチに工夫を導入する．図 6 には，導入した工夫の内容を示す．導入した工夫では，最初の生成トークンを選択する際に，生成確率が高い順に，上位 n 件のトークンを選択する．入力テキストの末尾に，選ばれた n 件のトークンのうち，1 トークンを追加し，グリーディサーチを行う．これを n 件分繰り返すことで，1 つの入力に対して，n 件の生成文を取得できる．本研究では，$n = 1000$ として生成を行った．生成文の文長は，各生成モデルにおけるトークン単位で 30 とした．

4.4　ベースライン手法 (Base)

本研究におけるベースライン手法は，入力した行動文 q と，データセット中の行動文 $action(d) \in D$ の類似度で，リスク文 $d \in D$ をランキングする手法である．ベースライン手法におけるランキング関数 $f(q, d)$ は，以下の様にして算出する．

$$cos(v(q), v(action(d_i)))$$

v はテキストをベクトル化する関数，cos はベクトルのコサイン類似度を計算する関数を示す．

本研究では，v として TF-IDF や BERT を用いた複数条件を試す．BERT は事前学習モデルと，4.6 節で後述する追加学習モデルを用いた 2 条件を試す．テキストをベクトル化する際の各条件を以下にまとめる．括弧内には各条件の略称を示す．

- TF-IDF でベクトル化（**TFIDF**）
- 事前学習 BERT でベクトル化（**PreBERT**）
- 追加学習 BERT でベクトル化（**AddBERT**）

BERT にトークン化したテキストを入力すると，各入力トークン毎にベクトルが出力される．BERT を用いてテキストをベクトル化する手法は，出力された各ベクトルをどのように扱うかで，複数考えられる．本研究では，出力されたベクトルのうち，特殊トークン以外のトークンに対応したベクトルの平均を用いた．

4.5　出力されたランキング結果に対する多様性の考慮

本研究では，ランキング結果の多様性を確保するため，MMR（Maximal Marginal Relevance）[1] による再ランキングを行

う．具体的にはリスク文の $|S|$ 位までのランキング S が与えられたときに $|S|+1$ 位のリスク文を以下の式によって選択し，S の末尾に追加する．

$$\arg\max_{d_i \in D \setminus S}[\lambda f(q, d_i) - (1-\lambda)\max_{d_j \in S}(\cos(v(d_i), v(d_j)))]$$

f はランキング関数である．v はテキストをベクトル化する関数である．MMR を計算する際には，4.6 節で後述する追加学習 BERT を用いた．ベクトル化の方法は，ベースライン手法と同様である．cos はコサイン類似度を計算する関数である．D は収集したリスク文の集合である．本研究では，全ての手法のランキング結果に対して MMR による再ランキングを行う．

上記の処理に加えて，本研究では 7 文字以下のリスク文や，同じ文字列が 3 回以上繰り返し出現するリスク文はランキング結果から除外した．

4.6 言語モデルに対する追加学習

4.2 節や 4.3 節では，BERT や GPT-2 といった言語モデルのファインチューニングについて述べた．本研究で取り扱った言語モデルは，いずれも Hugging Face の Transformers ライブラリ上で公開される事前学習済みモデルである．BERT は東北大学の乾研究室が公開するモデル「cl-tohoku/bert-base-japanese-whole-word-masking [4]」を取り扱った．GPT-2 は rinna 社が公開するモデル「rinna/japanese-gpt2-medium [5]」を用いた．

本研究では，上記の事前学習済みモデルをファインチューニングする前に，追加学習を行う．追加学習には，表 2 中の訓練データ，検証データ，行動リスク未検出データにおける事故説明文を用いた．行動リスク未検出データは追加の訓練データとして扱った．追加学習で解くタスクは，BERT は MLM タスクのみを用いた．GPT-2 は事前学習と同様のタスクで，追加学習した．さらに，BERT の追加学習では，事故説明文から確認された未知語をモデルに追加した．

5 実　　験

本節では，本研究で取り組んだ実験について述べる．まず，各アプローチにおける実験条件や，その評価手法について説明する．その後，実験結果に対する議論を行う．

5.1 実 験 条 件

表 4 には本実験で取り組む 9 種類の実験条件を示す．表中の略称において，冒頭が「Ret」が検索アプローチ,「Gen」が生成アプローチ,「Base」がベースライン手法であることを表す．本実験では，表 2 における 2 種類のテストデータを用いる．出力されるリスク文は，ランキング結果の内，上位 5 件のリスク文とした．

出力されたランキング結果を評価するため，本実験では，3

表 4　本研究における 9 種類の実験条件

実験条件の略称	リスク文の生成	ランキング時のスコアの計算
RetSTB	なし	STB
RetMTB	なし	MTB
GenGPT-STB	GPT	STB
GenGPT-MTB	GPT	MTB
GenB2B-STB	B2B	STB
GenB2B-MTB	B2B	MTB
BaseTFIDF	なし	TFIDF
BasePreBERT	なし	PreBERT
BaseAddBERT	なし	AddBERT

名の実験参加者を募り，人手による評価を実施した．評価タスクにおいて，評価者は入力した行動文と出力されたリスク文のペアを確認する．評価者は出力された各リスク文に対して，入力した行動文との関係の強さに応じてラベル付けを行う．用いたラベルと，評価者に対する説明を以下に示す．

ラベル「2」:「行動文」と「リスク文」は強く関係している
ラベル「1」:「行動文」と「リスク文」は多少関係している
ラベル「0」:「行動文」と「リスク文」はほとんど関係していない，または,「リスク文」が文として成立していない

上記のラベル「0」〜「2」は順序尺度である．3 名によるラベル付けの結果から，最終的な評価に使用するラベルは多数決により決定した．評価者全員のラベルが異なっていた場合は，ラベル「1」とした．

本実験では，テストデータ内の行動文を 1 つ入力すると，各実験条件ごとにランキング結果が出力される．各ランキング結果に対して，nDCG@5 を計算した．本実験では，全てのテストデータに対して計算された nDCG@5 の値の平均を求め，各アプローチの評価に用いた．

5.2 リスク文予測における結果と考察

表 2 における 9 種類の実験条件ごとの，テストデータに対する予測結果を表 2 に示す．本実験では，2 種類のテストデータを用いて評価を行った．いずれの結果においても，生成アプローチである「GenGPT-STB」が nDCG@5 において，最も良い結果を示した．nDCG@5 の値は，高いほど，入力した行動文と関係がある内容のリスク文を予測できたことを示す．従って，nDCG@5 による評価では，生成アプローチ「GenGPT-STB」が他条件と比べて，入力した行動と関係のある内容のリスク文が予測できる結果となった．

表 5 には，生成アプローチ「GenGPT-STB」と検索アプローチ「RetSTB」における実際の予測結果を示す．入力した行動文は,「二重瞼用糊を付けた」である．ランキング結果の 1 位に着目すると，検索アプローチでは「ゴムが外れて目にあたり充血した」であった.「ゴムが目に当たる」リスクは「二重瞼用糊を付けた」行動における「糊」の特徴を考慮できていないことが懸念される．一方で，生成アプローチでは,「当該糊が目に入り、眼科医の診察を受けた」が 1 位であった．これは，行動文

4：BERT 事前学習モデルの詳細，https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking
5：GPT-2 事前学習モデルの詳細，https://huggingface.co/rinna/japanese-gpt2-medium

表 5　行動文「二重瞼用糊を付けた」に対して予測されたリスク文

順位	RetSTB	GenGPT-STB
1	ゴムが外れて目にあたり充血した	当該糊が目に入り、眼科医の診察を受けた
2	眼瞼下垂等の重症	初めは痛かったが、2 週間程で取れた
3	ふさぎ込んでしまうようになった	マスカラが落ちない
4	染料が眼に入り、角膜上皮が広範囲剥離し、球結膜に化学傷が生じる重症	次第に瞼が腫れ、使用をやめた
5	止めて他院で切取り縫合した	埋没法で二重瞼にする手術を受けた

表 6　各テストデータに対する nDCG@5 の結果

実験条件	データバンク	あんぜんサイト
RetSTB	0.348	0.090
RetMTB	0.337	0.069
GenGPT-STB	**0.475**	**0.250**
GenGPT-MTB	0.414	0.188
GenB2B-STB	0.371	0.165
GenB2B-MTB	0.310	0.144
BaseTFIDF	0.340	0.033
BasePreBERT	0.354	0.030
BaseAddBERT	0.377	0.042

中の「糊」を考慮して予測された結果と考えられる．上記のことから，生成アプローチは，検索アプローチよりも，行動文の特徴を考慮したリスクの予測が可能であることが考えられる．

6　まとめと今後の課題

本研究では，日常的な行動を表す文を入力した時に，そこから考えられるリスクを予測する手法を提案した．予測されるリスクは，その内容が，自然言語で記述されたリスク文として出力を想定する．提案アプローチとしては，検索アプローチと生成アプローチの 2 種類を提案した．検索アプローチは，蓄積されたリスク文の中から，出力となるリスク文を検索するアプローチである．生成アプローチは，言語生成モデルを用いて，出力となるリスク文を生成するアプローチである．人手評価による結果では，生成アプローチの方が，検索アプローチよりも，入力した行動に関係するリスクが予測できる結果となった．しかし，生成アプローチで生成されたリスク文の中には，日本語として成立していない文も見受けられた．このため，今後は，文としての自然さを考慮した工夫の導入が必要であると考えている．

謝　　辞

本研究は JSPS 科研費 JP21H03775，JP21H03774，JP21H03554，JP18H03244，ならびに，2021 年度国立情報学研究所公募型共同研究（21S1001）の助成を受けたものです．ここに記して謝意を表します．

文　　献

[1] J. Carbinell and J. Goldstein. The use of MMR, Diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 2017 Special Interest Group in Information Retrieval Forum*, pp. 209–210, 2017.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 4171–4186, 2019.

[3] S. Gururangan, A. Marasovic, S. Swayamdipta, K. Lo, I. Beltagy, D. Downey, and N. A. Smith. Don ' t stop pretraining: Adapt language models to domains and tasks. In *Proceedings of the 2020 Annual Meeting of the Association for Computational Linguistics*, pp. 8342–8360, 2020.

[4] N. S. Keskar, B. McCann, L. Varshney, C. Xiong, and R. Socher. CTRL: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858*, 2019.

[5] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever. Language models are unsupervised multitask learners. Technical report, OpenAI.

[6] S. Rothe, S. Narayan, and A. Severyn. Leveraging pretrained checkpoints for sequence generation tasks. *Transactions of the Association for Computational Linguistics*, 8:264–280, 2020.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, U. Kaiser, and I. Polosukhin. Attention is all you need. In *Proceedings of the 2017 International Conference on Neural Information Processing Systems*, pp. 6000–6010, 2017.

[8] 永塚光一, 渥美雅保. 文生成の長さ制御のための条件付き言語モデルの学習. 情報処理学会第 82 回全国大会, 2020.

[9] 山田侑樹, 櫨山淳雄, 小川雄太郎. OSS プロジェクトの Issue 議論内容に対する BERT および AutoML を用いた文章分類の提案. 人工知能学会第 34 回全国大会, 2020.

[10] 山本建太郎. 救急業務に関連した事故事例及びヒヤリハット事例における心理的要因の分析. 日本救急医学会関東地方会雑誌, pp. 405–408, 2018.

[11] 小菅英恵. ヒヤリハット類型と日常運転行動，運転意識の関係—安全教育対策検討のためのヒューマンエラー分析—. 労働安全衛生研究, 11(1):25–37, 2018.

[12] 森泉慎吾, 臼井伸之介, 和田一成, 上田真由子. 急ぎ・焦りエラーに関する体験型教育の効果. 労働科学, 94(4):99–107, 2018.

[13] 村田信, 須藤信行, 平山悠太, 中村哲. Jr 東日本 「他山の石」置換え支援ツールに基づく練習船で発生したヒヤリハットの分析. 海技教育機構論文集, 5:68–73, 2019.

[14] 池田良磨, 土田靖高, 岡崎直宣. 雨天時におけるヒヤリハットの分析. 宮崎大学工学部紀要, pp. 269–273, 2020.

[15] 内藤勝太, 白松俊. Web 議論における BERT を用いた関連情報推薦エージェント. 情報処理学会第 82 回全国大会, pp. 637–638, 2020.

[16] 壹岐太一, 金沢輝一, 相澤彰子. 学術分野に特化した事前学習済み日本語言語モデルの構築. Technical report, 国立情報学研究所.

[17] 檜山明子, 中村惠子. 入院患者の転倒リスクが高い行動の分析. 日本看護研究学会雑誌, 40(4):657–665, 2017.