

# 事故概要説明文の文脈理解による傷病の程度の推定

川原 敬史<sup>†</sup> 橋口 友哉<sup>†</sup> 湯本 高行<sup>††</sup> 大島 裕明<sup>†,††</sup>

<sup>†</sup> 兵庫県立大学 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

<sup>††</sup> 兵庫県立大学 社会情報科学部 〒651-2197 兵庫県神戸市西区学園西町 8-2-1

E-mail: <sup>†</sup>{aa20m503, aa19j508, ohshima}@ai.u-hyogo.ac.jp, <sup>††</sup>yumoto@sis.u-hyogo.ac.jp

**あらまし** 本研究では、事故の概要を説明したテキストを入力として、当事者が受けた傷病の程度を推定する手法を提案する。入力の対象とするテキストは、数文程度の文書を想定する。ウェブ上で入手可能な大量の事故情報を収集し、それらを学習データとして機械学習を用いて分類問題を解くことで提案手法を構成する。提案手法では、入力テキストを BERT を利用して分散表現として表現する。BERT では事前学習モデルを用いるが、傷病の程度を表している文かどうかを判別する問題を解かせることによってファインチューニングを行う。また、本研究ではファインチューニングにおいて、別の分類タスクと同時に学習を行うマルチタスクラーニングや、分類クラスの順序を考慮した Ordinal Classification 等の複数の異なる手法を試す。そして、それらのモデルの分類精度を比較することで、傷病の程度を推定するタスクに適した手法の探索を行う。

**キーワード** 事故情報、傷病の程度の推定、深層学習

## 1 はじめに

我々の日常生活において事故が発生することは避けられないことである。しかし、過去に発生した事故を事前に認知していれば、その類似事故に対して、事前に警戒を強めることができる。これは、未来に発生しうることのリスクを低下させる効果が期待できる。近年では、ウェブ上における「Twitter」や「Yahoo! 知恵袋」といった CGM (Consumer Generated Media) の普及により、誰でも他人が遭遇した事故に関する情報を容易に得られるようになった。CGM に寄せられる事故情報の多くは、テキストによって事故の内容が表現されている。しかし、寄せられる事例の中には事故の状況を十分に表せていないものもある。例えば、「刺身を食べて食中毒になった」という事例の場合、「何の刺身を食べたのか」や、「治療に要した期間」等の情報が含まれていない。このような情報の不足は、閲覧者の事故に対する警戒意識にばらつきを発生させる原因となることが考えられる。事例を目にした閲覧者の中で、過去に刺身を食べて重度な被害を受けた者であれば、寄せられた事例を警戒すべき事例として受け取りやすい。一方で、過去に刺身を食べて健康被害が生じなかった閲覧者や、刺身を食べたことのない閲覧者は事故を軽視しがちになると考えられる。このような事故に対する軽視は、事故の再発に繋がりがかねない。そこで、本研究では事故に対する事前知識の乏しい閲覧者が事故への警戒の有無を判断する際、その正しい判断を補助するような付加情報の推定を目指す。

前述のように、過去に重大な被害を被った閲覧者は、類似事故に対して警戒を強めやすくなる傾向が考えられる。これは、再度同程度の被害を被りたくないという閲覧者の警戒感の高まりに由来することが考えられる。そこで、経験者が警戒するような被害の程度を事前知識の乏しい閲覧者とも共有できるよう

になれば、事故の再発防止に繋がることが期待できる。そのためには、事故で発生した被害の程度を定量的に示す必要があると考えた。本研究では、被害の程度を定量的に示す情報として当事者が事故で受けた傷病の治療期間に着目する。

以上のようなことから、本研究では事故の概要が記述されたテキストから、当事者が受けた傷病の治療期間を推定する問題に取り組む。研究に使用するデータは、事故情報データベースシステムが公開する事故事例データを取り扱う。このような、自然言語を取り扱った問題に対するアプローチとしては、まず、事前学習された BERT モデルをファインチューニングすることが考えられる。しかし、本研究で取り組む問題は事前に整えられたデータセットが用意されているような問題ではなく、社会における実問題である。このような問題においては、それ特有の改善へ向けた工夫が必要になる。そこで、本研究では以下に示すような 4 つの手法を用いてファインチューニングを行い、各種分類モデルを作成した。

- **マルチタスクラーニング**
  - － 複数の分類タスクを同時に学習
- **Ordinal Classification**
  - － 分類クラス間の順序を考慮した学習
- **クラスウェイト**
  - － 分類クラス毎の訓練データ量の偏りを考慮した学習
- **「商品名称」認識モデルへのファインチューニング**
  - － 目的のタスクを学習させる前に別のタスクを学習させたモデルを使用

また、これらの手法によるモデルの分類性能への影響は、手法を用いなかった場合の分類性能との比較により評価した。

以下、2 節では関連研究について述べる。その後の 3 節では実験に使用する事故事例データの説明と問題定義を行う。続く 4 節では基本となるファインチューニングの手法について説明した後、本研究で導入した手法の詳細について記述する。

## 2 関連研究

本研究では、事故に関する情報が記載されたテキストの分析を行い、そのリスクの把握を目指す。そこで、本節では過去に行われた事故情報を取り扱った研究や、何らかのリスクの把握を目的とした研究について紹介する。また、本研究において取り扱う言語処理技術に関連した研究についても紹介する。

### 2.1 事故情報解析

八十岡ら [20] の研究では消費者から国民生活センターに寄せられる事故情報の分析が行われている。この研究の目的は、寄せられた事例から健康被害が発生した確率が高い事例のみを抽出することである。ここで作成された分類モデルは約 9 割の分類性能を示した。さらにこの研究では、作成された分類モデルを用いて「Yahoo!知恵袋」における事故相談事例に対してラベル付けを行っている。そして、ラベルが付与されたデータを訓練データに加え、再度分類モデルを作成を行うことでモデルの性能が向上することが報告されていた。このことから、CGM 上の事故事例を活用することは、事故の分析において有効であることが考えられる。しかし、同論文では CGM 上の事例にはノイズとなる情報も多く、しきい値等の違いによってはその分類性能が低下することも懸念されていた。

八十岡らの研究では大規模な非公開データセットを用いて分類モデルを作成していた。一方で、公開されているデータセットのみを用いた研究も行われている。例としては、栗原ら [14] や、阿部ら [11] の研究が挙げられる。栗原らは Twitter に寄せられる tweet の中から自転車等の製品の不具合に関連する tweet の抽出に取り組んでいた。阿部ら [11] の研究では皮膚障害に関連した tweet を抽出し、その原因となりうる製品名の探索を試みていた。また、ウェブ上の情報を取り扱った研究ではないが、新聞に記載されているテキストから交通事故関連テキストを抽出する研究が酒井ら [17] によって行われている。抽出されたテキストは事故の要因の探索に使用される。

これらの過去の研究から、事故の内容が記載されたテキストを抽出、活用する事は事故の分析において有効であると考えられる。本研究で行うことは、事故に関連したテキストに対して、追加の判断材料となる情報を付与することである。このような本研究での取り組みは、CGM に寄せられる事例の分析において、これまで以上のアプローチを可能にすることが期待できる。

### 2.2 リスクの定量化

精神医療分野においては、池田 [18] によって、医療カルテに含まれる自殺リスクを定量的に把握する研究が報告されている。ここで扱われる医療カルテでは、患者の状態がテキストで表現されている。池田はこのようなカルテ内のテキストの特徴から患者の自殺リスクの推定に取り組んでいた。手法としては、患者の自殺リスクがラベル付された医療カルテのデータセットを用いて、ナイーブベイズ法による分類モデルの作成が行われた。この分類モデルは、約 8 割程度の正答率を示す結果であった。このことから、テキストを用いた分類手法がリスクの定量的な

把握に対して、有効であることが考えられる。

本研究で定量的な把握を目指すリスクは、事故で発生した被害の程度である。被害の程度を表す指標としては、当事者の受けた傷病の治療期間が挙げられる。以下では、傷病の治療期間にを扱った研究を紹介する。峰崎ら [21] の研究では看護事故を分析する際、患者の重症度は事故の被害把握において必要な情報として扱われていた。また、園田ら [12] の集団救急の効率化を目的とした研究では患者を最も適した医療機関に搬送するための判断指標として、患者の負傷程度（重症、軽傷等）を用いていた。さらに、吉田ら [13] の研究では交通事故における被害状況の早期把握のために、エアバック装置内に記録された情報から当事者の重症度の推定が行われていた。これらの過去の研究から、発生した事故におけるリスクの定量化において、事故で発生した傷病の治療期間を用いることは適正であると考えられる。また、本研究で取り扱う事故事例データでは、その項目の中に、事故の当事者が受けた傷病の治療期間を表す項目がある。そこで、本研究ではこのような項目を活用し、事故において発生したリスクの定量化を目指す。

### 2.3 使用する汎用言語処理モデル（BERT）

BERT は自然言語処理を目的としたディープラーニングモデルの一つとして、2018 年に Google の Devlin ら [3] によって提案された。これは、双方向の Transformer [10] に対して、大規模なコーパスを事前学習させることでモデルの作成を行う。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。Devlin らの研究では、既存の様々な自然言語タスクにおいて、従来手法よりも高い性能を示した。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 [15] [19]。

## 3 取り扱う事故事例データと問題定義

本節ではまず、本研究で利用する事故情報データバンクシステムについて述べ、研究に使用するデータセットの説明を行う。その後、本研究における問題定義を行う。

### 3.1 事故情報データバンクシステム

事故情報データバンクシステム<sup>1</sup> 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データベースである。ここでは、生命や身体に係る消費生活上の事故情報を各関係機関から一元的に集約し、消費者への情報発信が行われている。この取り組みは平成 22 年より行われており、令和 2 年 10 月現在では約 19 万件の事例データが公開されている。この事例データは「事故情報」と「危険情報」の 2 種類に大別される。このうち「事故情報」は当事者が事故により健康被害を患った事例データであり、「危険情報」は健康被害が発生しなかったと

1：事故情報データバンクシステム：<http://www.jikojoho.go.jp/>

考えられる事例データを指す。本研究では、事故によって発生した傷病の治療期間に着目するため、「事故情報」に属する事例データのみを取り扱う。また、公開される事例データは複数の項目から構成されている。それらの項目の種類を表 1 に示す。また、ここに示す項目は代表的な項目であり、事例によっては別の項目が追加されている場合もある。これらの項目のうち、本研究で扱うのは以下の 4 項目に記載される情報である。

- 自由記述
  - － 「商品など名称」項目
  - － 「事故の概要」項目
- カテゴリ
  - － 「傷病内容」項目
  - － 「傷病の程度」項目

以下では、これらの項目についての説明を行う。「事故の概要」項目では、その事故における背景や要因、発生した身体的被害等の情報がテキストで記述されている。以降、この項目に記載されるような事故の概要を表すテキストのことを**事故概要テキスト**と呼ぶこととする。事例データにおける事故概要テキストの文量は、基本的には 1～3 文で記載されていることが多い。本研究ではこの項目に記載される事故概要テキストの内容を CGM 上で見受けられる事故相談文や、事故報告文に相当するテキストとして扱う。「商品など名称」項目では、その事故の発生に関与したと考えられる物や施設、サービス等の名称が記載されている。この項目に記載される物やサービスの名称は対応する事故概要テキストにも含まれていることが多い。「傷病内容」項目では火傷や、骨折といったような事故で発生した傷病の種類が記載されている。記載される内容の種類は未記載の場合を含めて全部で 23 種類である。「傷病の程度」項目では事故によって当事者が被った傷病の程度について記載されている。この項目に記載される内容の種類は未記載の事例を含めて、8 種類である。これらのカテゴリの中で、「治療 1 週間未満」、「治療 1～2 週間」、「治療 3 週間～1ヶ月」、「治療 1ヶ月以上」の 4 種類においては、当事者が被った傷病の治療期間が示されている。本研究ではこれら 4 種類の治療期間を**治療期間ラベル**として扱う。治療期間ラベルが付与された事例データでは、事故の当事者が医師による診断を受けている。このため、これらのデータは事故で発生した被害程度の分析に用いることができる。一方、他のカテゴリに属する事例データでは、事故で受けた傷病の治療期間が不明瞭である。中でも、「傷病の程度」項目が「**医者にかからず**」とされる事例データでは、当事者が医師による診断を受けていないことが確かである。このため、属する事例データでは治療期間が短い軽度な傷病の事例と潜在的な重症事例が混在していることが伺える。しかし、「医者にかからず」カテゴリに属するデータの数は、治療期間ラベルが付与された事例データと同程度の量がデータバンクに存在する。従って、「医者にかからず」カテゴリに属する事例データは、事故で発生した被害程度の分析には扱えないが、寄せられる事故のバリエーションの把握には有用な事例であることが考えられる。

本研究では、上記のような公開される事例データの特徴を踏まえて、以下の条件を満たす事例データを事故情報データバン

表 1 公開される事故事例データにおける項目

項目名	記載内容
種別	事故情報 or 危険情報
発生日月日	事故の発生した年月日
発生場所	事故の発生した場所
商品など分類	事故に関連した商品等の種類
商品など名称	事故に関連した商品等の名称
事故内容	事故内容の種類
事故の概要	事故概要テキスト
傷病内容	当事者が受けた傷病の種類
傷病の程度	当事者が受けた傷病の治療期間
被害者年代、性別	当事者の年代や性別
情報提供元	データバンクに情報を提供した機関の名称
登録年月日	その情報が登録された年月日
受付年月日	その情報が受け付けられた年月日

クシステムから収集した。

- 収集の対象となる事例データ
  - － 「種別」項目が事故情報
  - － 「傷病の程度」項目が治療期間ラベル、「医者にかからず」

収集された事例データは 2020 年 9 月までに事故情報データバンクシステムで公開されたものであり、その総数は 91,005 件であった。収集したデータにおける「傷病の程度」項目の種類毎の内訳を図 1 に示す。収集した事例データのうち、「傷病の程度」項目に治療期間ラベルが記載されているデータは 44,961 件であり、「医者にかからず」データは 46,044 件であった。

事故情報データバンクシステムから収集した事例データの中には、事故概要テキストにおいて、「当該製品」という語を含むものが存在している。このようなテキストからは、事故の発生にどのような物が関与したのかを読み取ることができない。事例データにおける事故に関与した物の名称は表 1 における「商品など名称」項目に別途記載されている。そこで、データバンクから収集したデータのうち、事故概要テキストにおいて「当該製品」という語を含むものに関しては、その語から「商品など名称」項目に記載される語への置換を行った。その例を以下に示す。

- 事故概要テキストに対する前処理の例
  - － **処理前** 当該製品に乗って走行中、転倒し怪我をした。
  - － **処理後** 自転車に乗って走行中、転倒し怪我をした。

このような置換済みのテキストを本研究で扱うデータセットとした。

### 3.2 問題定義

本研究における問題定義を行う。本研究で取り組むのは、事故の概要を示すテキストがあるときに、そのテキストの特徴から事故で発生した傷病の治療期間の推定することである。その入力と出力は以下の通りである。

**入力** 事故概要テキスト

**出力** 治療期間ラベル（治療 1 週間未満 or 治療 1～2 週間 or 治療 3 週間～1ヶ月 or 治療 1ヶ月以上）

従って、本研究で取り組む問題は入力されたテキストを 4 つの

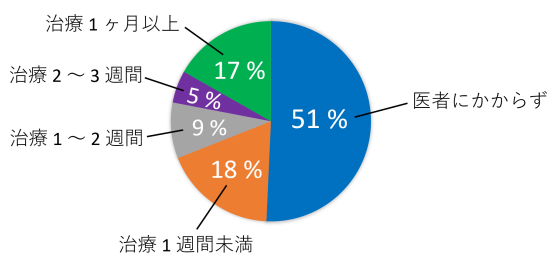


図 1 収集データ 91,005 件における「傷病の程度」項目の分布

カテゴリのいずれかに分類する 4 値分類問題である。また、推定の対象となるデータは収集した事例データにおいて治療期間ラベルが付与されている 44,961 件のデータである。

これらの結果については、正解率や F1, RMSE, 混同行列で評価するものとする。RMSE は本来、回帰問題を解く場合に用いられる評価指標である。しかし、本研究における出力の各分類クラスでは、従来の分類クラスとは異なり、それぞれに明確な順序が存在している。従って、本研究では出力されるクラスの順序を考慮した評価を行うために RMSE を用いる。

## 4 事故概要テキストを用いた被害程度の推定方法

本節では、入力される事故概要テキストから事故により当事者が受けた傷病の治療期間を推定する手法について説明する。まず、本研究で使用する BERT 事前学習モデルについて説明し、その後、モデルのファインチューニングについて記述する。

### 4.1 使用する BERT 事前学習モデル

本研究では、テキスト内の文脈を考慮した分散表現を用いて分類を行うため、BERT 事前学習モデルを使用する。本研究で扱うテキストは全て日本語であるため、日本語で事前学習されたモデルを使用する事が望ましいと考えた。そこで、本研究では自然言語処理ライブラリの一つである transformers ver3.10<sup>2</sup> に登録されている日本語 Wikipedia 事前学習モデル (cl-tohoku/bert-base-japanese-whole-word-masking) を使用する。このモデルは東北大学の乾研究室により提供された事前学習モデル<sup>3</sup>であり、日本語 Wikipedia を訓練データにしている。事前学習時のテキストに対する前処理としては、MeCab による形態素解析を行った後、WordPiece によるサブワードへの分割が行われている。従って、モデルへの入力はサブワード由来のトークン列である。語彙数は 32,000 語である。

### 4.2 治療期間推定モデルの作成方法

本研究における分類モデルの作成方法は、事前学習された BERT を、収集した事故データを用いてファインチューニングすることである。本節ではまず、基本となるファインチューニングの手法について説明する。

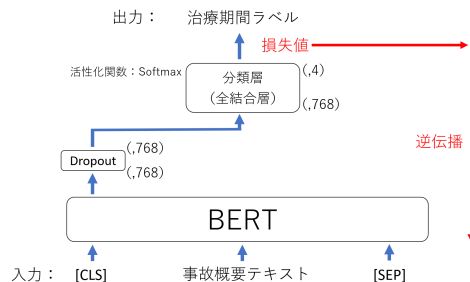


図 2 BERT モデルの基本的なファインチューニング

BERT モデルをファインチューニングする際には、図 2 に示すように、モデルの Transformer 層の上に追加の全結合層を設置する。以後、この層の事を分類層と呼ぶこととする。この分類層では BERT モデルから出力されるベクトルを入力として、各種分類ラベルの推論を行う。本研究では、BERT モデルから出力された [CLS] トークンに対応した 768 次元のベクトルに Dropout [9] を適用したものを入力として扱う。分類層からの出力は分類クラスの種類に応じた次元のベクトルである。4 値分類タスクの場合、分類層からは 4 次元のベクトルが出力される。本研究では分類層から出力されたベクトルの要素のうち、最も高い値を示す要素に着目して分類を行う。そして、その値が位置する次元に対応したクラスをモデルによる予測クラスとして扱う。例えば、分類層から出力されたベクトルが [0.06, 0.02, 0.88, 0.04] であり、正解のクラスが 0, 1, 2, 3 のいずれかである場合、予測クラスは 2 となる。正規化の際に使用した活性化関数は Softmax である。また、予測と正解の誤差を表す損失値は、損失関数クロスエントロピーロスで算出する。この損失値をモデルに対して逆伝播することで、モデル内のパラメータを各種分類ラベルの推定に適したものへと更新する。

以上が基本的なファインチューニングの流れである。BERT モデルに対してファインチューニングを行う手法は、他にも様々なものが考えられる。そこで、本研究では以下に示す 3 種類の手法を導入する。

- マルチタスクラーニング
- Ordinal Classification
- クラスウェイト

以下では、これらの手法の詳細について説明する。

#### 4.2.1 マルチタスクラーニング

ニューラルネットワークを用いた学習において、目的とするタスクと、それに関連のある別タスクを同時に解かせた場合、モデルの分類性能が改善することが報告されている [8]。このような学習手法はマルチタスクラーニングと呼ばれている。本研究で収集した事故データの中には、治療期間に関する情報以外に、事故で発生した傷病の種類が「傷病内容」として記載されている (表 1)。その例としては「火傷」や「骨折」等が挙げられる。このような「傷病内容」は傷病の治療期間と関連のある情報であることが考えられる。例えば、「擦り傷」事例と「骨折」事例では、後者の方が治療期間が長くなるような傾向が考えられる。そこで本研究では、傷病の治療期間と「傷病内容」

2 : transformers : <https://huggingface.co/transformers/v3.1.0>

3 : 提供元の GitHub : <https://github.com/cl-tohoku/bert-japanese>

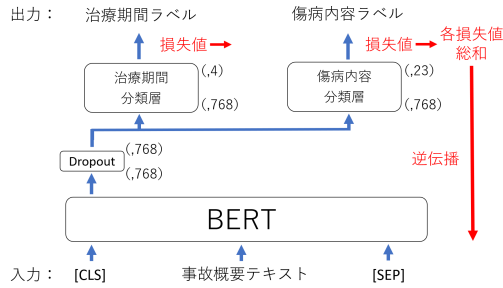


図3 マルチタスクラーニング

のマルチタスクラーニングに取り組む。各タスクにおける推定の対象を以下に示す。

- 推定の対象
  - － 治療期間ラベル：4種類
  - － 「傷病内容」を示すラベル：23種類

次に、本研究で実施するマルチタスクラーニングの概要を説明する。まず、学習全体の流れを図3に示す。前述のファインチューニングとマルチタスクラーニングの相違点は、BERTモデルの上に設置する分類層が二つという点である。これらの分類層への入力とは前述のファインチューニングにおける分類層への入力と同じである。そして、各分類層ではそれぞれタスクにおける推論が行われる。また、逆伝播に使用する損失値は、各分類層での推論結果から算出された損失値の和を用いる。

#### 4.2.2 Ordinal Classification

本研究で扱う分類クラスは、事故の当事者が受けた傷病の治療期間を示している。従って、異なるクラス間には明確な順序が存在している。このような順序付きのクラスを分類する際、その順序を考慮した学習手法として Ordinal Classification が提案されている [4] [7]。この手法と従来手法の大きな違いの一つは、分類クラスに付与するラベルにある。従来、分類クラスに付与するラベルは 0, 1, 2 というような数値である。以後、このような一般的なラベルのことを **NC ラベル** と呼ぶ [2]。一方、Ordinal Classification で使用されるラベルはクラスの種類 -1 の次元からなるベクトルである。以後、Ordinal Classification で使用するラベルを **OC ラベル** と呼ぶ。例えば、4 値分類問題における OC ラベルは [0, 0, 0], [1, 0, 0], [1, 1, 0], [1, 1, 1] となる。従って、OC ラベルにおける各要素の値は 0 もしくは 1 のみである。このことから、ラベル内の単一の要素を推定するタスクは二値分類タスクとして見なすことができる。Ordinal Classification では、このような二値分類タスクを並列して複数同時に解くことで、クラス間の順序を考慮した学習を行う。

本研究で取り組む Ordinal Classification の概要を図4に示す。また、本研究において各分類クラスに付与するラベルを表2に示す。OC ラベルを用いたファインチューニングの際には、BERT モデルの上に異なる三つの分類層を設ける。これらの分類層への入力は、基本的なファインチューニングにおける分類層への入力と同じである。これらの分類層ではそれぞれ二値分類が行われる。本研究では、これらの層から出力された各

表2 クラスに付与する分類ラベル

治療期間	NC ラベル	OC ラベル
1 週間未満	0	[0, 0, 0]
1～2 週間	1	[1, 0, 0]
3 週間～1ヶ月	2	[1, 1, 0]
1ヶ月以上	3	[1, 1, 1]

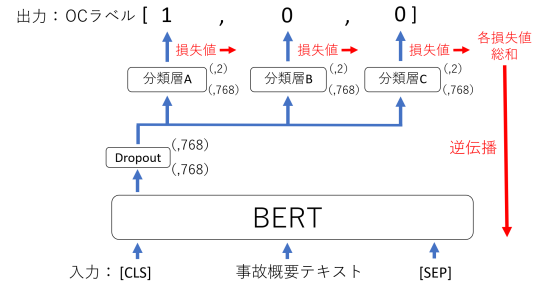


図4 OC ラベルを用いたファインチューニング

ラベルをリスト化したものをモデルから出力された OC ラベルとして扱う。また、逆伝播に使用する損失値には、各分類層の予測結果から計算された損失値の和を用いる。モデルの評価の際には、各 OC ラベル [0, 0, 0], [1, 0, 0], [1, 1, 0], [1, 1, 1] をそれぞれ各 NC ラベル 0, 1, 2, 3 に対応させる必要がある。本研究では OC ラベルにおける各要素の和を、対応した NC ラベルとして扱い評価を行う。その例を以下に示す。

- 分類層から出力された OC ラベル：[1, 1, 0]
- 対応する NC ラベルの算出：1 + 1 + 0 = 2

従って、モデルから出力された OC ラベルが [1, 1, 0] であった場合、それに対応する NC ラベルは 2 となる。

#### 4.2.3 クラスウェイト

本研究で扱う事例データは図1に示されるように、各クラス毎のデータ数が不均衡である。このような偏りは、学習データの数の多いカテゴリの分類に特化した分類モデルの作成に繋がりがねない。そこで、本研究ではデータ数の偏りによる分類性能への影響を緩和するため、分類層で計算される損失値に対して、各クラス毎の訓練データ数を考慮した重みをかける。本研究では、この重みの事をクラスウェイトと呼ぶ。本実験で使用する各クラス毎のクラスウェイトは以下に示す式により算出した。

$$\text{クラスウェイト} = \frac{\text{データ総数}}{\text{クラス数} \times \text{属するデータの総数}} \quad (1)$$

#### 4.3 「商品名称」認識モデルの作成

事故事例データにおける事故概要テキストには、閲覧者による事故理解を助ける要素が多く含まれている。例えば、「魚を食べて下痢をした」というような事例の場合、「魚」、「下痢」という要素は、その事故が食中毒に関連した事故であるという理解を助ける。しかし、BERT モデルがファインチューニングの際に、このような要素に注目しているかどうかは不明瞭である。そこで、本研究では治療期間を推定するタスクをモデルに解かす前に、事故理解を助けるテキスト中の要素を推定するタスクを解かせる。事故理解を助ける要素として、本研究では「商品



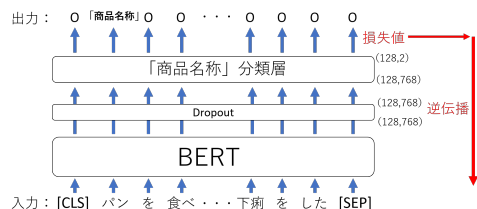


図 5 「商品名称」推定ファインチューニング

名称」に着目する。ここでいう「商品名称」とは一般的な商品名称を表すものではなく、事故の発生に関与したと考えられる物、場所、サービスの名称を表す要素である。そして、「商品名称」を推定するタスクを学習した BERT モデルのことを本研究では、「商品名称」認識モデルと呼ぶ。以下では、事故概要テキストに含まれる「商品名称」のラベル付けについて述べた後、「商品名称」を推定するファインチューニングについて説明する。

#### 4.3.1 テキストに対するラベル付け

事故概要テキストに含まれる「商品名称」は、表 1 における「商品など名称」項目に別途記載されていることが多い。そこで、本研究では「商品など名称」項目の記載情報から名詞を抽出し、それと共通の概要テキスト内の名詞に対して機械的にラベル付けを行った。その例を以下に示す。

- 事故概要テキスト：パンを食べて下痢をした
- 「商品など名称」項目：食パン
- ラベル付けされる語：パン

また、名詞の抽出には形態素解析器 MeCab を使用した。本研究では上記のような一致する名詞が無く、ラベル付けがされなかった事故概要テキストについては学習データから除外する。理由としては、一致する名詞が無い事例においても、その類義語等が事故概要テキストに含まれている可能性があるからである。

#### 4.3.2 「商品名称」の推定

まず、「商品名称」認識モデルの作成の流れを図 5 に示す。「商品名称」推定におけるファインチューニングでは、モデルから出力された全てのトークンに対応したベクトルに Dropout [9] を適用したものを分類層への入力として用いる。この分類層では、入力された各トークンベクトルに対し、それが「商品名称」か否かの二値分類を行う。そして、その損失値をクロスエントロピーロスで算出し、各層に対して逆伝播することでモデルのファインチューニングを行う。

このファインチューニングに使用するデータは基本的には傷病の治療期間推定に用いる訓練データ、検証データと同じである。しかし、事故情報データベースから収集した事例データのうち「傷病の程度」項目が「医者にかからず」であるデータも、この学習時には追加の訓練データとして扱う。

## 5 実 験

### 5.1 実験設定

本実験に使用するデータは事故情報データベースから

ら収集した事故事例データである。このうち、治療期間ラベルが付与されている 44,961 件を無作為に 8:1:1 に分割し、それぞれを訓練データ、検証データ、テストデータとして扱った。

BERT モデルへの入力形式はサブワード由来のトークン列である。このため、本実験におけるモデルへの入力テキストは同一のトークン列で表される必要がある。従って、本実験では入力テキストに対する前処理として、使用する事前学習済みモデルと同様のトークン化を行った。訓練データにおける各トークン列の入力長のヒストグラムを図 6 に示す。訓練データ内の最大の入力長は 321 であったが、他の大部分のデータにおいては 50 程度であった。本実験で使用する BERT モデルに出力されるトークン列の最大長は 128 である。このため、本実験で使用する訓練データはモデルによって特徴が正しく認識されると判断した。

本実験では以下に示す 2 種類の BERT モデルに対して、訓練データと検証データを用いてファインチューニングを行い、分類モデルの作成を行った。また、本実験ではモデルのファインチューニング時に複数の異なる手法を用いた。それらの条件も以下に示す。また、各条件における括弧内にはその略称を示す。

- 使用する BERT モデル (使用モデル)
  - － 事前学習モデル
  - － 「商品名称」認識モデル
- タスクの種類 (タスク)
  - － 治療期間ラベルのみを推定 (シングルタスク)
  - － 複数の分類ラベルを同時に推定 (マルチタスク)
- 使用ラベル
  - － NC ラベル (NC)
  - － OC ラベル (OC)
- クラスウェイト (CW)
  - － 損失値をそのまま逆伝播 (無)
  - － 損失値にクラスウェイトをかける (有)

これらの各条件で作成された分類モデルは、テストデータを用いて、その分類性能を検証した。性能の評価には正解率、Macro-F1, RMSE, 混同行列を用いた。

また、本実験でのファインチューニングにおけるハイパーパラメータは予備実験の結果と過去の研究 [1] [3] [6] [16] を参考にして以下のように定めた。

- ハイパーパラメータ
  - － batch size : 16
  - － 最適化手法 : Adam [5]
  - － 学習率 :  $2e-7$
  - － Dropout 率 : 0.5
  - － epoch : 条件により早期終了

ファインチューニングの早期終了条件は、学習の過程において、検証データにおける損失値が改善せずに 10epoch 経過した場合とした。ここではその例として、「事前学習モデル + シングルタスク + NC + CW 無」の学習時における損失値の推移を図 7 に示す。この例では、33epoch 学習を行ったモデルをテストデータでの検証に使用する。他の条件においても同様の手順により、テストデータでの検証に使用するモデルを選択した。

表 3 テストデータでの検証結果

使用モデル	タスク	使用ラベル	CW	正解率	Macro-F1	RMSE
事前学習 モデル	シングル タスク	NC	無	0.589	0.423	1.216
		NC	有	0.550	<b>0.488</b>	1.129
		OC	無	0.557	0.476	1.085
		OC	有	0.576	0.442	1.187
	マルチ タスク	NC	無	0.596	0.427	1.219
		NC	有	0.553	0.484	1.131
		OC	無	0.563	0.471	1.106
		OC	有	0.586	0.448	1.178
「商品名称」 認識モデル	シングル タスク	NC	無	0.590	0.427	1.203
		NC	有	0.539	0.481	1.119
		OC	無	0.552	0.470	<b>1.080</b>
		OC	有	0.580	0.448	1.170
	マルチ タスク	NC	無	<b>0.598</b>	0.430	1.222
		NC	有	0.540	0.474	1.131
		OC	無	0.565	0.478	1.096
		OC	有	0.582	0.444	1.190

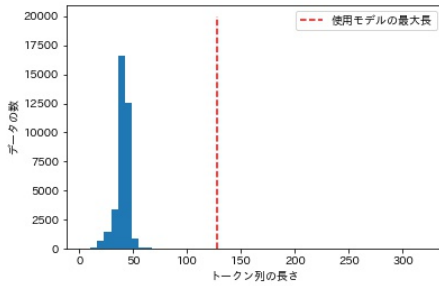


図 6 訓練データにおける入力長の分布

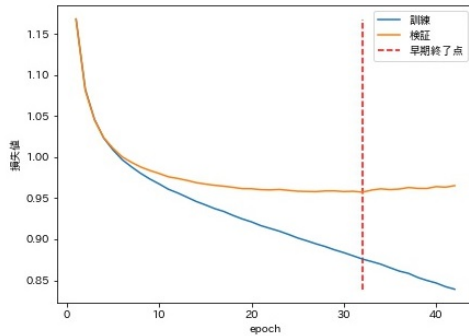


図 7 学習時における損失値の推移

## 5.2 実験結果

表 3 には、各 BERT モデルに対してファインチューニングを行った結果を示している。ベースラインに相当するモデルの分類性能は表の 1 行目の「事前学習モデル+シングルタスク+NC + CW 無」に示す。まず、使用モデルの違いによる分類性能への影響を確認したところ、それらの性能に大きな違いは見受けられなかった。シングルタスクとマルチタスクを比較した場合においても上記と同様の結果であった。しかし、使用するラベルの違いに着目した場合、OC ラベル + CW 無の条件では、他の条件よりも RMSE が小さいことが確認された。最も小さい RMSE を示した条件は「商品名称」認識モデルに対して、シングルタスク+OC ラベル + CW 無で学習を行わせた場合であった。しかし、事前学習モデルに対して同様の条件で学習を行った場合との差はわずかであったことから、RMSE の改善には OC ラベルの使用による影響が大きいと考えられる。また、RMSE は回帰問題で扱われる評価指標であるため、この評価値の改善が実際の分類結果のどのような特徴に由来しているのかを別途調べる必要がある。そこで本研究では、「事前学習モデル+シングルタスク+OC+CW 無」における混同行列の確認を行った。比較対象としてはベースラインにおける結果を用いた。それらの混同行列を図 8 と図 9 に示す。図示した混同行列において着目する箇所は、正解ラベル「1 週間未満」に対してモデルが「1 ヶ月以上」と予測したデータの数と、その逆の場合である。前者の事例の数は、ベースラインでは 216 件であるのに対し、OC ラベルを使用した場合では 121 件であった。その逆のケースでは、ベースラインモデルが 212 件、OC ラベルを用いた場合が 100 件であった。従って、OC ラベルを

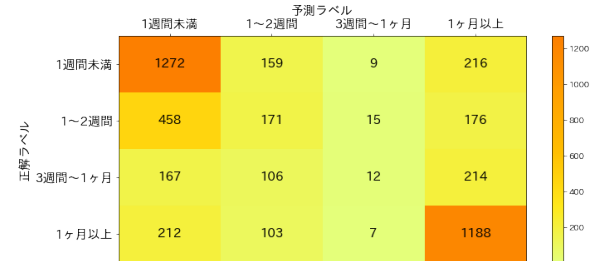


図 8 ベースラインモデルの混同行列

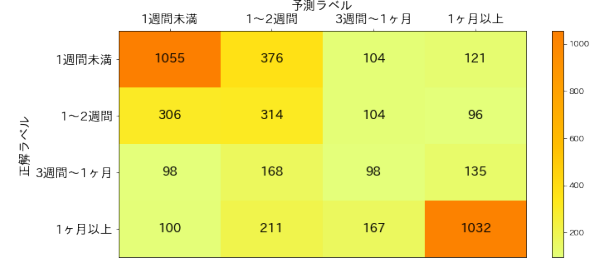


図 9 事前学習モデル+シングルタスク+ OC + CW 無

用いて作成された分類モデルは、ベースラインよりも、正解クラスとかけ離れた順序に位置するクラスに対する予測頻度が低下していることが確認された。

これらの結果をまとめると、ファインチューニングの対象となるモデルの違いや、タスクの違いによるモデルの分類性能への影響はほとんど見受けられなかった。しかし、クラスに付与したラベルの違いにおいては、正解クラスと予測クラスにおける順序間の距離に影響を及ぼすことが示唆された。

## 6 考 察

本実験で作成された分類モデルの正解率は、5~6 割程であった。これは決して良い性能のモデルとはいえない。この原因

としては、BERT への入力テキストにおける未知語の影響が考えられる。今回使用した BERT 事前学習モデルは、日本語 Wikipedia で事前学習されたモデルである。このため、本研究で扱った事故概要テキストでは、しばしば未知語が確認された。未知語の例としては「腎盂炎」や「脛」、「痒く」等が挙げられる。上記のような語は事故で発生した症状に関連した語である。このような語が未知語になることで、事故の特徴が正しく学習されなかったことが懸念として考えられる。従って、BERT 事前学習モデルに対して未知語を追加し、上記の様な特徴を正しく学習させることで性能の改善が期待できる。

本実験の結果からは、ファインチューニングの対象としたモデルの違いによる分類性能への影響はほぼ見受けられなかった。本研究において、分類層での予測クラス推定時には [CLS] トークンに対応したベクトルのみを使用している。従って、提案手法では分類の際に使用する [CLS] ベクトルに対して、「商品名称」の特徴があまり反映されなかったことが考えられる。改善策としては、[CLS] ベクトルに対して、「商品名称」に対応したベクトルの特徴をより直接的に反映させること等が考えられる。

## 7 まとめと今後の課題

本研究では、事故の概要が示されたテキストから当事者の傷病の治療期間を推定する手法を提案した。実験に使用したデータは事故情報データベースシステムから収集した事故事例データである。目的とする分類モデルは事前学習された BERT モデルをファインチューニングすることにより作成した。本研究では、このファインチューニングにおいて様々な工夫を用いて取り組んだ。その結果、異なる分類クラス間の順序を考慮した工夫によって、正解クラスとかけ離れた順序に位置するクラスへの予測頻度が低下することが示唆された。しかし、作成された分類モデルの正解率はいずれも 5~6 割程度であり、現時点では実用化には不十分である。この原因としては事故概要テキストに含まれる未知語の影響や、分類に使用した特徴ベクトルが事故を十分に表現できていなかったこと等が考えられる。従って、今後はこのような問題の解決へ向けたアプローチが必要であると考えている。前者の問題に対しては、事前学習モデルに対して未知語を追加することで改善が可能だと考えられる。後者の問題に対しては、分類の際に、入力テキスト内の他要素の特徴をさらに強く反映させることでモデルの性能の改善が期待できる。

## 謝 辞

本研究の一部は JSPS 科学研究費助成事業 JP18H03243, JP17H00762, JP18H03244 による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] Ashutosh Adhikari, Achyudh Ram, Raphael Tang, and Jimmy Lin. DocBERT: BERT for document classification. *arXiv preprint arXiv:1904.08398*, 2019.

- [2] Alan Agresti. Measures of nominal-ordinal association. *American Statistical Association*, Vol. 76, No. 375, pp. 524–529, 1981.
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina 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. 4920–4928, 2019.
- [4] Eibe Frank and Mark Hall. A simple approach to ordinal classification. In *Proceedings of the 2001 European Conference on Machine Learning*, pp. 145–156, 2001.
- [5] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the 2015 International Conference on Learning Representations*, 2015.
- [6] Jieh-Sheng Lee and Jieh Hsiang. PatentBERT: Patent Classification with Fine-Tuning a pre-trained BERT model. *arXiv preprint arXiv:1906.02124*, 2019.
- [7] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, and Gang Hua. Ordinal Regression with Multiple Output CNN for Age Estimation. In *Proceedings of the 2016 Institute of Electrical and Electronics Engineers conference on computer vision and pattern recognition*, pp. 4920–4928, 2016.
- [8] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.
- [9] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, Vol. 15, No. 56, pp. 1929–1958, 2014.
- [10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 2017 International Conference on Neural Information Processing Systems*, pp. 5998–6008, 2017.
- [11] 阿部健一, 吉田博哉. Twitter を用いた皮膚障害を引き起こす製品名等の抽出. 情報処理学会第 79 回全国大会講演論文集, pp. 529–530, 2017.
- [12] 園田章人, 井上創造, 岡賢一郎, 藤崎伸一郎. RFID を利用した救急トリアージシステムの実証実験. 情報処理学会論文誌, Vol. 48, No. 2, pp. 802–810, 2007.
- [13] 吉田傑, 長谷川卓, 富永茂, 西本哲也. 事故データによる傷害予測に関する研究. 自動車技術会論文集, Vol. 43, No. 2, pp. 275–280, 2012.
- [14] 栗原光平, 嶋田和孝. ブートストラップ法を用いた twitter からの不具合文抽出. 言語処理学会第 21 回年次大会発表論文集, pp. 341–344, 2015.
- [15] 山田侑樹, 樫山淳雄, 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 回全国大会論文集, 2020.
- [16] 柴田知秀, 河原大輔, 黒橋禎夫. BERT による日本語構文解析の精度向上. 言語処理学会第 25 回年次大会発表論文集, pp. 205–208, 2019.
- [17] 酒井浩之, 梅村祥之, 増山繁. 交通事故事例に含まれる事故原因表現の新聞記事からの抽出. 言語処理学会誌自然言語処理, Vol. 13, No. 2, pp. 99–123, 2006.
- [18] 池田伸. 自然言語処理による精神科入院患者の自殺リスク評価. 人工知能学会第 34 回全国大会論文集, 2020.
- [19] 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 回全国大会講演論文集, pp. 637–638, 2020.
- [20] 八十岡智章, 鳥海不二夫, 岡田将吾, 新田克己, 高橋久尚, 本村陽一, 田中智貴. WEB 上の相談事例とトラブルデータベースを利用した重要事案発見のための要因解析. 情報処理学会第 74 回全国大会講演論文集, pp. 649–650, 2012.
- [21] 峯崎智裕, 井上創造. 介護サービス向上に向けた介護事故事例テキストの分析. 情報処理学会論文誌, Vol. 58, No. 10, pp. 1701–1711, 2017.