

## 2機種の協働ロボット導入マニュアルへの機能分解木技術の適用 と非専門家向けマニュアル作成方法論の検討

### Investigation on Manuals for Non-Experts Based on Functional Decomposition Tree Technology for the Installation of the Two Types of Collaborative Robots

平岡 あおい<sup>1</sup> 山口 知彦<sup>2</sup> \* 笹嶋 宗彦<sup>1</sup>  
Aoi Hiraoka<sup>1</sup> Tomohiko Yamaguchi<sup>2</sup> Munehiko Sasajima<sup>1</sup>

<sup>1</sup> 兵庫県立大学 <sup>2</sup> iCOM 技研株式会社  
<sup>1</sup> University of Hyogo <sup>2</sup> iCOM Robotics Inc.

**Abstract:** The authors are studying electronic manuals for non-experts using functional decomposition trees in order to promote the use of collaborative robots, and have created manuals for non-experts using two types of collaborative robots, confirming that non-experts can perform tasks with the same accuracy as experts. Based on the findings, we are also investigating a methodology that enables experts themselves to create manuals for non-experts. This paper describes the process of manual prototyping and the comparison of manuals for two models of collaborative robots.

#### 1 はじめに

本研究の目的は、協働ロボットの作業マニュアルをモチーフとして、知識を持たない非専門家でも作業を行うことができるマニュアルを機能分解木を用いて作成すること、ならびに、そのようなマニュアルの作成方法を考案することである。

協働ロボットとは、比較的小型なロボットで、変種変量生産が可能なロボットであり、中小企業をはじめとした、製造現場の人手不足を解決する方法の一つとされている。しかし、協働ロボット等の産業用ロボットを扱うには、専用プログラムの知識やティーチングといった専門的スキルを持った人材がいないと導入が難しく[1]、現場への普及が進んでいないのが現状である。

そこで、筆者らは、協働ロボットの普及を目指して、専門家の持つ専門的知識、いわゆる暗黙知を盛り込んだ、協働ロボットを導入する企業(非専門家)が自力で作業を行うことを可能にする電子マニュアルを作成する研究を行っている。このマニュアル作成には、作業手順を構造化できる、機能分解木技術を用いている。先行研究[2]では、荷積み作業を行うロボットである「パレタイザ X」を題材として、非専門家向けマニュアルを作成した。そのマニュアルを用いて評価実験を行ったところ、非専門家でも専門家と同等の精度で、作業を実施できることが確認できた。

\*連絡先：兵庫県立大学社会情報科学部社会情報科学科  
〒 651-2197 兵庫県神戸市西区学園西町 8 丁目 2-1  
E-mail:ad23i059@guh.u-hyogo.ac.jp

一方で、こうした非専門家マニュアルの作成には、膨大なコストがかかることも判明した。先行研究[2]では、マニュアルに明記されていない暗黙知を専門家へのインタビューで明らかにしたり、それを非専門家にも分かる動画などの形でマニュアルに取り入れたりする必要があり、それらの作業のために、1年以上を要した。また、協働ロボットには、パレタイザ X の他にも多くの種類があり、新機種が出るたびに、それだけの時間をかけて、我々のような知識エンジニアがマニュアルを作り続けることは現実的ではない。

つまり、非専門家向けマニュアルを実用化するには、その作成作業自体のコストを下げ、専門家自身が、非専門家向けマニュアルを作ることができるような方法論が必要である。本研究に関連する研究の1つとして、マニュアルにあるべき構成要素を検討した、三波らの研究[3]がある。これはマニュアルの見せ方に関する研究であり、マニュアルに入れるべき要素についての研究ではない。その他にもマニュアルに関する先行研究について、筆者らが調べた限りでは、望ましいマニュアルコンテンツと、その構成方法、すなわち、マニュアル作成方法論と呼ぶべきものについての研究はない。

筆者らは、先行研究[2]で、知識エンジニアではなくても非専門家向けマニュアルを作成することができる方法論を検討し、パレタイザ X のマニュアル作成で得られた知見から、マニュアル構築ガイドラインを作成した。このガイドラインは1機種の協働ロボットの知見から作成したものであり、他の協働ロボットの機

種にも適用するために、一般性を高める必要がある。

そこで、さらなるマニュアル作成の知見を得るために、別機種の協働ロボットとして、溶接作業を行うロボットを対象としてマニュアル作成を行った。溶接は、専門テクニックや知識が必要とされる高度なタスクである。しかし、近年、溶接を行うことができる職人が不足し始めている。その問題を解決する一つの有効な方法として、職人ではない非専門家が、溶接を行う協働ロボットを使うことで溶接作業を行えるようにすることが挙げられる。本研究で扱う溶接ロボット「i ArcWeld S」は、溶接条件を適切に設定すれば、溶接に関する知識や特別なテクニックがない人でも、パネル操作をするだけで、簡単に溶接作業を実施することが可能である。しかし、非専門家は、溶接トーチを動かす速度などの溶接条件の設定ができない。また、溶接は、溶接対象の金属の材質や厚み、その日の気温等から、溶接の仕上がりが変化し、その仕上がりを見て溶接条件を試行錯誤して決定する必要があるが、非専門家は試行錯誤をどのように行えばよいかの知識がないため実施できない。

そこで本論文では、非専門家向けマニュアル構築方法論の一般化を目的として、溶接の専門家の協力のもと、専門家が持つ暗黙知を引き出し、非専門家が溶接ロボットを用いて溶接作業を実施できるようなマニュアルを。機能分解木技術を用いて作成したことについて報告する。試作した溶接作業向けマニュアルを用いて評価実験を行った後、マニュアル構築ガイドラインの改善を行った。例えば、専門家のノウハウをマニュアルのみで表現すると手順が多くなりすぎて、手順の見落としにつながることが分かったため、コアトレーニングという非専門家向けの事前講習をマニュアルと組み合わせて実施しても良い、という項目をガイドラインに追加した。

## 2 研究全体の方針

まず、研究チームで、協働ロボットで作業を行うための非専門家向け電子マニュアルを作成する。作成方法に関しては3節で紹介する。作成ができたら、そのマニュアルで作業が行えるかどうか確認するために、非専門家を被験者としたマニュアル評価実験を数回行う。その際に、実験ごとに実験条件を変更することで、マニュアルの改善をしていく。実験に関しては5節で説明する。マニュアル評価実験が一通り終わったら、変更した実験条件の中で、作業の成功に影響したものを分析し、それらを一般化して分類した後、その分類をもとに非専門家向けマニュアル構築ガイドラインを検討する。ガイドラインを作成した後に、今度は別の機種でマニュアル作成することで、ガイドラインの一般

性を高める。これを繰り返すことで、マニュアル構築ガイドラインの作成を行うが、本研究では、2機種の協働ロボットでのマニュアル作成を行い、その知見からマニュアル構築ガイドラインを作成した。そして、そのマニュアル構築ガイドラインを用いて、専門家を被験者としたガイドライン評価実験を行う。その後、実験とガイドラインの改善を繰り返すことで、マニュアル構築ガイドラインが完成する。マニュアル構築ガイドラインや、ガイドライン評価実験に関しては、6節で述べる。

## 3 電子マニュアルとその作成方法

本研究チームでは、機能分解木技術を用いて、協働ロボットを扱うための非専門家向け電子マニュアルを作成した。本節では、機能分解木の概要と、マニュアルの作成方法について説明する。

### 3.1 機能分解木の概要

機能分解とは、実現したい機能や目的を、それを達成できる部分機能の列に展開することをいい、それを木の形で表現したものを機能分解木という[4]。

本研究では、機能分解木の作成にあたり、所属研究室で開発された「FWTエディタ」を用いている。FWTエディタでは、通常の作業手順、場合分けを表す方式、作業実施時の前提条件やリスクとリスクの回避方法、複数手順から実行する手順を選択する際の適用条件を記述することができる。これらを用いて、通常の作業手順を部分行為に機能分解することで階層的に記述する。機能分解木でのマニュアルの書き方を説明するために、機能分解木でマニュアルを記述した一部を図1に示す。この図では、説明用に四角や丸囲み文字を追加しているが、本来の機能分解木にはない。図1のオレンジ色の四角で囲まれた作業を分解する際は、オレンジ色の①～④のように、オレンジ色の四角で囲まれた作業を達成するための作業を、左から順に記述する。さらに、水色の四角で囲まれた作業を分解する際は、もう1つ下の階層に、水色の①～④のように、作業を左から順に記述する。また、このFWTエディタでは、必要に応じて、各ノードに画像や動画、URLをリンクさせることもできる。

### 3.2 専門家からの暗黙知獲得の難しさ

専門家から暗黙知を獲得することは困難である。本研究でいう「暗黙知」とは、暗黙的な知識や技能の中で、専門家しかできない特殊なテクニックではなく、内容や方法を見たり聞いたりすれば誰でも理解可能な知識のことをいう。そのような簡単な知識や技能を引き

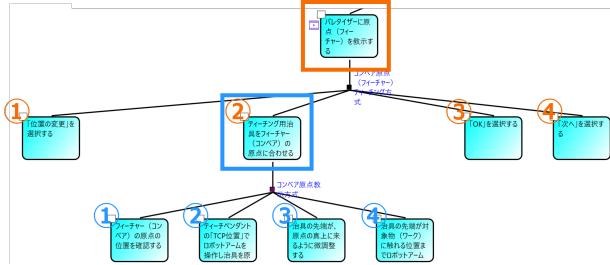


図 1: 機能分解木で記述されたマニュアル(一部)

出すのが何故難しいのか、理由は大きく3つある。1つ目は、専門家の保有している知識の多さである。専門家は、その分野において、知識を多く保有しており、状況に応じて適切に使い分けることができる。しかし、それゆえに、なにも準備せずに、いざ作業をマニュアルに書き出すとなると、専門家はどこから書き始めてよいのか分からず、作業が進まなくなる。2つ目は、専門家自身が暗黙知を正確に認識していないことである。専門家が持つ暗黙知は、これまでの経験や現場での工夫などの積み重ねから、自然に獲得されたものが多く、専門家自身が認識していないものも多い。このため、何も支援がない状況で、専門家自身が認識していない暗黙知を記述することは困難である。3つ目は、専門家の暗黙知をアウトプットしたことがないからである。専門家は暗黙知を無意識的に活用しており、それを他者に説明するために言語化する経験がほとんどない。その結果、自分が自然に行っている作業手順や考え方を、明確に他者に教えることが苦手といわれている。

専門家の暗黙知を獲得し保存する方法としては、専門家による作業過程を動画で記録する方法が考えられる。しかし、単に作業過程を撮影しただけでは、知識がない非専門家が理解することが難しく、専門家の暗黙知を伝えることが難しいことが分かっている[5]。

機能分解木を用いてマニュアル作成を行うことで、専門家から暗黙知を獲得することが可能になる。何故なら、作業手順を構造化することで、作業どうしの関係性が視覚的になり、どの作業が曖昧なのか、すなわちどの作業が暗黙知となっているのかが分かりやすくなるからである。以上の理由から、本研究では、機能分解木を用いてマニュアル作成を行う。

### 3.3 電子マニュアルの作成方法

ここでは、非専門家向けマニュアルの作成の手順について説明する。はじめに、現場で使用されている専門家向けマニュアルを、そのまま機能分解木で記述する。次に、専門家が使用する作業マニュアルに書かれている作業行為を実際の作業現場で見学し、専門家から作業行為の説明を受ける。その聞き出した作業の詳

細を、筆者らの手で機能分解し、最初に作った機能分解木に追加・修正を行う。機能分解木を作っていく中で、曖昧な箇所が出てきたら、専門家に質問して、ノードを追加する。これを繰り返すことで、機能分解木を完成させ、FWTエディタの機能により、機能分解木の形からマニュアルの形に変換する。

また、作成したマニュアルには、文字の説明を補う、画像や動画もリンクしている。作業内容を文字で説明しただけのマニュアルでは、用語の意味が分からぬために作業内容が分かりにくいものがあったり、手順が正しく実行されたのか確認できなかったりするため、非専門家向けマニュアルには適していない。そのため、注目すべきポイントにフォーカスした撮影をした画像や動画を、マニュアルにリンクすることで、文字の説明を補うことができる。この補完動画は、作業単位で撮影して編集を行った、1本あたり30秒~4分の動画である。

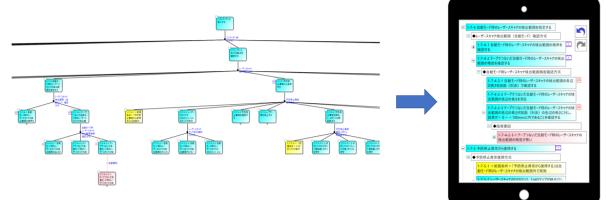


図 2: 機能分解木をマニュアルの形にするイメージ図

## 4 マニュアル作成方針

最初に、我々研究チームが考える、非専門家向けマニュアルの定義について述べる。また、マニュアル作成の対象とした2機種の協働ロボットの種類や、マニュアルの作成方針等を比較したものを表1に示し、4.2節以降で説明する。

### 4.1 非専門家向けマニュアルの定義

我々が目指している非専門家向けマニュアルは、非専門家自身がマニュアルを見て作業をスムーズに行うことができるマニュアルである。その作成において最も重要なのは、マニュアルを参照して作業する非専門家の安全を確保することである。作業についての知識があまりない非専門家は、危険性やしてはいけない行為が分からず、怪我や事故に巻き込まれる可能性が大きいにある。そこで、非専門家向けマニュアルを作成する際には、作業に潜む危険性や、必ずしなければならない行為、もしくはしてはならない行為を記述し、作業者が安全に作業できるようにしなければならない。その上で、作業者が専門的知識や経験を持たなくても、マニュアルを参照することで作業をスムーズに行うことができるマニュアルであることが必要と考える。

また、作業マニュアルは、効率的な作業の実現を目指すものでなければならない。マニュアルを見た作業者が作業内容について悩んでしまったり、間違った行為をしてしまったりするのは、作業を進めるのに非効率である。しかし、マニュアル利用者は知識のない非専門家なので、思わぬ行動をとったり、マニュアルを読み間違えたりして、作業が思うように進まない、もしくは作業が実施できない可能性がある。そこで、非専門家向けマニュアルは、マニュアルを読んだり、添付されている動画や画像を見たりすることで作業内容を理解し、作業者が正しく作業を実施できて、かつ、あまり迷わずスムーズに作業を進めることができるようなマニュアルを目指すべきであると考える。

先行研究、ならびに本研究では、この定義をもとに非専門家向けマニュアルの作成を行った。また、対象とした協働ロボットのタスクや作業の特徴などによって、その協働ロボットの作業に合ったマニュアルとなるよう、マニュアルの作成方針を変えて作成を行った。

## 4.2 先行研究：パレタイマー X

先行研究 [2] で扱った「パレタイマー X」というロボットは、共同研究先である iCOM 技研株式会社で開発された協働ロボットで、ベルトコンベアで流れてきた荷物をアームで持ち上げ、パレットという板に荷積みするタスクを行う。このロボットを現場で使用する前には、他の協働ロボットと同様にロボットを使える状態にする導入作業を行うが、この作業の中には、専門知識を要する作業がいくつか存在する。そこで、先行研究 [2] では、パレタイマー X の導入作業をマニュアル作成対象とし、専門家が持つ導入作業に関しての暗黙知を盛り込んだ、事前知識のない非専門家自身が導入作業を実施できるようなマニュアルを作成した。この導入作業は一度完了したら手戻りがない作業であり、マニュアルを最初から順に進めていけば作業が完了する。

また、共同研究先では、コアトレーニングという、協働ロボットを導入する人向けの事前講習を実施している。コアトレーニングでは、協働ロボット全機種共通の知識を 2 日間にわたって教育する。コアトレーニングには、個別の協働ロボットに関する内容は含まれておらず、先行研究では、マニュアルにパレタイマー X に関する知識を盛り込む必要があった。

表 1: 先行研究と本研究との比較表

	パレタイマー X (1 機種目)	溶接ロボット (2 機種目)
協働ロボットタスク	荷積み	溶接
マニュアル利用者	事前知識のない非専門家	
マニュアルの種類	導入作業	溶接作業
作業の特徴	手戻りが少ない	試行錯誤が必要
コアトレーニング	全機種共通知識	溶接に特化

## 4.3 本研究：溶接ロボット

現在研究で扱っている溶接ロボット「i ArcWeld S」は、パレタイマー X と同様、iCOM 技研株式会社で開発された協働ロボットで、金属を溶接するタスクを行う。パレタイマー X と違って導入作業は簡単であり、かつ、溶接作業のための操作自体は簡単であるが、溶接作業では、溶接トーチから出る電流の大きさや、動かす速度などの溶接条件を適切に設定する必要がある。こうした溶接条件は、溶接する金属の種類や厚さや、作業を行う環境の気温などに影響を受けるため、その設定には試行錯誤が伴う。溶接の知識がない非専門家にとってこれは非常に難しい。そこで、溶接ロボットを用いた溶接作業をマニュアル対象作業とし、非専門家自身が試行錯誤しながら溶接条件を設定し、溶接作業を成功させられるようなものを目指してマニュアル作成を行った。本研究では、はじめに、種類がいくつか存在する溶接作業の中から、比較的簡単とされる、TIG 溶接の直線溶接を対象とした。

溶接作業中に発生する試行錯誤については、判断の目安となる情報を加えることで対応した。マニュアルに目安表や画像を添付することで、非専門家でも判断できるようにした。試行錯誤して溶接条件を設定するといつても、知識がない非専門家は、最初どれくらいの数値にすればよいのかの目安も分からず。そこで、共同研究先から提供された、図 3 の左側にある、溶接条件に関する目安表をマニュアルに添付した。この目安表には、金属の種類や、板厚、溶接の向きごとに目安となる溶接速度や溶接電流の値が記載されている。非専門家はこの表を目安に数値設定をする。設定ができた後、その設定で溶接を行い、自分で溶接した物と、マニュアルに添付している溶接が成功した写真とを見比べて、溶接が成功しているかを判断する。もし、失敗していると判断した場合は、マニュアルの該当手順に戻り、失敗した原因に応じて溶接条件を調整し、再度溶接を行う。上記のように試行錯誤を手順化することで、非専門家でも溶接作業を実施できるようにした。



図 3: 溶接条件を試行錯誤するイメージ図

また、溶接ロボットのコアトレーニングは、溶接の内容に特化させたものにした。コアトレーニングの受講自体はパレタイマー X でも同様なのだが、前述した通り、全機種共通知識を教えるものだったため、マニュアルにパレタイマー X 固有の知識を取り込む必要があつ

たため、作成したマニュアルの手順数が膨大となった。先行研究[2]では、手順数が膨大になると、マニュアルを読み飛ばす失敗が起きることが分かった。そこで、溶接ロボットのコアトレーニングでは、溶接作業に特化した内容にした。協働ロボットを扱う上で必要な知識はもちろん、溶接方法や溶接の仕上がりの見方などの溶接に関する知識や、溶接ロボットのソフトウェアについての操作説明など、溶接ロボットを扱う人に絞った内容とした。この事前講習で教わる知識についてはマニュアルから除き、説明が煩雑化してマニュアルが膨大になることを防ぐことを狙いとした。さらに、パレタイザーXのコアトレーニングには、全機種共通知識を広く教えるため、2日間にわたって実施していたが、溶接に特化したコアトレーニングは、内容を絞ったことで、約2時間という短時間で溶接の知識を教えることができるというメリットもある。

## 5 マニュアル評価実験

筆者らが作成した非専門家向けマニュアルを使用して、非専門家を被験者とした、パレタイザーXと溶接ロボット、それぞれ数回実験を実施した。実験は、毎回少しづつ実験条件を変えて行い、それらの変更点が、被験者の作業実施にどう影響したかについて、比較し、分類した。本節では、それぞれの実験の実施概要や、各実験での主な変更点や分類について述べる。

### 5.1 全体の実験実施概要

ここでは、先行研究、ならびに本研究で行った評価実験で、共通している実施概要について説明する。

被験者は、事前知識がない非専門家とし、我々が作成したマニュアルを参照しながら、協働ロボットの作業を実施する。マニュアルは、Microsoft社のSurface Go 3のタブレットに表示した。

実験では、実験開始直前に被験者に対して、実験の実施概要と電子マニュアルの操作方法などについて、A4の紙を数枚配布し、教示を行った。実験中は、我々研究チームが立ち合い、ビデオカメラ2台で作業の様子を撮影記録した。また、実験の全ての工程について、被験者が作業を行っている付近に専門家を配置し、被験者に何らかの危険が予知された場合には、安全確保を最優先として実験に介入することとした。もし、被験者が間違った行動をしていると分かったとしても、被験者に危険が及ぶ可能性やロボットが壊れる可能性がある場合を除いて、作業手順に関する指示や指導は行わないようにした。さらに、作業中に不明点が生じた場合は、被験者の好きなタイミングで何度も、zoomによる専門家の遠隔支援が受けられることとした。被

験者は、webカメラで現場の様子を見せながら、zoomを通じて、専門家に質問することができる。実験後には、被験者インタビューを実施した。

作業は、マニュアルと遠隔支援によって、作業を終えることができたかどうかで評価を行った。遠隔支援に関係なく、必要な精度を達成して手順を完了をしていた場合は、正しく作業手順を完了したとみなした。逆に、必要な精度を達成せずに手順を完了した場合や、手順を読み飛ばしていたり、未完了のまま中断したりした場合は、作業が失敗したと評価した。

インタビュー、作業評価までが終わった後に、各実験で行った実験条件の変更点の中で、作業の成功に影響した要因を分析し、それらを一般化して分類した。

### 5.2 先行研究：パレタイザーX

先行研究であるパレタイザーXの非専門家向けマニュアルの評価実験は、予備実験1回と本実験5回の計6回実施した。ここでは、先行研究での実験の実施概要と、その結果について簡潔に述べる。詳しくは[2]や[6]を参照されたい。

予備実験では、本研究チーム2名が、非専門家向けマニュアルを参照しながら、パレタイザーXの導入作業を実施し、実験実施に関する安全確認と、作成した非専門家向けマニュアルの確認を行った。

本実験では、パレタイザーXの導入作業の中でも主要な作業である、「レーザースキナの動作を確認する」「電気信号の入出力を確認する」「ティーチングを実施する」の3つの作業を実験タスクとした。対象作業のマニュアル手順数は894手順である(全実験終了時点)。各実験で被験者は2名とし、2名1組で実験タスクに取り組ませた。

毎回の実験で、マニュアル内容等の実験条件を変更し、マニュアルや実験環境の改善を行った。その結果として、非専門家自身が、マニュアルと遠隔支援を使用することで、パレタイザーXの導入作業を行うことができることを確認した。

また、作業の成功に影響を及ぼした変更点を分析し、分類をした結果、8つの分類をすることができた。その1つに「多角的な表現」がある。協働ロボットの導入作業においてティーチングという作業では、ロボットアームが稼働するにあたって基準となる位置を、誤差を数mm以内におさめなければならない。そこで、ティーチングが要求する精度が誤差が数mm以内であるということを、文字だけではなく、図で多角的に表現することで、ティーチングの精度を改善できた。これを一般化すると、作業の成功を直接表現するパラメータとそれにかかる作業については、マニュアルで多角的に表現すること、となり、ガイドラインの1つとしている。

### 5.3 本研究：溶接ロボット

本研究の溶接ロボットマニュアル評価実験は、予備実験1回と本実験2回の計3回実施した。ここでは、実験の実施概要に加えて、作業成功に有効だった変更点の例を説明する。

#### 5.3.1 予備実験

本実験を行う前に、実験タスクの確認と、作成した非専門家向けマニュアルの最終確認のために、マニュアルを作成した筆者が、実際に、溶接ロボットマニュアルを用いて、溶接作業を行った。筆者は、専門家の作業風景を見学したことはあったが、実際に作業するのはこの時が初めてであった。作業した結果、溶接を成功させることができたが、マニュアルに改善すべき点が見られたので、本実験に向けて修正を行った。

#### 5.3.2 本実験実施概要

本実験では、溶接ロボットや溶接機の電源を投入し、溶接作業を行い、電源を落とすまでの、溶接ロボットを使用する一連の作業を実験タスクとした。また、溶接は、初心者でも実施しやすい、角継手の片引きをすることとした。本稿執筆時点で、マニュアル手順数は104手順である。1回の実験につき、被験者を2名募集し、約2時間の専門家(iCOM技研スタッフ)によるコアトレーニングを受講した後、マニュアルを参照しながら、1名ずつ溶接作業を実施させた。さらに、溶接作業では、失明や火傷する恐れがあるので、溶接面や手袋を使用するよう、実験前に説明した。

#### 5.3.3 実験結果

第1回と第2回の実験結果を表2, 3に示す。被験者4名とも、間違って作業完了させたり、作業を読み飛ばしてしまったり、などのいわゆる作業失敗をすることなく、すべての手順を正しく完了させることができた。このことより、溶接に特化したコアトレーニングとマニュアルによって、非専門家自身が溶接ロボットで溶接作業を実施できることを確認した。また、第1回実験からコアトレーニングやマニュアル内容を改善したことで、第2回実験の方が遠隔支援の回数が減少していることが分かった。しかし、第2回の方が短縮されると予想していた所要時間については、第1回実験と大きな差は見られなかった。これは、被験者の性格によるものではないかと考えている。被験者Bは、マニュアルに添付されているの動画を一通り視聴してから作業していたので、所要時間が長くなった。それに対し、被験者AやDは、作業中に分からぬことがあった時だけ動画を視聴していたため、所要時間が短い結果と

なった。慎重な作業者があまり頻繁に動画を参照しなくても済むように、コアトレーニングと作業時に視聴する動画をなるべく同じようにして安心させる方法や、動画を効率よく視聴できるように分割の仕方を工夫するなどの方法を考えている。

表2: 第1回評価実験結果

被験者	正しく作業完了		作業失敗	所要時間
	遠隔なし	遠隔あり		
A	90手順	2手順	-	1時間 6分
B	83手順	9手順	-	1時間 58分

表3: 第2回評価実験結果

被験者	正しく作業完了		作業失敗	所要時間
	遠隔なし	遠隔あり		
C	104手順	-	-	1時間 36分
D	102手順	2手順	-	1時間 10分

#### 5.3.4 第1回評価実験からの変更点

第1回実験で被験者が作業進行に詰まったり、遠隔支援を利用したりした手順を抽出し、第2回に向けてコアトレーニングやマニュアルの内容を変更した。第2回実験終了後、作業の成功に影響した変更点を分析し、それらを一般化して分類した。溶接ロボットマニュアル評価実験で得た分類と、それに対する変更点の例を表4に示す。以下では、主な分類を2つ紹介する。

##### 分類「コアトレーニング」

まず、1つ目は、「コアトレーニング」である。iCOM技研株式会社では、今まで協働ロボット全般知識を教えるコアトレーニングしか実施したことがなく、とある作業に特化した内容で実施するのは初の試みであった。そのため、第1回実験時のコアトレーニングは、アームの動かし方やロボットの設定方法など、協働ロボット全般知識について説明を約55分、溶接ソフトの使い方や溶接ロボットでのティーチング方法など、溶接ロボットの知識についての説明を約47分で実施した。その結果、被験者がコアトレーニングで習った内容を使わずに作業したり、習った内容を遠隔支援で質問したりすることが起こった。また、第1回実験後の被験者インタビューで「コアトレーニングと実際の作業とのギャップがあり、作業中に戸惑った」という声があがった。

この結果を受けて、第2回実験では、溶接に関する内容を増やし、協働ロボット全般知識についての説明を約38分、溶接ロボットの知識についての説明を約1時間3分で実施した。また、コアトレーニングの最後に溶接ロボットを直接見学する時間を約15分設けた。その結果、被験者は、マニュアルには記載していない、コアトレーニングで習った知識を活かして作業した。第2回実験後の被験者インタビューでは「時間配分に関しての不満はなく、不要な知識もなかったと感じる」という意見が出た。

表 4: 分類とそれに対する変更点の例

分類	効果があった変更点の例
具体的な指示	動画の字幕に書いてある内容をマニュアルにも記載する(設定をどう確認するのか等)
作業手順	仮止めの順番や溶接トーチの角度をマニュアルで指定
リンクコンテンツ	画面の中で確認すべき箇所を、動画や画像で四角で強調表示
配布資料	目安表の中で、今回使用するロボットにおいて不要となる部分を削除
コアトレーニング	溶接に関する内容の割合を多くし、溶接ロボットを見学する時間を設けた

のことから、このコアトレーニング内容の変更点は作業成功に影響したと判断し、分類「コアトレーニング」とした。

#### 分類「具体的な指示」

2つ目は、「具体的な指示」である。第1回実験時のマニュアルは、コアトレーニングや添付動画の内容はあまり詳しく記載せず、作業の粒度を粗めに作成した。しかし、それが原因で、被験者が具体的な作業内容を理解できず、遠隔支援を何回も利用することになった、また、作業内容を把握するために、毎回動画を視聴する必要が出てくるため、余計に作業時間を要してしまうことが分かった。

そこで、第2回実験時のマニュアルでは、動画の字幕に書いてある作業、言いかえれば、字幕に埋め込むことでマニュアルから除いていた手順を、マニュアルにも作業手順として記載するようにした。その分手順数が増えてしまったが、結果として、マニュアルの粒度が小さくより適切な大きさとなり、被験者が作業内容を理解できるようになって、被験者自身のみで作業を達成することができた。

のことから、この作業の粒度を適切な細かさにするという変更点は作業成功に影響したと判断し、分類「具体的な指示」とした。

## 6 マニュアル構築ガイドラインとその評価実験

前節で述べた実験に基づいて、専門家自身が、非専門家向けマニュアルを作成できるような、マニュアル構築ガイドラインを、先行研究で作成したものから改善した。本節では、そのガイドラインの紹介と、ガイドラインの評価実験の実施概要の説明を行う。

### 6.1 マニュアル構築ガイドライン案

下記に、作業を成功に導くようなマニュアルを機能分解木で構築するためのガイドラインを示す。ガイドラインの各項目の後の括弧内には、そのガイドラインの根拠となった、変更点の分類を示している。このガイドラインは、協働ロボットマニュアルをモチーフとして作成したが、対象は協働ロボットマニュアルに限

らず、作業マニュアル全般に使われることを想定している。また、マニュアルの作成手順を記載しているのではなく、非専門家向けマニュアルを作成する上での心得のようなものを表している。なお、このガイドライン利用者は、機能分解木を用いて、作業手順を表現することは出来るものとする。

#### マニュアル構築のガイドライン案：

1. 非専門家向けマニュアルは、作業者の安全を守るのは大前提として、その上で、非専門家自身がマニュアルを見て作業をスムーズに行うことができるマニュアルを目指す。
2. 作業マニュアルは、1回目は危険回避、2回目は機能的な問題、3回目以降は分かりにくい言葉の追記など、機能分解木を用いて数回試作するものである。被験者による利用実験を複数回行うことでき、さらに改善を行うこと。
3. 作業者の安全を守るための手順と、専門家と非専門家で定義が異なる言葉については必ず書く。(マニュアルコンテンツ)
4. 指示の粒度は荒すぎず細かすぎず、適切な細かさにする。実験を行い、被験者が迷った手順については、指示の粒度を小さく、具体的にする。(具体的な指示)
5. 作業をより行いやすくする手順についても、記載を心がける。(作業手順)
6. 作業を達成する方法が複数あり、どの方法でもよい場合、非専門家が迷わないよう、「どちらでもよい」という曖昧な表現は避け、どれか1つの方法を推奨する。(作業手順)
7. 作業全体の精度を決定づけるような重要な作業については、マニュアルの指示を多角的にする。(多角的な表現)
8. その作業で考えられるエラーや、その手順を実施しなかった場合に起こる不具合は、阻害要因としてマニュアルに示す。(阻害要因の提示)
9. 文面だけで説明しにくい作業は動画や画像コンテンツで説明を補う。その際、マニュアルの内容と相違がないよう心がける。特に注目すべき箇所は強調する。(リンクコンテンツ)
10. 作業を行うために必要な、部品の名称や基礎的な作業の仕方など周辺知識についての資料は、マニュアルもしくはマニュアルに付属させるドキュ

- メントに含める。なお、ドキュメントは作業に必要な情報のみが載っているものにする。特に注目すべき箇所は強調する。(配布資料)
11. 作業をスムーズにすることが見込めるのであれば、マニュアル周辺機器についても指示をする。(タブレット周辺機器)
  12. 初期の実験で、マニュアルが見づらい、読みづらいなどの課題が明らかになった場合は、マニュアルのインターフェースそのものを変更する可能性を考える。(タブレットインターフェース)
  13. 現場のマニュアルが常に正しいとは限らない。実験を重ねて、現場マニュアルの誤りが発覚した場合には、マニュアルそのものを修正するとともに、現場にフィードバックする。(マニュアルコンテンツ)
  14. マニュアルを補う目的でコアトレーニングを実施してもよい。コアトレーニング(事前講習)を行う場合、実施する予定の作業に即した内容にする。座学だけでは伝わりにくい場合、実際に作業しながら説明を行う。(コアトレーニング)

## 6.2 ガイドライン評価実験

上記で示したガイドラインを用いて、専門家に、非専門家向けのマニュアルを作成させる、マニュアル構築ガイドラインの評価実験を現在実施している。その実験の実施概要について説明する。

被験者はiCOM技研株式会社のスタッフ1名で、溶接コアトレーニングの講師を担当している人である。この被験者に、マニュアル構築ガイドラインを参照しながら、機能分解木を用いて、溶接コアトレーニング実施のマニュアルを作成してもらう。作成するマニュアルの仮定としては、マニュアル利用者は溶接ロボットの知識はあるが、コアトレーニングの講師をしたことがない非専門家を想定し、マニュアルを見て、溶接コアトレーニングを最後まで実施することができるものである。被験者に2週間程度で作成を依頼しており、作成終了後、評価をする予定である。

## 7 おわりに

本研究では、機能分解木を用いたマニュアル構築ガイドラインの一般化を目的として、2機種目となる溶接作業を行う協働ロボットの非専門家向けマニュアルの試作を行った。1機種目のパレタイザーXのマニュアルと違い、コアトレーニングを溶接に特化した内容にしたり、溶接条件を試行錯誤する手順をマニュアルに反映させたりした。その結果、計3回のマニュアル評価実験から、溶接コアトレーニングとマニュアルで、非専門家でも溶接作業が実施できることを確認した。ま

た、実験を行う中で、実験条件を変更し、作業の成功に影響した変更点を分類した。その分類をもとに、先行研究で作成したマニュアル構築ガイドラインを改善した。

今後は、マニュアル構築ガイドラインの評価実験として、専門家がガイドラインを参照しながら作成したマニュアルを、我々が見て、コアトレーニングが実施できそうかどうかでマニュアルの評価を行う予定である。その後、被験者にインタビューを実施し、マニュアルの評価と合わせて、ガイドラインの有効性を確認したいと考えている。

また、このガイドラインは、機能分解木が記述できることを前提としているが、機能分解木の作成方法に關しては、未だにあまり手順化されておらず、機能分解木を利用したことがない人にとっては、マニュアルどころか、そもそも機能分解木を作成することが難しい。そこで、本研究チームでは、機能分解木構築ガイドラインも検討している。これと本研究のガイドラインを組み合わせることで、専門家が簡単に非専門家向けマニュアルを作成できると考える。

## 参考文献

- [1] 一般社団法人日本ロボット工業会, ロボット産業ビジョン 2050 Ver.0, [https://www.jara.jp/publications/img/vision/visionver0\\_booklet.pdf](https://www.jara.jp/publications/img/vision/visionver0_booklet.pdf), (2023年5月刊行)
- [2] 平岡あおい, 他, 機能分解木を用いたマニュアル作成方法論についての一検討-非専門家による協働ロボット導入作業の適用事例-, 第37回人工知能学会全国大会予稿集, 3R1-GS-3-05, 2023
- [3] 三波千穂美, 他, マニュアルの構成要素から見た高齢者向け携帯電話マニュアルの現状と課題:高齢者の身体・心理的特性から想定される問題への対応に関して, 情報メディア研究, 12巻, 1号, pp.14-27, 2013
- [4] 來村徳信, 他, オントロジーに基づく機能的知識の体系的記述とその機能構造設計支援における利用, 人工知能学会論文誌, 17巻, 1号, 2002
- [5] 笹嶋宗彦, 知識工学の手法を用いた協働ロボット導入支援の試み, 人工知能学会誌, Vol.37, No.3, pp.292-298, 2022
- [6] 松本健汰, 他, 協働ロボットの遠隔での非専門家による構築を実現するための専門家ノウハウの収集とマニュアル化の検討, 第36回人工知能学会全国大会予稿集, 4M3-GS-3-02, 2022

# ChatGPTを用いた対話相手の好感度を高められるコミュニケーション支援システム

## Communication Support System for Enhancing Dialogue Partner Likability using ChatGPT

野呂 悠斗<sup>1\*</sup> 砂山 渡<sup>2</sup> 服部 峻<sup>2</sup>  
Yuto Noro<sup>1</sup> Wataru Sunayama<sup>2</sup> Shun Hattori<sup>2</sup>

<sup>1</sup> 滋賀県立大学大学院工学研究科

<sup>1</sup> Graduate School of Engineering, The University of Shiga Prefecture

<sup>2</sup> 滋賀県立大学工学部

<sup>2</sup> School of Engineering, The University of Shiga Prefecture

**Abstract:** While the increase in the number of SNS users has made it easier to interact with others, messages and posts that hurt or offend others have become a problem. Some of them are objectively understandable, such as slander, while others are personally offensive, even if generally unproblematic. In addition, few existing communication support tools directly modify input comments. In this study, we aim to realize better communication by using ChatGPT, a system that, in addition to objective criteria, extracts dialogue partners based on the dialogue history with them and presents a sentence modification plan tailored to the dialogue partner.

### 1 はじめに

近年、SNS等のサービスの利用者が増加傾向にある。総務省による調査[1]では、日本のソーシャルメディア利用者数は2028年には1億1360万人になると予測されている。それに伴い、SNSでのコミュニケーションの機会が増加している。SNSの利用におけるメリットとして、他者と簡単に対話でのコミュニケーションが取れることが挙げられる。しかしデメリットとして、他者を傷つける、あるいは不快にさせるようなメッセージが書き込まれる可能性がある。このような書き込みの例として、第一に挙げられるのが誹謗中傷であり、客観的に適切でないものだと分かる場合が多い。しかし、対話相手や対話する二者の関係、状況次第では、客観的に見て問題のないようなメッセージであっても相手を不快にする、あるいは傷つける場合がある。そこで本研究では、ChatGPTを利用して、客観的な基準と、対話履歴から抽出した相手が好感あるいは不快感を示す文章の特徴の2つを用いて、二者の対話における、ユーザが送信しようとするメッセージについて、必要に応じて対話相手に合わせた文章の修正案を提示するシステムを提案する。

### 2 関連研究

対話相手に対して持つ好感の程度から、自身の感情を制御し、人間らしい返信を行う研究がある[2]。この研究では、対話の質問文や返答文からロボットが対話者に抱く好感度を推定し、そのときの感情を決定し、その感情に対応する会話文を発話するシステムが提案されており、より人間らしい感情表現するロボットの実現を目指している。しかし、この研究では、相手のロボットに対する怒りや苦しみなど、負の感情による発話が行われる可能性がある。本研究では、相手を不快にさせる可能性のある文章を、実際に発信する前に修正を促すことを目的としている。

パーソナライズされた嗜好プロフィールに関する研究[3]や、嗜好プロフィールに対してユーザが不快に感じるコンテンツをフィルタリングする研究がある[4]。これらの研究では、インターネット上のプラットフォームでの閲覧履歴をもとに、ユーザの嗜好を作成するシステムや、ユーザを不快にするレコメンドを嗜好プロファイルからフィルタリングするシステムが提案されている。本研究では、対話履歴をもとにしたプロファイルの作成を行うことで、対話相手から的好感度を上げることを目的としている。

発話に含まれる品詞の関係から、名詞に対する感情を推定する研究[5]や、発話の構文情報や音響的情報を

\*連絡先：滋賀県立大学大学院工学研究科電子システム工学専攻  
〒 522-8533 滋賀県彦根市八坂町 2500  
E-mail: on23ynoro@ec.usp.ac.jp

もとに発話対象の好感度を計算する研究[6]がある。これらの研究では、構文情報をもとに好感度や感情の推定を行っているが、文章全体の内容を把握することは難しい。本研究では、ChatGPTを用いることで文章全体の意味を理解して好感度推定を行う。

### 3 好感度を判定する基準

本章では、対話相手が発話者に対して抱く好感度を判定する基準について述べる。

#### 3.1 客観的基準

本研究における客観的基準として、ユーザが入力した文章に対して対話相手がどう感じるかを、対話相手に依らない共通の指標を用いる。Horshamらの研究[7]によると、他者への好奇心と好感度には正の相関があることが示されている。また、他者に対する好感度の変化と相手への理解度の関係に関する研究[8]によると、対話相手が自身に対して理解していると認知することで好感度が高まるとしている。本研究では、「共感」、「関心」、「好意」、「ポジティブ」、「親しみやすさ」の5つの要素について判定を行う。「関心」については、その内容に対して否定的である場合も想定されるため、「ポジティブ」という要素を加える。また、SNSでの利用を想定し、親しみやすい文章である必要があるため、「親しみやすさ」という要素を加える。

#### 3.2 主観的基準

本研究における主観的基準として、ユーザが入力した文章に対して対話相手がどう感じるかを、対話相手の特徴による指標を用いる。対話履歴をもとに、ユーザが入力した文章に対して対話相手が示す反応を、あらかじめ対話履歴から抽出した4つの特徴「p(ポジティブ)」、「n(ネガティブ)」、「e(ニュートラル)」、「a(無関心)」をもとに判定する。

## 4 対話相手に合わせたメッセージ修正案提示システム

本章では、提案する対話相手に合わせたメッセージ修正案提示システムについて説明する。「LINE」等の1対1での対話を想定しており、ユーザの入力するメッセージが相手を傷つける、あるいは不快にさせるかどうかを2つの観点からChatGPTにより判定し、それをもとに修正案を提示する。

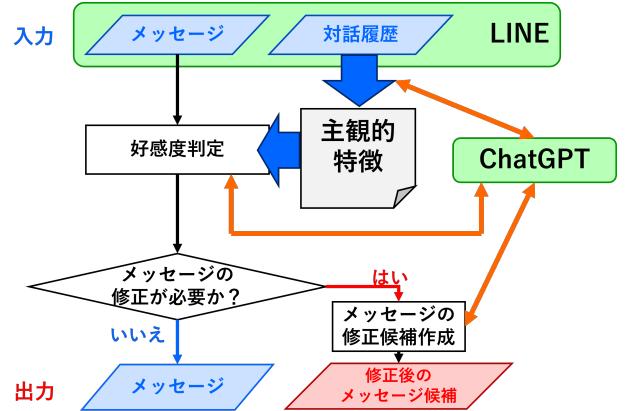


図1: 対話相手に合わせたメッセージ修正案提示システムの構成

#### 4.1 対話相手に合わせたメッセージ修正案提示システムの構成

図1に、対話相手に合わせたメッセージ修正案提示システムの構成を示す。本システムは「LINE」などの1対1でのテキストでのメッセージのやり取りを想定している。

ユーザが送信しようとしているメッセージと、ユーザと対話相手との対話履歴を入力データとしてシステムに入力する。まず、対話履歴から、対話相手がポジティブ、ネガティブ、無関心に感じるメッセージの主観的特徴を抽出する。その後、客観的基準、および主観的基準の特徴から、ユーザが送信しようとしているメッセージについて修正が必要かどうかを判定する。修正が必要とされた場合、いくつかの修正案をユーザに提示し、ユーザが必要に応じて選択する。

#### 4.2 主観的基準の特徴抽出

ユーザと対話相手の対話履歴から、対話相手がポジティブネガティブ、あるいは無関心に感じるメッセージの特徴を抽出する。

##### 4.2.1 対話履歴のメッセージのポジネガ分類

ユーザのメッセージと、それに対する対話相手のメッセージを1組のデータとし、これらを表1のプロンプトに挿入し、ChatGPTに入力することで、対話相手のメッセージの内容から、ユーザのメッセージが、「p(ポジティブ)」、「n(ネガティブ)」、「e(ニュートラル)」、「a(無関心)」のどのグループに当たるかを判定し、分類する。

表 1: 対話相手のメッセージから関心・無関心、およびポジネガを判定するプロンプト

	プロンプト
無関心の判定	<p>文章から心理分析を行うプロとして、入力された&lt;Aさん&gt;の発言に対する&lt;Bさん&gt;の発言が、相手に関心のあるものならば「1」を、無関心ならば「0」を出力してください。1桁の数字で出力してください。</p> <p>1(関心のある発話) -&lt;Aさん&gt;: そのアニメ観てるよ。 &lt;Bさん&gt;: 本当!?僕も観てるよ。 ...</p> <p>0(無関心な発話) - &lt;Aさん&gt;: 次は東京に行ってみたいなあ。 &lt;Bさん&gt;: 別にいいんじゃない? ...</p>
ポジネガ判定	<p>文章から心理分析を行うプロとして、入力された&lt;Aさん&gt;の発言に対する&lt;Bさん&gt;の発言が、喜ぶ表現であれば「p」を、喜んでいない表現であれば「n」を、特に感情が無ければ「e」を出力してください。アルファベット1文字で出力してください。</p> <p>p(相手がポジティブに感じる表現) ... n(相手がネガティブに感じる表現) ... e(どちらでもない表現) ...</p>

まず、ユーザのメッセージが対話相手にとって関心があるものかどうかを判定する。無関心であると判定された場合、ユーザのメッセージは「a (無関心)」に分類される。関心があると判定された場合、対話相手にとって「p (ポジティブ)」な内容なのか、「n (ネガティブ)」な内容なのか、そのどちらでもない「e (ニュートラル)」な内容なのかを判定し、分類する。

#### 4.2.2 特徴の作成

4.2.1項で作成した4つのグループそれぞれについて、グループ内のメッセージに共通する話題や特徴を抽出する。表2に示す特徴の作成に用いるプロンプトにメッセージ集合を挿入し、ChatGPTに入力することで判定を行う。プロンプトでは抽象的な表現にする指示を与えることにより、重要度の高い情報のみの特徴への反映を図っている。これにより、「p (ポジティブ)」、「n (ネガティブ)」、「e (ニュートラル)」、「a (無関心)」の4つの特徴が作成される。

### 4.3 入力メッセージの好感度推定

#### 4.3.1 客観的基準による好感度推定

入力されたメッセージについて、3.1節で定義した5つの客観的基準の項目について、その項目を満たす度合いを1-5の5段階で評価する。表3に示すプロンプト

表 2: 主観的特徴を作成するプロンプト

プロンプト
#指示文
あなたは文章から心理分析を行うプロです。文章を分析し、共通する話題や特徴を特定し、箇条書きで出力してください。分析する文章は、{#参照}に示します。話題や特徴を特定する際、{#ルール}に従ってください。出力形式は、{#形式}に従ってください。
#形式
- 箇条書き - 箇条書き以外の文章は削除する
#ルール
- {#参照}の文章をそのまま出力しない - 固有名詞は使わず、抽象的な表現を使用する
#参照
...

表 3: 客観的基準の「共感」を判定するプロンプト

プロンプト
文章から心理分析を行うプロとして、入力された文章について、相手の発言に共感や理解を示す程度を5点満点で評価してください。1桁の数字で出力してください。
<5点>
-めっちゃ分かる。 -だよね~。 -その通りだと思います。
<4点>
-確かに。すごく良さそう。 ...

トと入力メッセージをChatGPTに送信することで判定を行う。なお、プロンプト内の各項目の説明文については表4に示したものを使用し、各項目に合わせた点数ごとの例文を使用する。

本システムでは、客観的基準の5項目について、点数が「4」以上であればその項目を満たし、「3」以下であれば満たさないものとする。

#### 4.3.2 主観的基準による好感度推定

4.2.2項で得られた特徴をもとに、入力されたメッセージの内容が「p (ポジティブ)」、「n (ネガティブ)」、「e (ニュートラル)」、「a (無関心)」のいずれに当たるかを判定する。表5に示す抽出した特徴によるメッセージ判定のプロンプトに入力メッセージと各項目の特徴を挿入し、ChatGPTに送信することで判定を行う。

### 4.4 メッセージの修正判定

4.3節で得られた入力メッセージに対する好感度の結果をもとに、入力されたユーザのメッセージについて、修正の必要の有無の判定を行う。4.3.1項の結果をもとにした条件1、条件2と、4.3.2項の結果をもとにした条件3について判定を行い、いずれか1つでも満

表 4: プロンプトで使用する客観的基準5項目の説明文

項目	説明
共感	共感や理解を示す程度
関心	関心・興味を持つ程度
好意	好ましく思う程度
ポジティブ	積極的・肯定的な度合い
親しみやすさ	親しみを感じさせる度合い

表 5: 抽出した特徴による入力メッセージの関心・無関心、およびポジネガを判定する判定プロンプト

プロンプト
#指示文 文章から心理分析を行うプロとして、「[入力メッセージ]」という発言が、相手に関心がありポジティブな表現ならば「p」を、相手に関心がありネガティブな表現ならば「n」を、相手に関心があり中立な表現ならば「e」を、無関心ならば「a」を出力してください。{#参照} を例にして出力してください。
#参照 p(ポジティブ) ... n(ネガティブ) ... e(中立) ... a(無関心) ...

たした場合、入力メッセージの修正が必要であると判定する。

**条件 1** 「親しみやすさ」を満たさない

**条件 2** 「共感」を満たさない、かつ「関心」と「ポジティブ」のどちらかを満たさない、かつ「好意」を満たさない

**条件 3** 「n (ネガティブ)」または「a (無関心)」に当てはまる

条件 1 の条件を設けた理由は、SNS 等での運用を想定しており、硬い文章ではなく、軟らかい表現の使用を促すためである。条件 2 については、相手に対して興味を持つような表現を増やすことにより対話相手が感じる好感度を高めるためのものである。

## 4.5 メッセージの修正

4.4 節の修正判定結果をもとに、文章の修正に用いるプロンプトを書き換え、文章の修正を行う。本システムでは、ChatGPT にプロンプトを 3 回入力し、修正案を 1 つずつ、合計 3 つ作成する。表 6 にメッセージ修正に用いるプロンプトを示す。基本文の内容に加え、条件 1 から条件 3 の修正条件のうち満たしたものに対応する文章をプロンプトに追加する形となる。

3 つ修正案を作成する際、条件 2 を満たしている場合には、1 回目は「共感」のみ、2 回目は「関心」と「ポジティブ」のみ、3 回目は「好意」のみを修正させるよう、プロンプトに文章を追加する。

表 6: 入力メッセージを修正するプロンプト

プロンプト
あなたは文章から心理分析を行うプロです。入力された文章について、次の条件に従って修正した文章を 1 つ作成してください。 - 入力された文章と同程度の文字数にしてください。
条件 1 - 相手に対して親しみを感じさせる文章にしてください
条件 2 - 相手の発言に共感や理解を示す文章にしてください - 相手に対して好ましく思う文章にしてください - 相手に対して積極的・肯定的な文章にしてください - 相手の発言に相手の発言に対して関心・興味を持つ文章にしてください
条件 3 - {#参照} を参考にし、ポジティブな表現を増やした文章にしてください - {#参照} を参考にし、喜ばない表現、無関心な表現を避けた文章にしてください

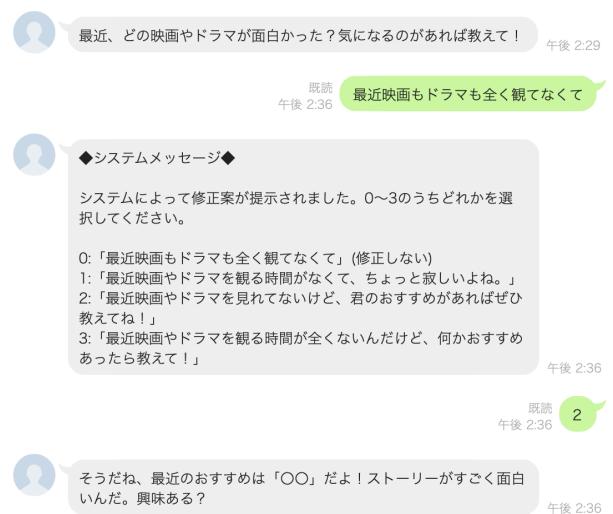


図 2: LINE 上での対話相手に合わせたメッセージ修正案提示システムのデモ

「好意」のみ、3回目は「好意」のみを修正させるよう、プロンプトに文章を追加する。

なお、3つすべての条件に当てはまらなかった場合、そのメッセージの修正は必要ないと判定し、元のメッセージのままユーザに出力する。

## 4.6 ユーザへの出力

4.5 節で、3 つの修正案が作成され、これらがユーザに提示される。ユーザは、提示された修正案から、任意で 1 つの案を選択し、それを対話相手に送信することができる。ユーザは必要に応じて修正後のメッセージを追加修正できる。図 2 に、LINE 上での対話相手に合わせたメッセージ修正案提示システムのデモの様子を示す。修正が必要な場合、システムから修正案が提示され、入力した数字に合わせてメッセージが送信される。

表 7: 実験で使用する対話相手のメッセージ作成用プロンプト

プロンプト	
#指示文	あなたは 20 代くらいの人間です。20 代の友人を相手にした対話をします。対話文を 1 つ生成してください。{#制約条件} に従ってチャットボットとして振る舞ってください。
#制約条件	<ul style="list-style-type: none"> <li>- 自分が AI であることは隠してください。</li> <li>- 会話を終わらせず、続けてください。</li> <li>- 堅い文章にしてください。</li> <li>- 文章を短く、形式的にしてください。</li> <li>- 中立的やネガティブ、否定的な表現を入れてください。</li> <li>- 50 文字以下にしてください。</li> <li>- カウントした結果、50 文字以下であるときに限ってタスクを終了してください。</li> <li>- カウントした結果、50 文字以下でない場合は、50 文字以下になるまで文字を追加したり削除して処理を繰り返してください。</li> </ul>

表 8: 比較システムで使用するメッセージ修正用プロンプト

プロンプト	
#指示文	「[入力メッセージ]」というメッセージを、条件に従って丁寧で好感度が高くなるように修正してください。出力形式は {#形式} を、条件は {#ルール} を参考にしてください。
#形式	<ul style="list-style-type: none"> <li>- テキスト形式</li> <li>- 50 文字以下</li> <li>- 修正後のメッセージのみを出力</li> <li>- 「、や」は削除する</li> </ul>
#ルール	<ul style="list-style-type: none"> <li>- 共感：相手の立場や気持ちに寄り添う表現を加えてください。</li> <li>- 関心：相手に关心・興味を持つ表現を加えてください。</li> <li>- 好意：相手に対して親しみや好ましく思う表現を加えてください。</li> <li>- ポジティブ：積極的・肯定的な表現を使ってください。</li> <li>- 親しみやすさ：フレンドリーで親しみやすいトーンにしてください。</li> </ul>

## 5 対話相手に合わせたメッセージ修正案提示システムの評価実験

本章では、提案システムの評価実験について述べる。メッセージアプリ「LINE」上の被験者(人)と対話相手(ChatGPT)の一対一のテキスト形式の対話において、対話相手(ChatGPT)が提案システムを使った場合と、比較システムを使った場合の、被験者の対話相手に対する好感度の違いに着目し、好感度の違いについて、対話や結果のログ、及び被験者に対するアンケートの結果から検証した。

今回の実験で被験者と対話する対話相手は ChatGPT を利用した bot(被験者には bot ではなく対話相手と伝える)であり、モデルは gpt-4o-2024-05-13 を用いた。通常 ChatGPT はポジティブな表現を生成する傾向にあるため、提案システムや比較システムを利用した際の好感度改善の効果が分かりにくくなる可能性がある。そのため、表 7 に示すように、対話文を生成するプロ

ンプトを、ネガティブな表現や無関心な表現などを出力させるように設定した。

提案システムと比較する比較システムとして、修正の有無の判定や修正案の作成の際、2.2 の好感度や特徴を用いず、プロンプトの指示のみで行うものを用意した。

実験の被験者は、14 名の大学生と大学院生とした。

### 5.1 実験手順

#### 5.1.1 主観的特徴の抽出

被験者(人間)と対話相手(ChatGPT)による対話をを行い、提案システムで使用する被験者のポジネガ特徴の抽出を行う。この際、対話相手に提案システムや比較システムは利用しない。対話は最初にユーザが「こんにちは」というメッセージを送信し、対話相手が「こんにちは。最近大学はどうですか?」と返信するところからスタートし、その後は自由に対話を行ってもらった。なお、「ポジティブ」、「ネガティブ」、「ニュートラル」、「無関心」と判定された対話相手(ChatGPT)のメッセージがそれぞれ 3 文以上かつ対話が 25 往復以上になったとき、あるいは対話が 40 往復に達したときに終了とする。

#### 5.1.2 システム利用時の対話

まず、被験者を、グループ 1 とグループ 2 の 2 つに分ける。以降、提案システムを利用した対話相手を「対話相手 A」、比較システムを利用した対話相手を「対話相手 B」と呼称する。その後、グループ 1 の被験者は対話相手 A、対話相手 B の順、グループ 2 の被験者は対話相手 B、対話相手 A の順番で対話をを行う。5.1.1 項と同様に、対話は最初にユーザが「こんにちは」というメッセージを送信し、対話相手が「こんにちは。最近大学はどうですか?」と返信するところからスタートし、その後は自由に対話を行ってもらった。各対話は 20 往復に到達すると終了する。なお、表 8 に、比較システムで使用するメッセージ修正プロンプトを示す。比較システムでは主観的基準の特徴は使用せず、客観的基準の単純なプロンプトのみで修正を行う。

各対話相手について、対話相手が送信する文章の修正前後での客観的基準および主観的基準の変化を調べる。

また、1 回目および 2 回目の対話終了後、対話した対話相手について以下の質問に 5 段階評価及びその理由を回答してもらうアンケートを実施した。

- 親しみやすい相手だったか
- よく共感や関心、好意を示してくれたか

表 9: 対話相手のメッセージの客観的基準の評価結果

5段階評価		対話相手A(提案システム)					対話相手B(比較システム)				
修正前	修正後	共感	関心	好意	ポジ	親しみ	共感	関心	好意	ポジ	親しみ
3以下	4以上	49.1%	41.3%	77.0%	70.7%	69.9%	41.1%	21.6%	60.5%	70.5%	40.8%
	3以下	50.9%	58.7%	23.0%	29.3%	30.1%	58.9%	78.4%	39.5%	29.5%	59.2%
	4以上	85.0%	80.6%	90.2%	88.3%	80.6%	69.0%	66.7%	92.0%	95.8%	82.7%
4以上	3以下	15.0%	19.4%	9.8%	11.7%	19.4%	31.0%	33.3%	8.0%	4.2%	17.3%
	3以上	0.8%	1.7%	0.0%	0.8%	0.0%	3.5%	10.1%	0.0%	1.0%	2.6%

表 10: 比較システムにおける修正前後のメッセージと客観的基準「関心」の評価

	評価	文章
修正前	3	それは大変ですね。試験勉強と課題のバランスは難しいですね。
修正後	2	それは大変ですね。応援していますので、無理せず頑張ってくださいね。
修正前	3	頻度が少ないと効果も限定的ですね。少しずつでも増やせるといいですね。
修正後	2	頻度が少ないのは仕方ないですし、少しずつ増やせるといいですね。
修正前	3	その通りですね。リラックスも適度に、バランスが大事だと感じます。
修正後	2	その通りですね。お話を共感します。リラックスも適度が大事ですね。

- ポジティブなものだったか
- 引き続き対話をしたいと思ったか

## 5.2 実験結果

### 5.2.1 客観的基準の評価

表 9 は、対話相手 A, B がユーザに送信した文章について、修正前および修正後の客観的基準の 5 段階評価とその割合を示したものである。なお、数値はすべて修正前を前提とした条件付確率であり、対話相手 A (提案システム) については、修正が必要だったものみを母数としている。

修正前が 3 以下だったもののうち、4 以上になった割合がすべての項目において、提案システムが比較システムを上回った。「好意」、「親しみ」については比較システムに比べて 15% 程度向上したが、「共感」、「ポジティブ」については約 6% 程度の向上にとどまった。関心についても 16.6% の向上がみられたが、41.3% と低い結果となった。

また、修正前が 3 以上だったもののうち、2 以下に悪化している文章の割合は、比較システムでは「好意」を除く項目で数パーセント見られ、特に「関心」では 10.1% と高い割合で発生していたが、提案システムではそのような文章がほとんど見られなかった。

表 10 は、比較システムにおいて修正後の評価が修正前より低くなった例である。比較システムでは、すべての項目に対して修正を促すようなプロンプトになっ

ており、他の項目の修正により、関心を示すと判定された表現が他の表現に置き換わってしまった可能性が高い。提案システムでは修正が必要な項目のみに対して修正を促しているので、点数の悪化が少なかったと考えられる。

### 5.2.2 主観的基準の評価

表 11 は、対話相手 A, B がユーザに送信した文章について、修正前および修正後の基準のポジネガ・関心無関心を、対話相手の返答メッセージから判定した結果である。なお、対話相手 A (提案システム) については、修正が必要だったもののみを母数としている。

修正前に「無関心」または「ネガティブ」と判定されたメッセージは、比較システムでは 131 文あり、そのうち改善されたのは 30 文 (22.9%) と低かった。一方で提案システムでは、26 文あり、そのうち改善されたのは 7 文 (26.9%) となり、大きな改善は見られなかった。また、提案システムでは「無関心」または「ネガティブ」と判定された文の数が少なくなっていた。

提案システムで「無関心」、「ネガティブ」と判定された文が少ない原因として、対話相手 B が生成する修正前の文章を生成する際に用いた対話履歴に、提案システムで修正した後の文章が入っており、その修正後の文章の影響を受けて文章を生成していためであると考えられる。

### 5.2.3 アンケート結果

表 11 は、1 回目および 2 回目の対話終了後のアンケートの結果である。引き続き対話をしたいと回答した人は提案システムを利用した対話相手 A の方が多い結果となったが、その他の 3 つの質問に関しては、大きな差が見られず、「親しみやすい相手だったか」という質問に対しては 2 や 1 と回答した人が対話相手 B より多かった。

対話相手 A の「親しみやすい相手だったか」という質問の回答に対する理由では、対話相手の口調に対して親しみを感じた人と、逆に口調に親しみを感じなかつた人が混在していた。親しみやすさを感じる基準が被験者によって異なったためにこのような結果になったと考えられる。

表 11: 対話相手のメッセージの主観的基準の判定結果

		修正後									
		対話相手 A (提案システム)				対話相手 B (比較システム)					
修正前	p (ポジティブ) e (ニュートラル) n (ネガティブ) a (無関心)	p	e	n	a	計	p	e	n	a	計
		62	8	0	0	70	71	7	0	17	95
		6	31	0	2	39	13	29	2	10	54
		0	6	9	0	15	0	4	22	4	30
	a (無関心)	1	0	0	10	11	14	12	7	68	101
	計	69	45	9	12	135	98	52	31	99	280

表 12: 対話相手に対する好感度 (5段階評価)

質問	対話相手	5	4	3	2	1
親しみやすい相手でしたか	A	1	6	2	2	2
	B	0	7	4	2	1
よく共感や関心、好意を示してくれましたか	A	2	9	1	1	0
	B	3	7	2	2	0
ポジティブなものでしたか	A	6	6	1	0	0
	B	7	5	1	1	0
引き続き話をしたいと思いましたか	A	1	7	3	0	2
	B	0	4	5	3	2

## 6 おわりに

ChatGPT を用いて、二者のテキストでの対話における客観的な好感度の基準と、対話履歴から抽出した主観的な特徴をもとに、ユーザが送信しようとしているメッセージに修正が必要かどうかを判定し、修正案を提示するシステムを提案した。

比較システムとの比較実験の結果、提案システムが全体的に上回った結果となり、提案システムの有用性が確認できたが、客観的基準では大きな差が見られなかった項目があり、改善の余地が見られた。

今後は、人間がシステムを利用する評価実験や、アンケートの結果をもとにした口調の好みを判定する仕組みやプロンプトの修正を行うことでコミュニケーション支援をより良くすることを目標にしたい。

## 参考文献

- [1] (URL) <https://www.soumu.go.jp/johotsusintoeki/whitepaper/ja/r06/pdf/n2170000.pdf>, 総務省情報通信白書令和6年版, pp.152–157 (2024)
- [2] 竹内将吾, 酒井あゆみ, 加藤昇平, 伊藤英則: 対話者好感度に基づく感性会話ロボットの感情生成モデル, 日本ロボット学会誌, Vol.25, No.7, pp.1125 – 1133, (2007)
- [3] Zhao, Z., Fan, W., Li, J., Liu, Y., Mei, X., Wang, Y., ... Li, Q. : Recommender Systems in the Era of Large Language Models (LLMs), IEEE Transactions on Knowledge and Data Engineering, Vol.36, pp.6889 – 6907 (2024)

- [4] Liu, J., Shao, Y., Zhang, P., Li, D., Gu, H., Chen, C., ... Gu, N. : Constructing and Masking Preference Profile with LLMs for Filtering Discomforting Recommendation, arXiv preprint, arXiv:2410.05411, (2024)
- [5] 小林峻也, 萩原将文: ユーザの嗜好や人間関係を考慮する非タスク指向型対話システム, 人工知能学会論文誌, Vol.31, No.1, pp.DSF-A.1 (2016)
- [6] 杉原穂, 目良和也, 黒澤義明, 竹澤寿幸: 対話を通じて話者の嗜好情報を学習する音声対話システム, 言語処理学会第23回年次大会発表論文集, Vol.23, pp.226 – 229 (2016)
- [7] Horsham, Z., Haydock-Symonds, A., Imada, H., Tai, H. C., Lam, L. W., Lui, S. T., ... Feldman, G. : Does learning more about others impact liking them?: Replication and extension Registered Report of Norton et al.(2007) 's Lure of Ambiguity, OSF (2024)
- [8] 村越琢磨: 他者に対する理解度の認知が好感度に及ぼす効果, 日本心理学会大会発表論文集, Vol.85, No.7, pp.PI-042-PI-042 (2021)

# 生成 AI を用いた広告クリエイティブにおける ジェンダー関連炎上の可能性検出の検討

## A Study of Detecting the Potential for Gender-Related Flaming in Ad Creative Using Generative AI

平野徳麻<sup>1</sup> \* 笹嶋宗彦<sup>1</sup>  
Tokuma Hirano<sup>1</sup> Munehiko Sasajima<sup>1</sup>

<sup>1</sup> 兵庫県立大学  
<sup>1</sup> University of Hyogo

**Abstract:** 近年、企業や自治体の広告が炎上し、その企業や自治体の評判が低下することが問題となっている。そこで本研究では、ジェンダーの観点に着目し、広告の種類をオンライン広告のバナー広告に限定した広告の炎上防止を目的とする。その目的の達成のためにジェンダー関連の広告炎上原因を構造化したオントロジーを構築し、生成 AI (ChatGPT-4o) を用いてバナー広告やオントロジーの解析を行い、最終的にバナー広告の炎上可能性の検出を行う。

### 1 はじめに

バナー広告とはインターネット上の広告枠に画像や動画などで表示される形式の広告で、私たちがウェブサイトなどを閲覧する際によく見ることができる。バナー広告は企業や自治体にとって重要なマーケティング手段となっており、松田ら（2007）は、バナー広告の単純接触効果が商品評価と購買意図に与える影響を実験的に検討し、バナー広告の反復呈示により好意度および購買意図が向上することを示した[1]。また松田ら（2008）は、バナー広告の反復呈示により商品記憶および好感度や再認の評価が長期的に保持されることを示した[2]。しかし、広告表現の内容が不適切であることが原因で炎上し、メディアに取り上げられることが増加している。広告では、電車内に貼られていた栄養ドリンクの広告で男女それぞれ2種類の広告がある中、女性側の広告にだけ「仕事、育児、家事。3人自分が欲しくないですか？」とあり、その内容が女性への家事や育児の押し付けと見られ、時代錯誤で性差別だとSNSを中心に炎上した例がある[4]。企業や自治体の広告が炎上することにより、その企業や自治体の商品やサービス、または企業や自治体そのものの評判が低下することが問題であると考えた。

そこで本研究では、炎上の原因をジェンダーの観点に着目し、対象をバナー広告に限定した広告の炎上防止を目的とし、生成 AI を用いた炎上可能性の検出を行う。具体的には、オントロジーを用いて体系化した

ジェンダー関連での広告の炎上原因をマイ GPT に知識としてアップロードし、バナー広告の画像の炎上可能性をカスタマイズしたマイ GPT を用いて検出する。その結果、オントロジーで体系化した炎上原因に沿った、広告の炎上可能性の検出をすることができた。

### 2 関連研究

#### 2.1 バナー広告の効果

前節で示した通り、バナー広告は企業や自治体にとって重要なマーケティング手段である。ユーザーが様々なウェブサイトを閲覧する際に、バナー広告を何回も反復して目にして単純接触効果により商品やサービスの好感度や認知度が上がり購買意欲に結びつく。さらに中岡ら（2008）は人がバナー広告にどのようなイメージを抱いているかを分析するために、SD法による感性評価実験を行い、バナー広告を評価する基準として、「目にとまる」、「心をひかれる」の2つが重要視されることを示した。バナー広告の背景色や文字色、背景色の彩度、アニメーションやイラストが感性に影響を与える、そのデザインを工夫することで、ユーザーにクリックされやすくなる[3]。

#### 2.2 炎上のメカニズム

ウェブで発生する炎上について平井（2012）は、「炎上とは、ブログ、ミクシィ（mixi）、ツイッター（Twitter）などに投稿されたメッセージ内容、ならびに投稿者に

\*連絡先：兵庫県立大学社会情報科学部社会情報科学科  
〒 651-2197 兵庫県神戸市西区学園西町 8 丁目 2-1  
E-mail:sasajima@sis.u-hyogo.ac.jp

対して批判や非難が巻き起こる現象」と定義し、炎上が起こる背景として電子掲示板の2ちゃんねるを中心に形成された文化と、若年層による携帯電話のやりとりを通じて形成された文化の2つの文化がウェブサービスの発展に伴い、双方が拡張し、接続してきたことであると考察している[5]。また、山口(2016)は、炎上を「ある人や企業の行為・発言・書き込みに対して、インターネット上で多数の批判や誹謗中傷が行われること」と定義した。さらに、炎上加担行動の動機について定量的な分析を行い、「頻発している炎上について、ごく一部の人が、正義感から加担しており、かつ、そのような人が複数の炎上事例に加担し、何度も書き込んでいる」ことを示した[6]。ソーシャルメディア上の個人攻撃を類型化して背後にある構造を理解することを目的とした瀬川ら(2021)の研究では、Sentence-BERTを利用して日本語大規模ツイートデータから網羅的に攻撃事例を抽出し、攻撃パターンが趣味趣向が近いユーザー間で起こる場合と、そうでないユーザー間で起こる場合の2つがあることを明らかにした。趣味趣向が近いユーザー間の攻撃は、一部のコミュニティで起こっている場合が多く、そうでないユーザー間の攻撃は、政治的な思想が強いユーザーや、普段の投稿内容がネガティブなユーザーによって起こされやすく、また、集団攻撃となっている傾向が強く、普段の会話においてインタラクションを行っていることが多いことを示した。[7]。

### 2.3 広告の炎上

広告の炎上について水村(2023)は、人種・性差別な表現で炎上した事例を基に、広告を見てSNS上で批判・攻撃モードを発動させる人の心の中はどうなっているのかについて様々な角度から検討した。また、不適切な広告表現については、「広告主・広告代理店・媒体社各社が予防的な措置を講じるべきである。」と述べた上で、「実際問題として、広告を見て何を連想するのかは三人三様で、どのような表現が人の心の機微に触れるのかを正確に予測することは難しい。」と広告の炎上を防ぐことは困難であると言及している。[8]

### 2.4 ジェンダー関連の炎上

ジェンダー関連で炎上したCMについて瀬地山(2020)は、過去炎上したCMのパターンを分類し、4象限に図式化した。縦軸に「女性」「男性」といった「商品や描写内容の訴求対象」をとり、横軸には「性役割」「容姿・性的メッセージ」といった「炎上ポイント」をとった。以上の4象限の図式を図1に示す。

まず、右上の第1象限は、「女性」を応援する意図の広告であったはずが「性役割」の固定化・強化と受け取られ炎上したパターンである。次に、左上の第2象限は、「女性」を応援したつもりなのに「外見・容姿」の面で性差別と受け取られ炎上したパターンである。左下の第3象限は、一般的に受け取られるように作られたが、「男性」への性的なメッセージが強く炎上したパターンである。最後に右下の第4象限は、「男性」への共感がかえって「性役割」の固定化・強化につながってしまい炎上したパターンである。[9]

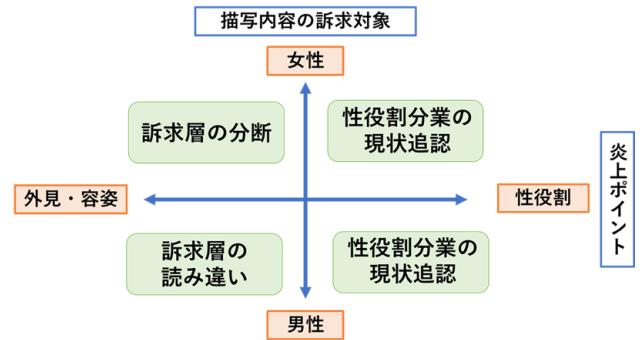


図1: ジェンダー関連で炎上したCMの4象限

これまでに発表された研究では、炎上の背景や原因といったメカニズムに関する研究は行われているが、炎上を予見したり、防いだりするためのシステムの研究はされておらず、炎上を防ぐためには広告主や広告代理店、媒体社各社などの現場で対策しなくてはならない。

### 3 提案手法

本研究では、バナー広告がジェンダーの観点で炎上する可能性の有無を、ジェンダー関連での広告の炎上原因を体系化したオントロジーの知識を持たせた生成AIを用いて検出するシステムを提案する。バナー広告の制作やインターネット広告の運用者が広告を作成して出稿する際に、事前に提案システムを利用することで、炎上の可能性の有無が分かるようになり、予見された炎上に備えることができるようになると考えている。本研究では生成AIとしてChatGPT-4oを用いる。2024年11月時点のモデルである。また、生成AIを用いた一連の処理についてはChatGPTに搭載されているGTPsという機能で、特定の処理に特化したチャットボットを作成することのできるマイGPTを作成した。マイGPT内に構築したオントロジーをXMLファイル形式でアップロードし、プロンプトもプロンプトの原則26ヶ条[10]を参考に事前に記述した。それにより、提案手法は作成したマイGPT上にてバナー広告

の画像を1枚ずつアップロードし入力され、検出の結果が出力される。

本章では、提案手法について、オントロジーの構築、バナー広告の解析、バナー広告の炎上可能性の検出、検出の結果の4節に分けて示す。提案手法で用いたプロンプトは付録に示す。

### 3.1 オントロジーの構築

ジェンダー関連での広告の炎上原因を体系化したものとして、ジェンダー広告炎上原因オントロジーを構築した。オントロジーとは、共通語彙（概念）を提供する体系化された辞書のようなもの、つまり、対象の知識の意味を共通する語彙で表し体系化したものである。オントロジーが知識のバググラウンドにある暗黙的な情報を明示することで、知識が表している内容が明確になり、相互理解の助けになる[11]。ジェンダー広告炎上原因オントロジーは、瀬地山（2020）が図式化した4象限[9]を基に構築した。ジェンダー広告炎上原因オントロジーを構築することで、個人的な感想に偏らず、ジェンダーに着目した広告の炎上原因の知識を明確化できた。以上のジェンダー広告炎上原因オントロジーを図2に示す。

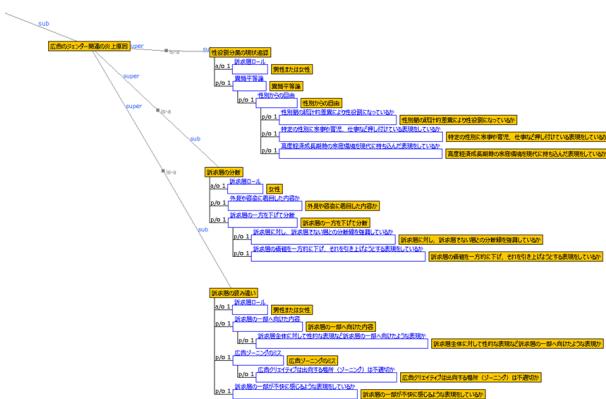


図2: ジェンダー広告炎上原因オントロジー

### 3.2 バナー広告の解析

バナー広告の炎上可能性を検出するためには、バナー広告の内容を適切に解析し、テキスト情報や視覚的要素を抽出する必要がある。バナー広告の画像データを対象に解析を行い、炎上可能性を評価するための特徴量を抽出した。本節では、バナー広告解析の具体的な手法と手順について示す。

バナー広告の画像は主に、商品説明やコピーなどのテキストと商品自体や背景のイラストで構成されてい

る。そのテキストやイラストの情報を解析するために、付録??で示したプロンプトでは、まず、バナー広告内のテキストを抽出させ、文字集合ごとにリスト化させる。次にバナー広告のイラストについてレビューさせ、その詳細を形状、色、構図などに着目して説明させる。また、炎上可能性の検出には、それが誰に向けた広告であるかを表す、訴求対象の情報が必要なので、得られたテキストとイラストの情報を基にバナー広告の訴求対象を判断させた。

### 3.3 バナー広告の炎上可能性の検出

バナー広告に含まれる要素をもとに、炎上可能性の有無を判断する。本節では、炎上可能性の検出の手順とプロンプト設計について示す。

前節で示した、バナー広告のテキストとイラストの情報。そして、その2つの情報を基に取得した訴求対象の情報の、合わせて3つの情報と3.1で構築したジェンダー広告炎上原因オントロジーを用いてバナー広告のジェンダーの観点での炎上可能性の検出を行う。プロンプトでは、オントロジーの解析の際に構築したオントロジーの「広告のジェンダー関連の炎上原因」の概念の下位概念を用いることを指示した。また、プロンプトの原則26ヶ条[10]に従い、指示と情報の分割やFew-shotプロンプトの使用、「ステップバイステップで考えてください」プロンプトの使用といった工夫をプロンプトに反映させた。

### 3.4 検出の結果

提案手法を用いて、バナー広告の炎上可能性を検出した。本節では、検出の詳細と結果について述べ、手法の有効性および課題を考察する。

検出に用いたバナー広告の画像は

1. 炎上原因に該当する自作バナー広告（図3）
2. 炎上原因には該当しない自作バナー広告（図4）
3. クローリングして収集した炎上原因に該当する既存のバナー広告（図5）
4. クローリングして収集した炎上原因には該当しない既存のバナー広告（図6）

の4種類を用いた。炎上原因に該当する、または該当しないの判断は筆者の予測である。

検出の結果をバナー広告と共に図3から順に示す。図3、図4、図5の炎上可能性の検出に関しては、筆者の予測通りの結果になった。しかし、図6の炎上可能性の検出に関しては、筆者の推測とは違う結果になった。

図 6 の化粧品の広告は図 1 では「訴求層の分断」という炎上原因に該当することが多いが、図 6 のバナー広告は訴求層の一方を落とすような表現は見られず、炎上原因に該当しないと考えた。しかし、容姿に着目した内容かつ女性しか広告内に登場していないため、女性が美しくあるべきという性役割の固定・強化をしてると ChatGPT が判断したと考えられる。



図 3: 炎上原因に該当する自作バナー広告

- 炎上可能性: あり
- 炎上可能性の原因:
  - 「25歳すぎると顔にできる」という表現が「若さ」を過剰に強調し、年齢による美の価値観を押し付けていると捉えられる可能性があります。
  - 年齢を基準に美しさを評価することが、訴求層の一部に不快感を与える可能性があります。また、外見や容姿への過剰なフォーカスは、ジェンダー関連の炎上要因である「外見や容姿に着目した内容」に該当します。



図 4: 炎上原因には該当しない自作バナー広告

- 炎上可能性: なし、または著しく低い
- 炎上可能性の原因: このバナー広告にはジェンダー関連で炎上する原因となる要素は含まれていません。しかし、困惑した表情がターゲット層にとって誤解を招く可能性はあるため、文脈を十分に伝える補足テキストを加えることを検討すると良いでしょう。



図 5: クローリングして収集した炎上原因に該当する既存のバナー広告

- 炎上可能性: あり
- 炎上可能性の原因:
  - 「女性が育児・教育を行う」という構図が性別役割分業を暗に支持していると受け取られる可能性。
  - 男性やその他のジェンダーが全く描かれていない点で、ジェンダー的な偏りを感じる視聴者がいるかもしれない。



図 6: クローリングして収集した炎上原因には該当しない既存のバナー広告

- 炎上可能性: あり
- 炎上可能性の原因:
  - バナー広告は「外見や容姿に着目した内容」となっており、ジェンダー広告炎上オントロジーの「性役割分業の現状追認」や「外見に依存した表現」の項目に該当するリスクがあります。
  - 女性のみをターゲットとし、美しさや肌の透明感を前提とするメッセージが一部の層にステレオタイプ的であると認識される可能性があります。
  - さらに、スキンケアが女性専用であるかのような暗黙のメッセージが含まれており、「性別による役割の固定化」と受け取られるリスクが存在します。

## 4 おわりに

本研究では、ジェンダー関連の広告の炎上原因をオントロジーを用いて体系化し、生成AIによるバナー広告の炎上可能性の検出することを提案した。本研究は予測が困難と言われる炎上がある程度検出可能であることを示すことができた点で、広告の炎上防止に貢献していると言える。

しかし、検出の結果が筆者の予測と異なる場合もあるため、不適切な広告表現の検出は、確実であるとは言えない。また、炎上の原因をジェンダーの観点に、広告の種類をバナー広告に限定して考えたため、それ以外の原因で起きる炎上には対応出来ない。今後は、オントロジーやプロンプトを改善することで、人種差別など他の不適切な表現、映像広告といった広告形態など、幅広い範囲で広告の炎上を防げるようなシステム構築を目指す。

## 参考文献

- [1] 松田憲, 平岡齊士, 杉森絵里子, 楠見孝 (2007). バナー広告への単純接触が商品評価と購買意図に及ぼす効果. *Cognitive Studies*, 14(1), 133-154.
- [2] 松田憲, 小林剛史 (2008). バナー広告の反復呈示が短期・長期記憶と商品評価におよぼす効果. 日本心理学会第 72 回大会発表論文集. 2AM097.
- [3] 中岡伊織, 氏平淳子, 串田淳一, 亀井且有 (2008). 「ついクリックしたくなる」バナー広告の感性評価およびそのモデル構築. 第 24 回ファジィシンポジウム. TD2-4.
- [4] やまもとゆか. 「鬼すぎない?」 大正製薬の広告が“性差別”と物議…… 男女の“非対称性”に「昭和かな?」「時代にあってない」. ねとらぼ. 2024-07-04. <https://nlab.itmedia.co.jp/n1/articles/2407/03/news185.html>. (参照 2024-11-24).
- [5] 平井智尚 (2012). なぜウェブで炎上が発生するのか—日本のウェブ文化を手がかりとして. 情報通信学会誌, 29 卷, 4 号, 61-71.
- [6] 山口真一 (2016). 炎上加担動機の実証分析. 2016 年社会情報学会 (SSI) 学会大会. [https://www.sgu.ac.jp/soc/ssi/papers/32.pdf?utm\\_source=chatgpt.com](https://www.sgu.ac.jp/soc/ssi/papers/32.pdf?utm_source=chatgpt.com)
- [7] 瀬川友香, 浅谷公威, 坂田一郎 (2021). ユーザーに着目した SNS 上の攻撃とそのメカニズムに関する分析. 2021 年度人工知能学会全国大会 (第 35 回). 1I4-GS-4c-04.
- [8] 水村典弘 (2023). 不適切な広告表現の研究-人種・性差別的な表現で炎上した事例の分析-. 日本経営倫理学会誌, 第 30 号, 61-74.
- [9] 瀬地山角 (2020). 炎上 CM でよみとくジェンダー論, 光文社新書.
- [10] Sondos Mahmoud Bsharat, Aidar Myrzakhan, Zhiqiang Shen (2023). "Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4". <https://doi.org/10.48550/arXiv.2312.16171>
- [11] 溝口理一郎, 古崎晃司, 來村徳信, 笹島宗彦 (2006). オントロジー構築入門, オーム社.

## 付録

### 【提案手法で用いたプロンプト】

バナー広告画像から指定された情報を抽出し、その情報をもとにバナー広告のジェンダー関連での炎上可能性を検出し、指摘してください。

- バナー広告内の文字を抽出し、文字集合ごとにリスト化してください。
- バナー広告内のイラストや描写の詳細な説明を提供してください。
- 抽出した文字とイラスト描写を基にバナー広告の訴求対象を考えてください。
- 入力されている『ジェンダー広告炎上オントロジー v1.xml』という XML ファイルからオントロジーの構造を解析してください。
- バナー広告の文字集合、イラスト説明、訴求対象と XML ファイルのオントロジーからバナー広告の炎上可能性を検出し、指摘してください。

### # Steps

1. バナー広告画像を解析し、文字を抽出します。
2. 検出された文字を一つの文字集合として整理し、一つのリストとして記載します。
3. バナー広告のイラストをレビューし、その詳細を説明します。形状、色、構図などに着目してください。
4. step 2 と step 3 で得られた情報を基にバナー広告の訴求対象を考えてください。
5. step 2, step 3, step 4 で取得した情報をまとめ、バナー広告の情報を取得してください。
6. 入力されている『ジェンダー広告炎上オントロジー v1.xml』という XML ファイルを解析し、オントロジーの構造を取得してください。
7. step 5 で取得したオントロジーの構造のうち、「広告のジェンダー関連の炎上原因」の概念の下位概念のオントロジーを用いて、step 5 でまとめたバナー広告がジェンダー関連で炎上する可能性があるか考えてください。
8. 炎上の可能性がある場合は、その炎上原因についてオントロジーを基に考えてください。炎上の可能性がない場合はその旨を記述してください。

### # Output Format

- 炎上可能性: テキスト形式
- 炎上可能性の原因: 詳細なテキスト記述形式

### # Examples 1

#### \*\*入力:\*\*

- バナー広告画像

#### \*\*出力:\*\*

- 炎上可能性: あり
- 炎上可能性の原因: 「このバナー広告では訴求対象が女性で、イラストに女性だけが家事を行っている描写があり、昭和のステレオタイプである『家事は女性がするもの』ということを追認しています。すなわち女性に対して家事をすることを押し付けることを肯定してしまっており、『性役割の現状追認』の点で炎上する可能性があります。」

### # Examples 2

#### \*\*入力:\*\*

- バナー広告画像

#### \*\*出力:\*\*

- 炎上可能性: なし、または著しく低い
- 炎上可能性の原因: このバナー広告にはジェンダーに関して炎上する原因となる描写は見つかりませんでした。しかし、ジェンダー以外の観点で炎上してしまう可能背はゼロではないので、文字・文章表現やイラストの描写などの確認をしましょう。

### # Notes

- テキスト解析の精度は画像の質に依存することがあります。
- イラスト説明には、主観が入らないように具体的な要素にフォーカスすることが重要です。
- 両方の情報がバランスよく提供されるように努めてください。
- バナー広告の広告ゾーニングは「不特定多数の人の閲覧の可能性がある WEB」です。
- オントロジーは「広告のジェンダー関連の炎上原因」の概念の下位概念を用いて、炎上の検出はそれぞれの下位概念の最下層の葉の概念から準備検討してください。
- 入力されたオントロジー知識の情報で炎上の可能性を判断してください。
- 出力は日本語で出力してください。
- ステップバイステップで考えてください。
- 出力は Examples の出力例に沿った出力をしてください。

# 瞳検出技術を用いた動画学習支援システム

## An Eye Pupil Detection-Based Review Support System for Video-based Learning

DUAN ZHENZHONG<sup>1\*</sup> SHAN JUNJIE<sup>2</sup> 西原陽子<sup>2</sup>

<sup>1</sup> 立命館大学大学院 情報理工学研究科

<sup>1</sup> Graduate School of Information Science and Engineering, Ritsumeikan University

<sup>2</sup> 立命館大学 情報理工学部

<sup>2</sup> College of Information Science and Engineering, Ritsumeikan University

**Abstract:** 近年、動画学習を利用する機会が増えている一方、対面授業で勉強するより学習者は集中力を維持しにくくなる傾向がある。本研究では、デバイスの正面カメラでユーザの視聴状況を計測し、学習動画への視聴状況を可視化するグラフを提供する動画学習の支援システムを提案する。20名の被験者に動画学習を実施させ、事後テストの正解率により提案システムを評価した結果、実験群の平均正解率は93%、対照群の平均正解率は82%となった。

### 1 はじめに

近年、コロナウイルスの流行により、世界各地ではオンライン学習が実施されている[1][2][3]。オンライン学習の一環として、「動画学習」が時間や場所などの制限がないため、学習者に人気があって幅広く利用されている[4][5]。

しかし、動画学習にはデメリットがある。伝統的な対面授業と違い、教師やクラスメイトが近くにいないため、学習者は外部からの妨害で集中が途切れやすくなる。さらに、長時間の学習動画の場合、学習内容は多岐にわたるため、学習者個人の記憶のみで各自の復習すべき時間帯を把握するのも容易ではないと考えられる。結果として、復習時間を増やしても、期待される復習効果が得られない場合もたくさん発生する[6]。

そこで、本研究は瞳検出技術を用いて、動画学習の利用者の復習を支援するシステムを提案する。提案システムは、学習者が学習動画を視聴すると同時に瞳を検出し、その学習者の視聴状況<sup>1</sup>を計測する。計測した視聴状況を学習動画の時間軸と整合し、ユーザの視聴状況を反映する視聴状況図を描画する。ユーザは視聴状況図から学習動画の復習すべき具体的な時間帯の情報を簡単に把握できる。これにより、提案システムを通して、ユーザは動画学習への復習効率を向上させる効果が期待される。

本研究では、以下3つの内容を行った：

(1) 人間の瞳を検出するためのYOLOv5 モデルを作成した。

(2) 瞳の検出結果により、ユーザの視聴状況を計測及び学習動画の時間軸に合わせて視聴状況図を描画する復習支援システムを実装した。

(3) 被験者実験を実施して、提案システムの動画学習への復習支援効果を確認した。

### 2 関連研究

本章では、学習支援に関する既存研究を述べ、本研究の位置付けを行う。

オンライン学習の利用者の学習効果と満足度について、Ikhsan らの研究がある[7]。この研究では、Structural Equation Modeling を用いて、オンライン授業を受ける大学生の学習の効果と満足度の決定要因を探求した。潜在的な決定要因として、教師からの促進、技術的な支援、カリキュラムの構造、教師からのフィードバック、自己啓発、仲間からの協力があり、6つの側面から調査を行った。結果として、教師からの促進、技術的な支援、教師からのフィードバック、自己啓発は学生の学習効果に影響を与えることが示された。この研究から、オンライン学習を受ける学習者にとって、技術的な支援は重要で有益であることが示された。本研究では技術的な支援として、瞳検出技術を用いた復習すべき時間帯を学習者に知らせ、その効果を検証する。

Yeh らは、学習者を監視する研究を行った[8]。この研究では、Dlib1 を用いて瞳検出を行い、まばたきを指標にすることで学習者の不正行為を検出し、瞳の角度

\*連絡先：立命館大学情報理工学研究科

〒 567-8570 大阪府茨木市岩倉町 2-150

<sup>1</sup>本文では視聴状況を「ユーザは学習動画を視聴しているかどうか」とする。

を用いて学習者が適切な学習態度や行動を示しているかどうかを判断する。Yeh らの研究では、デバイスのカメラを利用して瞳検出を行い、学習者のオンライン学習を監視した。瞳検出の結果からユーザの視聴状況を分析できることを示した。他に、Abdulkader らは、オンライン学習を利用する学生の授業に対する集中力を分析する研究を行った [9]。このシステムは顔認識アルゴリズムとディープラーニングモデルである「DLIP」を用いて、対象の目と口の特徴を抽出する。しかし、この 2 つの研究の目的は学習者の学習状態を教師にフィードバックすることであるため、教師が存在するライブ配信授業を前提としていた。また、学習者を監視することで教師の授業を支援するため、学習者に直接的な支援を提供しなかった。この 2 つの研究と異なり、本研究は教師が存在しない動画学習を対象とし、学習者に視聴状況図を提供することで、学習者に直接的な支援を提供する。

Raca らは、対面授業中にクラス全員の集中状況を監視し、教師にフィードバックできるシステムを提案した [10]。このシステムを利用することで、教師はリアルタイムで学生がどの程度授業を受けているかを把握し、授業のスケジュールを調整することができる。Abdulrahman らは、コンピュータビジョン技術を用いて、学生の顔を特定し、対面授業に集中しているかどうかを判断して、教師に示せるシステムを提案した [11]。この 2 つのシステムと異なり、本研究の提案システムは対面授業ではなく、オンライン学習の動画学習に注目し、直接学習者に復習支援の視聴状況図を提供する。

### 3 提案システム

本章では提案システムについて説明する。

#### 3.1 提案システムの流れ

図 1 に提案システムの流れを示す。はじめに、提案システムを起動し、デバイスの正面カメラでユーザの学習動画を視聴する時の顔画像を撮る。次に、撮った顔画像に対して、再訓練された YOLOv5 モデルで瞳検出を行い、検出結果を記録する。システムは検出結果からユーザの視聴状況を判断する。学習動画の視聴が終了した後、ユーザの学習動画を視聴している時間帯と視聴していない時間帯を特定し、ユーザが理解しやすいように、視聴状況図を可視化する。

#### 3.2 顔画像からの瞳検出

顔画像から瞳検出を行う方法を説明する。提案システムでは YOLOv5 を用いて瞳検出を行った。

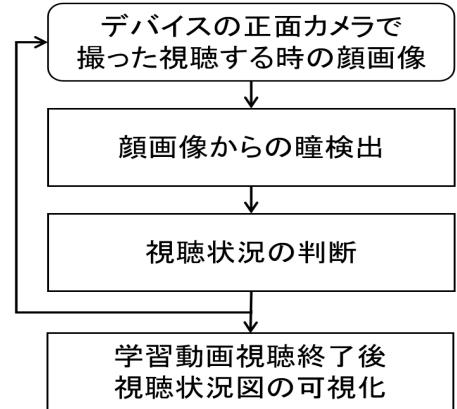


図 1: 提案システムの流れ

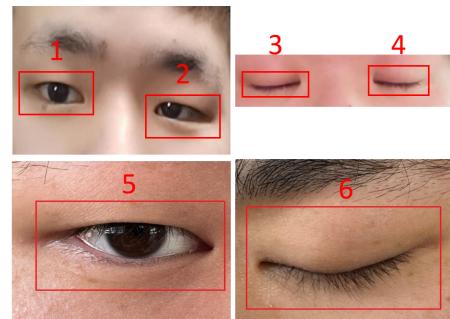


図 2: 教師データの例

YOLOv5 公式サイトで公開されたモデルは人間の瞳を検出できないので、再訓練する必要がある。教師データとして、「開いている両目 (1, 2)」、「閉じている両目 (3, 4)」、「開いている片目 (5)」、「閉じている片目 (6)」(図 2)、4 種類の目の画像各 248 枚を用意し、手作業でラベリングした。各画像のラベルは目が開いているか閉じているかを示す 0 または 1 の数値である。目が開いていると 1、目が閉じていると 0 とする。加えて、瞳の位置情報を画像にアノテーションとして付与した。位置情報として 4 つの数値があり、それぞれは画像の幅と高さを基に標準化された瞳の中心点の x 座標、中心点の y 座標、瞳の範囲を示す枠の幅と枠の高さである。画像に含まれる目の数に応じて、記録された目のラベルと位置情報の数も変化する (表 1)。

YOLOv5 は事前訓練モデルを 5 つ提供している。それぞれは「YOLOv5n」、「YOLOv5s」、「YOLOv5m」、「YOLOv5l」、「YOLOv5x」といい、検出精度が順に高くなるが、処理速度が順に遅くなる。本研究では、処理速度と検出精度を両立させる「YOLOv5m」モデルを用いて、教師データで再訓練を行った。その際に設定したパラメータを表 2 に示し、再訓練の結果を表 3 に示す。再訓練されたモデルを用いることにより、精度の高い瞳検出が可能になる。

表 1: 図 2 の画像に対する目のラベルと位置情報

目の ID	ラベル	x 座標	y 座標	幅	高さ
1	1	0.20	0.42	0.34	0.29
2	1	0.78	0.26	0.33	0.30
3	0	0.19	0.57	0.25	0.21
4	0	0.68	0.49	0.32	0.21
5	1	0.53	0.48	0.85	0.55
6	0	0.49	0.36	0.88	0.64

表 2: YOLOv5m の再訓練のパラメータ

パラメータ	数値
GPU Memory	7.2 GB
Batch Size	16
Epochs	60
Learning Rate	0.01

### 3.3 視聴状況の判断

瞳検出の結果から、視聴状況を判断する。瞳検出は 1/30 秒ごとに行われる。はじめに瞳検出の結果を 4 種類に分ける。

- (1) 瞳が検出されなかった
- (2) 閉じている両目が検出された
- (3) 開いている両目が検出された
- (4) 開いている片目が検出された

(1) 瞳が検出されなかった、(2) 閉じている両目が検出された場合は、ユーザは学習動画を視聴していないとみなす。反対に、(3) 開いている両目が検出された場合は、ユーザは学習動画を視聴しているとみなす。(4) 開いている片目が検出された場合は、片目をつぶって学習動画を視聴する学習者が現実には存在しにくいと考えられる。瞳検出モデルの動作を確認したところ、ユーザの目が髪の毛やメガネなどに遮られるなどして、瞳検出に失敗することがあった。そこで本研究では(4) の場合は、ユーザは学習動画を視聴しているとみなす。

### 3.4 視聴状況の可視化

ユーザの復習すべき学習動画の時間帯を直観的に示すため、本研究では、学習動画の時間軸に合わせてユーザの視聴状況を可視化する。

時刻  $t=0$  を学習動画視聴開始時刻とする。縦軸に 3 つの値を設け、一番下を(1) 瞳が検出されなかった、または(2) 閉じている両目が検出された、真ん中を(4) 開いている片目が検出された、一番上を(3) 開いている両目が検出されたとする。1/30 秒ごとに動画の終了まで検出結果をプロットする。

表 3: YOLOv5m の再訓練の結果

指標	数値
Bounding Box Regression Loss	0.01533
Objectness Loss	0.004425
Classification Loss	0.02435
Mean Average Precision 50	0.952

続いて、時間軸を幅  $T$  ごとに区切り、 $T$  の範囲内で閉じている両目が検出されたあるいは瞳が検出されなかった割合を算出する。割合が事前に設定した閾値  $V$  以上であれば、該当の  $T$  の範囲を復習すべき時間帯と判断し、点線で囲う。本研究では、 $T$  を 5 秒に、 $V$  を 80% に設定している。

図 3 に視聴状況図の例を示す。青い点が閉じている両目が検出された、あるいは瞳が検出されなかったところで、赤い点が開いている片目が検出されたところ、緑の点が開いている両目が検出されたところになる。灰色の点線で囲われた部分が、ユーザが復習すべき時間帯を示す。

## 4 提案システムの評価実験

20 名の大学生に被験者の協力を依頼し、提案システムの評価実験を行った。実験目的は、提案システムの動画学習における復習支援の効果を検証することである。

### 4.1 評価実験の手順

- P1) 被験者を実験群と対照群の両群に分ける。
- P2) 被験者に学習動画を視聴させる。学習動画を視聴していない状況を再現するため、動画の再生の間に、被験者のスマートフォンに妨害メッセージを送って、返信を要求する。
- P3) 動画学習が終わったら、被験者は 3 分間の復習をする。実験群は提案システムを利用し、提供された視聴状況図を参考にしながら復習する。対照群は提案システムを利用せず、個人の記憶に基づいて復習する。
- P4) 学習動画に関する事後テストを行う。テストの正解率を用いて、提案システムの評価を行う。

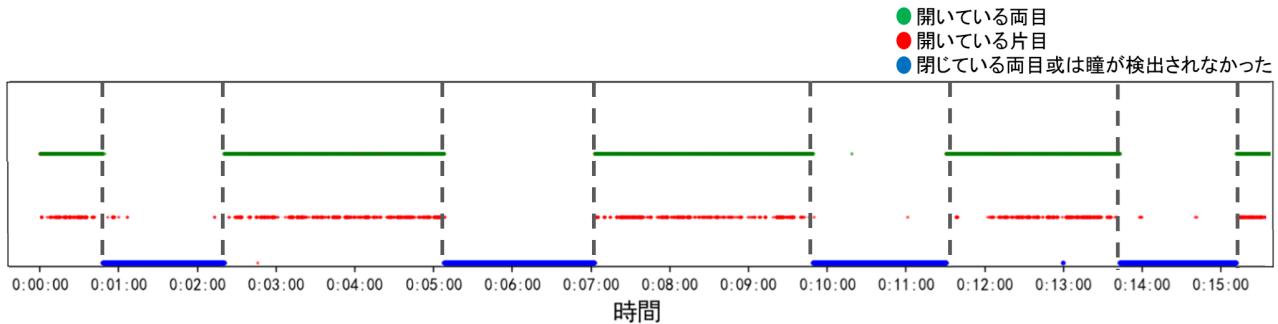


図 3: 視聴状況図の例

表 4: 学習動画のデータ

平均文字数	平均時間	合計文字数	合計時間
128	26 秒	5122	17 分 51 秒

表 5: 妨害メッセージの種類と説明

妨害メッセージ	説明
自己紹介	自己紹介文をスマホで返信させる
動画視聴	送った動画をスマホで視聴させる

## 4.2 実験手順の詳細

評価実験に用いた学習動画は第一著者が作成したもので、内容は「動物おもしろ雑学集」<sup>2</sup>と「ほんよま」<sup>3</sup>2つのウェブサイトを参考にして作成した「動物の雑学40選」であった。それぞれの雑学を説明する際の平均文字数と平均時間、学習動画の合計文字数と合計時間は表4に示す。

評価実験に用いた学習動画は18分程度であったため、最後まで集中力を保てる被験者が居た可能性が高い。そこで、意図的に被験者の学習動画を視聴しない状況を作る必要がある。本実験で採用した方法は、「学習動画の再生の間に、被験者のスマホに妨害メッセージを送って、返信を要求する」であった。妨害メッセージの種類と説明は表5に示す。妨害メッセージは自己紹介文を要求するものと、学習に関係ない動画視聴を要求するものの2つであった。

評価実験の最後に、被験者に学習動画の内容について、選択肢問題40個を出題した。各被験者の事後テストの正解率を以下の3種類に分類した。

(1) FCR: 学習動画を視聴している時間帯の問題の

表 6: 両群の各正解率 (FCR: 学習動画を視聴している時間帯の問題の正解率, UCR: 学習動画を視聴していない時間帯の問題の正解率, TCR: 全体正解率, \*p<0.05, \*\*p<0.01)

	FCR	UCR	TCR
実験群	90.2%	93.7%**	92.8%*
対照群	92.3%	67.0%	81.9%

### 正解率

(2) UCR: 学習動画を視聴していない時間帯の問題の正解率

(3) TCR: 全体正解率

表6に両群の各正解率の数値を示す。

## 5 考察

両群のFCR(学習動画を視聴している時間帯の正解率)について、実験群は90.2%、対照群は92.3%、両群の平均値の差は2.1%であった( $t = 0.476, p = 0.643 > 0.05$ )。これは、学習動画を視聴している時、両群の被験者は同等の学習を行なっていることを示している。本研究では、両群の被験者の学習能力は同等であると考えられる。

しかし、学習動画の視聴が妨害されると、両群の被験者の学習成果に差が現れた。両群のUCR(学習動画を視聴していない時間帯の正解率)について、実験群は93.7%、対照群は67.0%、両群の平均値の差は26.7%であった( $t = -5.032, p = 0.002 < 0.05$ )。この結果から、提案システムは被験者の学習動画を視聴していない時間帯の復習を支援できたことが分かった。

両群のTCR(全体正解率)について、実験群は92.8%、対照群は81.9%、両群の平均値の差は10.9%であった( $t = -2.811, p = 0.01 < 0.05$ )。この結果から、提案システムを利用した実験群の被験者は動画学習が効果的に行えることが確認された。

<sup>2</sup><https://zooing.honpo21.net/> 2024年11月29日アクセス確認。

<sup>3</sup><https://sanctuarybooks.jp/webmag/> 2024年11月29日アクセス確認。

この3つの結果から、提案システムはユーザの動画学習の復習に支援を提供することで、学習動画を視聴していない時間の復習が効果的に行え、学習成果を向上させることができることが確認された。

## 6 おわりに

本研究では、オンライン学習の一種である動画学習の復習効率問題を注目し、瞳検出技術を用いた動画学習中の視聴状況を表示する学習支援システムを提案した。提案システムは瞳検出を実装することで、ユーザの学習動画への視聴状況を計測し、学習動画の再生時間に合わせる可視化図をユーザに提供する。視聴状況の可視化図を参照することにより、ユーザが効率的に復習でき、動画学習の効果を向上させると期待される。

提案システムの有用性を確認するため、20名の被験者を募集し、実際の動画学習を模倣した評価実験を行った。実験の結果として、実験群の学習動画を視聴していない時間帯の問題の正解率は対照群より26.7%高くなった。この結果から、提案システムの利用により、ユーザは限られた時間に学習動画の視聴していない部分をより効率的に復習できることが確認された。

今後の課題として、大学の講義のようにより長い時間での動画学習における提案システムの効果を検証することが考えられる。

## 参考文献

- [1] Cathy Mae Toquero: Challenges and opportunities for higher education amid the COVID-19 pandemic: The Philippine context, *Pedagogical Research*, Vol. 5, No. 4, pp. 1–5 (2020)
- [2] Razzaqul Ahshan: A Framework of Implementing Strategies for Active Student Engagement in Remote/Online Teaching and Learning during the COVID-19 Pandemic, *Education Sciences*, Vol. 11, No. 9, pp. 1–24 (2021)
- [3] Murat Ertan Dogan, Tulay Goru Dogan, Aras Bozkurt: The Use of Artificial Intelligence (AI) in Online Learning and Distance Education Processes: A Systematic Review of Empirical Studies, *Applied Sciences*, Vol. 13, No. 5, pp. 3056 (2023)
- [4] Samad, Albertus Laurensius Setyabudhi: Application of Video-Based Learning Media with Applications to Improve Learning Activities of Vocational High School Students, *Technical and Vocational Education International Journal (TAVEIJ)*, Vol. 3, No. 1, pp. 22–27 (2023)
- [5] Samy Cheikh Youssef, Abdullatif Aydin, Alexander Canning, Nawal Khan, Kamran Ahmed, Prokar Dasgupta: Learning Surgical Skills Through Video-Based Education: A Systematic Review, *Surgical Innovation*, Vol. 30, No. 2, pp. 220–238 (2023)
- [6] Jon-Chao Hong, Yue Liu, Yinsheng Liu, Li Zhao: High School Students' Online Learning Ineffectiveness in Experimental Courses During the COVID-19 Pandemic, *Frontiers in Psychology*, Vol. 12, pp. 738695 (2021)
- [7] Ridho Bramulya Ikhsan, Listya Ayu Saraswati, Brian Garda Muchardie, Vional, Andrianto Susilo: The Determinants of Students' Perceived Learning Outcomes and Satisfaction in BINUS Online Learning, *2019 5th International Conference on New Media Studies*, pp. 68–73 (2019)
- [8] ChunHsiu Yeh, Wei-Cheng Shen, Hung-Yu Chi, Chin-En Lin, Jong-Shin Chen: Enhancing Online Learning Monitoring with Novel Image Recognition Method Using Dlib for Eye Feature Detection, *2023 12th International Conference on Awareness Science and Technology*, pp. 340–345 (2023)
- [9] Rasheed Abdulkader, Firas Tayseer Mohammad Ayasrah, Venkata Ramana Gupta Nallagattla, Kamal Kant Hiran, Pankaj Dadheech, Vivekanandam Balasubramaniam, Sudhakar Sengen: Optimizing student engagement in edge-based online learning with advanced analytics, *Array*, Vol. 19, pp. 100301 (2023)
- [10] Mirko Raca, P.Dillenbourg: System for assessing classroom attention, *Proceedings of the 3rd International Learning Analytics and Knowledge Conference*, pp. 265–269 (2013)
- [11] Abdulrahman Renawi, Fady Alnajjar, Medha Parambil, Zouheir Trabelsi, Munkhjargal Gochoo, Sumaya Khalid, Omar Mubin: A simplified real-time camera-based attention assessment system for classrooms: pilot study, *Education and Information Technologies*, Vol. 27, pp. 4753–4770 (2022)

# 再購入者予測に関するデータ分析と特徴量についての考察

## Data Analysis and Investigation of Features for Re-purchase Prediction

董子華  
Zihua Dong

柴田祐樹  
Hiroki Shibata

高間康史  
Yasufumi Takama

東京都立大学  
Tokyo Metropolitan University

**Abstract:** This paper reports on the results of analyzing a dataset for repurchase prediction competition. This paper sets two research questions: identifying important time periods during which actions taken contribute to the prediction and effective features in terms of acquisition costs and privacy protection. Experimental results show that actions taken within one month from Singles' day is important for the prediction. It is also confirmed that the prediction accuracy do not decrease without product-level features.

## 1 はじめに

本稿では、ユーザが同じ販売者から再度購入するかを予測するタスク（再購入者予測タスク）を対象としたデータセットを分析した結果について報告する。

近年、Amazon<sup>1</sup>や天猫（Tmall）<sup>2</sup>に代表される電子商取引（Eコマース）プラットフォームが発展し、多数の販売者がオンライン・チャネルを通じて商品を販売したり、サービスを提供したりするようになった。販売者にとって、ユーザのロイヤリティを推定することで、ターゲットを絞った販売促進活動を行うことが重要になっており、機械学習の適用が研究されている。関連するタスクの一つである再購入者予測は、ユーザが販売者から商品を購入した後に、同じ商品を再び購入したり、再びこの販売者から購入したりするかどうかを予測するタスクである[1,2]。このタスクに関して、Tmallのデータに基づくコンペティションが2015年に開催されている<sup>3</sup>。このコンペティションでは、匿名ユーザの11月11日（独身の日）に開催される大規模セールでの購入者が、6か月以内に同じ販売者から再度購入するかを予測する。11月11日当日から過去6ヶ月のログデータが公開されており、分類モデルの学習に用いることができる。優勝者を含む数人の研究者による論文が発表されているが、6か月ものデータが予測に必要かについては検討されていない。また、公開されているデータは匿名データではあるものの、各ユーザがアクセスした商品IDまでが記録されおり、かなり詳細なデータといえる。このうち、予測のためにはユーザに

関するデータがどの程度詳細に必要なのかについても議論されていない。

本稿では、再購入者予測タスクに用いる特徴に関する知見を得ることを目的として、Tmallのデータセットを分析する。具体的には、以下の2つのリサーチクエスチョンを設定する。

- **RQ1**：どの時期の行動が、11/11 当日の行動に影響しているか
- **RQ2**：取得コストやプライバシー保護の観点から、再購入予測に有効な特徴は何か

RQ1については、特徴量の計算に用いるデータの収集期間を変更して複数の分類モデルを構築し、その予測精度を比較する。実験結果より、11/11に近い時期（1か月以内）の情報は重要であること、それ以前の情報は性能にあまり影響しないことを示す。

RQ2については、予測対象販売者に最後にアクセスしてからの経過時間など、商品レベルのデータを必要としない特徴を用いてモデルを構築し、予測精度を比較する。実験結果より、商品レベルの特徴を利用しなくても予測精度は低下しないことを示す。

## 2 関連研究

### 2.1 再購入者予測の研究

再購入者予測に関する研究は、多くの領域で注目されているが、それらはタスクによって分類できる。再購入者予測タスクでは、一定期間内に対象の販売者から購入したユーザが、将来の一定期間内にその販売者から再購入するか否かを予測することを目的とする。同

連絡先：高間 康史、東京都立大学システムデザイン研究科、〒191-0065 東京都日野市旭が丘6-6、ytakama@tmu.ac.jp

<sup>1</sup><https://www.amazon.co.jp/>

<sup>2</sup><https://www.tmall.com/>

<sup>3</sup><https://tianchi.aliyun.com/competition/>

じ商品を再度購入するかを予測するか、あるいは異なる商品を購入する場合も対象とするかは研究によって異なる。Liu らは E コマースを対象として、同じ販売者から将来再度購入するかを予測する手法を提案しており、同じ商品でない場合も対象としている [1]。Zhang らも E コマースを対象としているが、同じ商品を再度購入するかを予測対象としている [2]。

## 2.2 再購入者予測手法

機械学習を用いた再購入者予測に関する研究は、特徴の検討、予測モデルの改良の 2 つのアプローチに大別できる。特徴の検討に関して、Zhang らは本稿と同じ Tmall のデータセットを対象として、ユーザー、販売者、ユーザーと販売者のインタラクションに関する特徴 147 種類の特徴を定義し、ランダムフォレストを用いた埋め込み法や ANOVA を用いたフィルター法などの特徴選択を適用し、予測に有効な特徴について考察している [3]。分析の結果、有効な特徴上位 10 種類のうち、6 種類はユーザーの特徴であること、販売者やユーザーと販売者の組み合わせに関する特徴の中では、購入クリック率と重複購入率に関する特徴が重要であることを報告している。Liu ら [4] も Zhang ら [3] と同様に、ユーザーの特徴、アイテムの特徴やインタラクションに関する特徴について検討している。各販売者、ブランド、カテゴリごとにユーザーが購入した日数などの平均値、分散、最大値を計算し、集約的特徴として定義している。これらの統計量に関する特徴だけでなく、各ユーザーを、購入先販売者を単語とする文書と見なし、主成分分析 (PCA) と潜在ディリクレ配分 (LDA) を適用することで、上位 10 個の主成分座標と 40 個のトピックも特徴として定義している。また、販売者を文書、購入ユーザーを単語とみなして同様の特徴抽出も行っている。合計 1364 種類の特徴を定義し、XG-Boost を用いた埋め込み法による特徴選択の結果、有効な特徴上位 20 種類のうち、10 種類はユーザーの特徴であること、の中でも、販売者あたりの平均購入回数と平均クリック数が重要であること、販売者に関する特徴の中では、ユーザーあたりの平均購入日数とその標準偏差に関する特徴が重要であることを報告している。

予測モデルの改良に関して、Zhu らは CNN(convolutional neural network) と LSTM(Long-Short Term Memory neural network) を組み合わせた予測手法を提案している [5]。CNN は短期間におけるクリックや購入の頻度などのユーザーの行動に見られる局所的な特徴を抽出するために用いられ、LSTM はある商品の購入頻度が時間とともに増加するパターンなどといった、行動傾向の長期的な変化をモデル化するために用いられている。CNN と LSTM を単体で用いた場合と比較した結果、予測精度が向上することを報告している。Yang らはランダムフォレストと LightGBM (Light Gradient Boosting Machine) の予測結果を Soft voting により結合する手法を提案している [6]。過学習に強いランダムフォレストの特徴と、予測精度の高い LightGBM の特徴を組み合わせることで、それぞれ単体で用いた場合よりも予測精度が高くなることを報告している。

これらの研究では、予測精度を高めることが目的であり、データ収集におけるユーザーのプライバシーに関する問題やデータ収集コストは議論されていない。

## 3 分析目的・方法

### 3.1 分析目的

本稿では、再購入予測に利用する特徴に関する知見を得ることを目的として、以下の 2 点をリサーチクエスチョンとする。

- **RQ1**：どの時期の行動が、11/11 当日の行動に影響しているか
- **RQ2**：取得コストやプライバシー保護の観点から、再購入予測に有効な特徴は何か

RQ1 について調査するため、本稿ではベースラインとなる分類モデルを定め、各特徴量の計算に用いるデータの収集期間を変更するなどして構築した各モデルの性能を比較する。ベースラインは、ユーザー、販売者、およびユーザーと販売者のインタラクションに関して、基本的な統計量を計算し、特徴として用いる。また、特徴選択を適用して用いる特徴数を削減する。ベースラインの特徴については、4.1 節で詳細に説明する。

RQ1 は、以下の 2 種類の実験により調査する。

- **実験 1.1**：ベースラインの特徴のうち、データ収集期間を変更して計算可能なものについて、11/11 当日からどこまで遡ってデータを収集するとよいかについて調査する。
- **実験 1.2**：実験 1.1 と同じ特徴について、どの月のデータが予測に影響を与えるかについて調査する。具体的には、各月のデータを除いて特徴量を計算し、分類モデルを構築して性能を比較する。

RQ2 については、ユーザーに関する詳細な情報を用いたほうが予測精度が高くなることが予想されるが、プライバシーの問題が懸念されたり、データ収集コストが高くなるなどの問題が発生する。そこで、RQ1 と同じベースラインから商品レベルの特徴を除去して分類モデルを構築し、除去前と性能を比較する。販売者レ

ベルの情報だけを利用することで、情報取得コストやプライバシー保護の点で利点があると考える。

### 3.2 評価指標

実験では、予測・分類タスクで一般的に使用される3つのモデル:LR, MLP, XG-Boostを学習して、予測性能を比較する。LRとMLPはscikit-learn, XG-BoostはXGBoost<sup>4</sup>を利用した。予測精度の評価にはscikit-learnを利用して、ROC(Receiver Operating Characteristics)曲線の下部分の面積であるAUC(Area Under the ROC Curve)を採用する。

### 3.3 データセット

本稿で使用するデータセットは、中国のオンラインショッピングTmall<sup>5</sup>のデータに基づくコンペティションで用いられたデータセット<sup>6</sup>である。データセット内の各データは、11月11日(独身の日)に商品を購入したユーザの属性情報及び、5月11日から11月11日までのTmall上での行動履歴から構成されている。データを取得した年については明らかにされていない。表1にTmallデータセットに含まれる属性を示す。ユーザの年齢(Age)は9段階に分類され、0は不明、1は18歳未満、2は18歳~24歳、3は25歳~29歳、4は30歳~34歳、5は35歳~39歳、6は40歳~49歳、7、8は50歳以上をそれぞれ表す。本論文では50歳以上は7に統一して扱う。ユーザがとる行動(Action)は4種類に分類され、0はクリック、1はカートに追加、2は購入、3はお気に入りに追加をそれぞれ表す。

表1: データセットに含まれる属性

属性	値	説明
Userid	数値(6桁)	ユーザのID
Itemid	数値(7桁)	商品のID
Catid	数値(4桁)	商品カテゴリのID
Merchantid	数値(4桁)	販売者のID
Brandid	数値(4桁)	商品ブランドのID
Age	1-8	年齢の範囲
Gender	0,1	男性(0)／女性(1)
Time	0511-1111	行動をとった日(mmdd)
Action	0,1,2,3	行動の種類

このコンペティションでは、11月11日にある販売者から初めて購入したユーザが、その後の半年間で同じ販売店から再度購入するかを予測するタスクを対象としている。説明変数(ラベル)は0, -1, NULL, 1の4種類であり、1が再度購入、0が再度購入なし、-1

<sup>4</sup><https://xgboost.ai/>

<sup>5</sup><https://www.tmall.com/>

<sup>6</sup><https://tianchi.aliyun.com/>

は新規顧客ではない(11月11日の購入が初めてではない)ことを意味する<sup>7</sup>。コンペティションではラベルがNULLのデータについて予測した結果を提出すると、推薦精度が評価される。ラベルがNULLのデータに対する正解は公開されていないため、本稿では0, 1のラベルを持つデータを訓練データ、テストデータに分割して用いる。

表2: データセットの概要

ユーザ数	販売者数	商品数	カテゴリ数	ブランド数
424170	4995	1090390	1658	8444

表3: 各行動をとった回数

Action	0	1	2	3
回数	48550713	76750	3292144	3005723

表2はデータセットに含まれるユーザ数などである。ここで、カテゴリとは商品を分類するために設定されたグループであり、1658種類のカテゴリが存在するが、idのみのためどの様なカテゴリであるかを具体的に知ることはできない。ブランドとは、特定の会社やメーカーが提供する商品の名称や商標を指す。これもカテゴリと同じく、idしか公開されていない。表3は、4種類の行動(Action)の回数を示す。表3より、クリック(0)はユーザの最も一般的な行動であるが、購入(2)に至る行動は少ないことがわかる。カートに追加(1)した回数は購入よりも大幅に少ないため、カートに入れずに購入している行動が多いことがわかる。これは、ECサイトで提供されている「即時購入」の機能を用いることで、カートに入れずに直接購入する場合が多いことを意味している。セール期間中は、大幅に値引きされた商品が数量限定で販売されているため、ユーザは商品の売り切れを避けるために、カートを通らずに直接購入することが多いと考える。

表4: 訓練データとテストデータ

データ	データ数	ユーザ数	販売者数	再購入者の割合
訓練	208597	169649	1993	6.09%
テスト	52267	42413	1984	6.21%

分析において、0あるいは1のラベルがついたデータの80%を訓練データ、20%をテストデータにランダムに分割する。このとき、ラベルの比率が訓練データ、テストデータでほぼ同じなるように分割する。表4に訓練データとテストデータの規模を示す。データセットには、同じユーザのデータが複数個存在する場合もあるが、コンペティションでは訓練データに存在しないユーザがテストデータに用いられている。これはデー

<sup>7</sup>新規顧客でない場合は予測対象ではないが、分類モデル構築に利用しても構わないとしてデータセットに含められている。

タリーキを避けるためと考えられるため、本実験でもユーザの重複がないように分割した。販売者については両データで共通である。再購入者（ラベル1）の割合は6%程度であり、ほとんどのユーザが再購入していないことがわかる。モデルの学習においては10分割クロスバリデーションを適用し、訓練データの20%をバリデーションデータとして用いる。

## 4 分析結果

### 4.1 ベースライン

ベースラインは、ユーザ、販売者、およびユーザと販売者のインタラクションに関して、基本的な統計量を計算し、特徴として用いる。また、実験結果の解釈を容易にするため、特徴選択を適用してベースラインに用いる特徴数を削減する。

#### 4.1.1 特徴の説明

ユーザ  $u$  の属性を以下に示す。

- $u_1$ :  $u$  の年齢区分。表1のAgeの値を用いる。
- $u_2$ :  $u$  の性別。Genderの値を用いる。
- $u_3$ :  $u$  がインタラクションした回数。Actionの回数をカウントする。
- $u_4$ :  $u$  が行動をとった日数。Timeに基づき求める。
- $u_5, u_6, u_7, u_8$ :  $u$  がインタラクションした販売者数、商品数、カテゴリ数、ブランド数。Merchantid, Itemid, Catid, Brandidに基づき求める。
- $u_9, u_{10}, u_{11}, u_{12}$ :  $u$  がクリックした回数、カートに追加した回数、お気に入りに追加した回数、購入した回数。同じ商品に対する同じ行動が複数回あった場合はそれぞれカウントする。Actionの回数を行動の種類ごとにそれぞれカウントする。
- $u_{13}$ :  $u$  の購入クリック比（式(1))。

$$u_{13} = \frac{u_{12}}{u_9} \quad (1)$$

- $u_{14}$ :  $u$  の重複購入率（式(2))。

$$u_{14} = \frac{\text{repurchase\_merchants}(u)}{\text{purchase\_merchants}(u)} \quad (2)$$

式(2)において、repurchase\_merchants( $u$ )、purchase\_merchants( $u$ )はそれぞれ  $u$  が複数回購入したことがある販売者数、一度でも購入したことのある販売者数を意味する。

販売者  $m$  の属性を以下に示す。

- $m_1$ :  $m$  に対するインタラクションの回数。
- $m_2$ :  $m$  に対するインタラクションがあった日数。
- $m_3, m_4, m_5, m_6$ : インタラクションがあった  $m$  の商品数、ユーザ数、商品カテゴリ数、ブランド数。
- $m_7, m_8, m_9, m_{10}$ :  $m$  の商品がクリックされた回数、カートに追加された回数、お気に入りに追加された回数、購入された回数。同じユーザから同じ行動が複数回あった場合はそれぞれカウントする。
- $m_{11}$ :  $m$  の購入クリック比（式(3))。

$$m_{11} = \frac{m_{10}}{m_7} \quad (3)$$

- $m_{12}$ :  $m$  の重複購入率（式(4))。

$$m_{12} = \frac{\text{repurchase\_users}(m)}{\text{purchase\_users}(m)} \quad (4)$$

式(4)において、repurchase\_users( $m$ )、purchase\_users( $m$ )はそれぞれ  $m$  から複数回購入したユーザ数、一度でも購入したユーザ数を意味する。

ユーザと販売者の組み合わせに関する属性  $um$  を以下に示す。

- $um_1$ :  $u$  が  $m$  の商品にインタラクションした回数。
- $um_2$ :  $u$  が  $m$  の商品にインタラクションした日数。
- $um_3, um_4, um_5$ :  $u$  がインタラクションした  $m$  の商品数、カテゴリ数、ブランド数。
- $um_6, um_7, um_8, um_9$ :  $u$  が  $m$  の商品をクリックした回数、カートに追加した回数、お気に入りに追加した回数、購入した回数。同じ商品に対する同じ行動が複数回あった場合はそれぞれカウントする。
- $um_{10}$ :  $u$  の  $m$  の商品に対する購入クリック比（式(5))。

$$um_{10} = \frac{um_9}{um_6} \quad (5)$$

#### 4.1.2 特徴選択

本稿では、LR, MLP, XG-Boost を実験で用いるため、特定の分類モデルに有利な特徴選択とならないよう、推定器としてこれらとは異なる決定木を用いてRFE(Recursive Feature Elimination)により特徴選択を行う。重要度の計算には情報ゲインを採用し、特徴数が

表 5: 選択された特徴と重要度

特徴	重要度
$u_3$	0.046
$u_4$	0.051
$u_6$	0.051
$u_7$	0.052
$u_8$	0.043
$u_9$	0.046
$u_{13}$	0.077
$m_{11}$	0.022
$m_{12}$	0.040
$um_1$	0.031

10 となるまで削減する。表 5 に、RFE によって選択された特徴と重要度を示す。

表 5 より、ユーザの特徴が多く選択されているが、これは先行研究 [3,4] の結果と一致する。一方、販売者に関する特徴では、比率に関する特徴（購入クリック比、重複購入率）のみが選択されていることから、ユーザと販売者では異なる性質を持つ特徴が選択されていると言える。また、ユーザと販売者の組み合わせに関する特徴は一つだけであり、特定のユーザ、販売者の組み合わせに限定されるため、データ数が少ないと考えられる。

表 6 は実験で用いる各モデルについて、4.1 節で定義した全ての特徴を用いた場合と、表 5 に示した特徴のみを用いた場合の AUC を示す。表 6 より、どちらの場合も予測精度は低く、本タスクの難易度が高いことがわかるが、特徴数を減らしても性能の低下はほとんどないことがわかる。

表 6: 全ての特徴を使用した場合と特徴選択後の AUC

モデル	全ての特徴を使用	特徴選択後
LR	0.67	0.65
MLP	0.65	0.66
XG-Boost	0.68	0.67

## 4.2 行動時期に関する分析

3.1 節で述べた RQ1 について調査するために、4.1.2 節で選択した各特徴について、データの収集期間を変更して計算する。

### 4.2.1 実験 1.1

実験 1.1 では、11 月 11 日当日からどこまで遡ってデータを収集するとよいかについて調査する。11 月 11 日当日のみ、11 月 11 日の 1 週間前、2 週間前、1 か月前、6 週間前、7 週間前、2 か月前、10 週間前、11 週間前、3 か月前、15 週間前、16 週間前、4 か月前、20 週間前、21 週間前、5 か月前、24 週間前、25 週間前までの各期間をそれぞれデータ収集期間と設定して各特徴を計算する。ベースラインの特徴のうち、 $u_4$  は 11 月 11 日当日のみの場合に全てのデータで 1 となるため、予測精度が低下することが考えられる。その場合、収集期間が与える影響の分析が正しく行えない可能性があるため、本実験では対象外とする。また、割合に基づく特徴である  $u_{13}$ 、 $m_{11}$ 、 $m_{12}$  も 11 月 11 日当日のみの場合に意味がないと判断し、対象外とした。図 1 に、LR、MLP、XG-Boost の各モデルについて、データ収集期間の違いによる AUC の変化を示す。

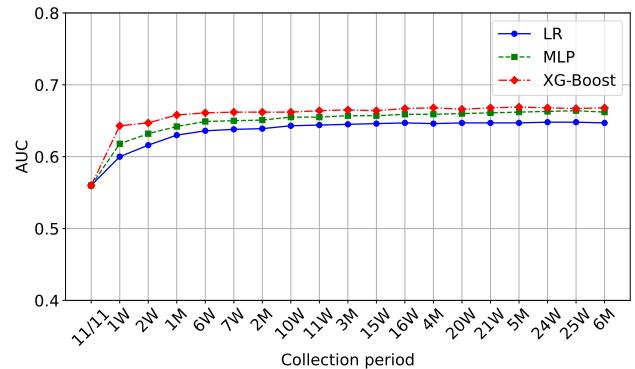


図 1: データ収集期間による AUC の変化

図 1において、一番右の 6M はデータセットの全期間のデータを用いた場合であり、表 6 に示した結果（特徴選択後）と同一である。図 1 より、どのモデルでも 11 月 11 日当日のデータのみでは不十分であることがわかる。また、1 か月前までは AUC が向上するが、それ以前のデータは性能にあまり影響しないと言える。

### 4.2.2 実験 1.2

実験 1.2 では、どの月のデータが予測に影響を与えているかについて調査する。5 月から 11 月のうちの各月を除いてそれぞれ各特徴を計算する。図 2 に、各月を除去した場合の各モデルの AUC を示す。

図 2 において、一番左の None はデータセットの全期間のデータを用いた場合であり、表 6 に示した結果（特徴選択後）と同一である。図 2 より、11 月 11 日当日に近い月（11 月、10 月）のデータを除いた場合に AUC は低下するが、9 月以前のデータは除去しても性能にあまり影響しないと言える。

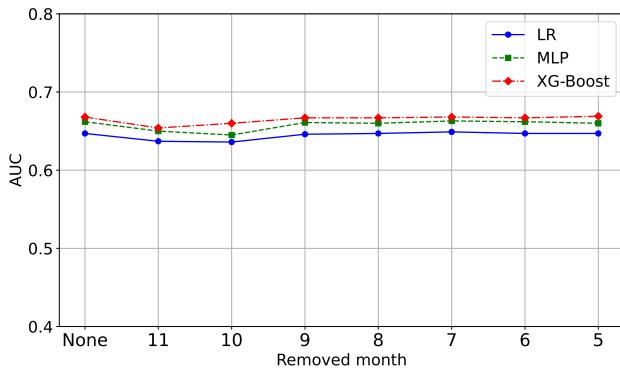


図 2: 各月を除去した場合 AUC の変化

実験 1.1 と実験 1.2 両方の結果から、11月 11 日当日から 1か月前くらいまでの情報が重要であると言える。

### 4.3 ユーザ情報に関する分析

実験 2 では、RQ1 と同じベースラインで用いているユーザに関する特徴のうち、商品レベルの情報を利用するものを除去して分類モデルを構築し、除去前と性能を比較する。具体的には、Itemid, Catid, Brandid をそれぞれ利用する  $u_6$ ,  $u_7$ ,  $u_8$  を除去する。表 7 は実験で用いる各モデルについて、商品レベルの特徴を利用した場合（ベースライン）としない場合（商品レベル特徴不使用）の AUC を示す。

表 7: 商品レベルの特徴の有無による AUC の比較

特徴	LR	MLP	XG-Boost
ベースライン	0.65	0.66	0.67
商品レベル特徴不使用	0.65	0.66	0.67

表 7 において、モデルの予測精度に変化はなく、商品レベルの特徴は収集不要と言える。

## 5 結論

本稿では、再購入者予測タスクを対象として、Tmall のデータセットを分析し、どの時期の行動が、11/11 当日の行動に影響しているか、および取得コストやプライバシー保護の観点から有効な特徴は何かについて検討した。実験により、11月 11 日当日に近い時期（1か月以内）の情報は重要であること、それ以前の情報は性能にあまり影響しないことを示した。後者に関しては、商品レベルの特徴を利用しなくても予測精度は低下しないことを確認した。

今後の計画としては、11/11 当日のログデータを分析し、予測対象となる販売者のアイテムに対し連続し

て実行された行動系列に基づく特徴を導入する予定である。この特徴は 1 セッションの情報のみが得られる場合を想定したものであり、アカウントがないユーザにも適用できるため、プライバシー保護効果が高いと考える。また、予測精度の向上を目指すため、他の分類モデルでの実験も計画している。

## 謝辞

本研究は JSPS 科研費 JP22K19836, JP23K21724, and JP23K24953 の助成を受けたものである。

## 参考文献

- [1] Liu, Y., Zhang, H., and Ren, H.: An integrated learning-based prediction model for purchasing propensity of jingdong visitors, *Highlights in Science, Engineering and Technology*, Vol. 70, pp. 60-66 (2023).
- [2] Zhang, W. and Wang, M.: An improved deep forest model for prediction of e-commerce consumers' repurchase behavior, *Plos one*, Vol. 16, Issue 9 (2021).
- [3] Zhang, M., Lu, J., Ma, N., Cheng, T.E., and Hua, G.: A feature engineering and ensemble learning based approach for repeated buyers prediction, *International Journal of Computers Communications & Control*, Vol. 17, No. 6 (2022).
- [4] Liu, G., Nguyen, T.T., Zhao, G., Zha, W., Yang, J., Cao, J., Wu, M., Zhao, P., and Chen, W.: Repeat buyer prediction for e-commerce, KDD2016, pp. 155-164 (2016).
- [5] Zhu, C., Wang, M., and Su, C.: Prediction of consumer repurchase behavior based on LSTM neural network model, *International Journal of System Assurance Engineering and Management*, Vol. 13, Suppl 3, pp. 1042-1053 (2022).
- [6] Yang, L., Niu, X., and Wu, J.: RF-LighGBM: a probabilistic ensemble way to predict customer repurchase behaviour in community e-commerce, *arXiv preprint arXiv*. 2109.00724 (2021).

# 広告アクセスログからユーザのモーメントを推定する 可能性についての検討

## An Investigation of the Method for Estimating User Moments from Online-Advertising Access Logs

大藏真奈美<sup>1\*</sup> 笹嶋宗彦<sup>1</sup>  
Manami Ohkura<sup>1</sup> Munehiko Sasajima<sup>1</sup>

<sup>1</sup> 兵庫県立大学  
<sup>1</sup> University of Hyogo

**Abstract:** デジタル化が進展し、ウェブ広告の重要性が高まっている。Google が提唱する ZMOT 理論によれば、インターネット上で購買行動を行う消費者は、商品に出会う前に購買意思決定を終えているとされており、広告を適切なタイミングで表示する意味でのターゲティング精度向上が求められている。本研究では、ウェブ上のユーザ行動を分析し、購買の兆候を捉えることを目指す。アサエルの購買行動類型を用いてクリックデータからユーザを分類し、最適な広告配信のタイミングとコンテキストターゲティングへの応用可能性を検討する。

### 1 はじめに

社会のデジタル化が進展する中で、インターネットは人々の生活や消費行動に大きな影響を与えており、広告業界においてもその重要性はますます高まっている。その結果、総広告費に占めるインターネット広告費は着実に増加しており、日本総務省「令和6年版情報通信白書」[1]によれば、2023年には総広告費全体の45.5%を占めるに至っている。この傾向は、従来のテレビや新聞といったマスメディア広告の影響力が相対的に低下する一方で、インターネット広告が広告市場の中心的な存在になりつつあることを示している。また、スマートフォンやタブレット端末などのデバイスの普及により、ユーザがオンラインで情報にアクセスする頻度が増加している点も、この成長を後押ししている。

広告マーケティングにおいて、消費者は従来のように単に受動的に広告を受け取るだけでなく、購買行動に至る前の段階で積極的に情報収集を行うことが知られている。特に、Google が提唱する ZMOT 理論 (Zero Moment of Truth) [2] によれば、消費者は製品やサービスに「出会う」前の段階で、インターネット上の口コミ、レビュー、SNS での情報などを基に購買の意思決定を終えている場合が多いとされている。この理論は、現代の消費行動において、消費者が製品やブランドと初めて接触する前にオンラインで膨大な情報を処理している現状を反映している。

\*兵庫県立大学社会情報科学部社会情報科学科  
〒 651-2197 兵庫県神戸市西区学園西町 8 丁目 2-1  
E-mail:fa21y013@guh.u-hyogo.ac.jp

しかし、こうした消費者の行動に対して、従来のターゲティング手法は十分に対応できているとは言えず、現行の広告配信技術では、消費者が既に購買の意思決定を終えた後に広告が配信されるケースが少なくない。実際、ユーザが閲覧するウェブページの内容に基づいて広告を表示するコンテキストターゲティングや、ユーザの過去の閲覧履歴やクリックデータに基づいて広告を表示する行動ターゲティングは、既にユーザがその製品について情報探索を行ったデータに由来している可能性がある。このような広告配信は、意思決定の前段階で広告を配信する場合と比較して、広告効果が低下することが考えられる。また、心理学者の Brehm が提唱した心理的リアクタンス理論 (Psychological Reactance Theory) [3] に基づき、消費者にとって不要なタイミングで配信された広告は、無関心で終わるだけでなく、場合によってはブランドに対する否定的な印象を与えるリスクすら伴うことが考えられる。このような課題を解決するためには、消費者の購買行動における意思決定プロセスをより詳細に理解し、そのプロセスに即した広告配信が求められる。

そこで、本研究ではインターネット上のユーザ行動データを追跡し、購買行動における「商品を求める瞬間」の予兆を予測することを目指す。この「商品を求める瞬間」とは、消費者が具体的な購入行動に移る直前のタイミングを指し、このタイミングでの広告配信が最も効果的であると考えられる。本研究では、消費者行動の分析において、アサエルの購買行動類型 [4] を基盤とし、クリックデータやアクセスログから購買行

動や購買意思決定の予兆を捉えるための手法を検討する。特に、ユーザが購買意思決定に至る過程で購買意識のトリガー（トリガーモーメント）[5]となる行動を抽出し、それらがどのような役割を果たしているかを明らかにする。

さらに、「商品を求める瞬間」の予兆を捉えることによるターゲティングの精度向上や、予兆となりやすいサイトの抽出によるコンテキストターゲティングへの応用可能性を検討する。本研究の成果は、企業にとって効率的な広告配信の実現を後押しするとともに、消費者にとっても興味やニーズに適した広告を届けることで、広告に対する満足度の向上や購買体験の質の向上をもたらすことが期待される。

## 2 関連研究

消費者の購買意思決定が一貫したパターンで行われるわけではなく、さまざまな要因によってそのプロセスが影響を受けるとされている。オンラインショッピングにおける購買プロセスに関して、秋山らが提唱する AISAS モデル [6] が存在する。AISAS モデルでは、ユーザの購買行動が、Attention (注意) → Interest (興味) → Search (検索) → Action (行動) → Share (共有) という段階を踏むとされている。

ユーザの購買行動の過程において、購買意思決定が状況や感情、情報提示の方法にどのように影響されるかについては、さまざまな理論が提唱されている。以下に代表的な理論を示す。

心理学者 Higgins が提唱した制御焦点理論 (Regulatory Focus Theory) [7] によれば、消費者の意思決定は「促進焦点（利益を最大化）」と「予防焦点（損失を最小化）」のいずれかのモチベーションに影響を受ける。消費者の状況や感情に応じて、どちらの焦点が優勢となるかが異なり、それによって意思決定プロセスも変化するとされている。

次に、社会心理学者 Petty らによるエルム説得モデル (Elaboration Likelihood Model, ELM) [8] では、消費者が意思決定時に情報を処理する際、高関与時には論理的で詳細な情報を重視する中心ルートを用い、低関与時には感情や単純な手がかりを重視する周辺ルートを用いるとされている。

行動経済学者 Kahneman らが提唱するフレーミング効果 (Framing Effect) [9] によると、消費者の意思決定は、情報の提示方法によって大きな影響を受ける。例えば、「成功率 80%」と提示された場合と「失敗率 20%」と提示された場合では、同じ事実を伝えているにもかかわらず、受ける印象が異なることが知られている。

また、消費者行動分野の研究者である Russel が提唱した状況的要因モデル (Situational Influence Model)

[10] では、消費者の意思決定が、時間の制約や物理的環境、他者の存在などの状況的要因に強く影響し、状況によっては通常の意思決定プロセスが変更される場合があるとされている。

しかしながら、これらの理論は消費者の感情や状況に依存する要因を主に取り扱っており、インターネット上の消費者行動データを直接的に読み取って適用するには限界がある。そこで、本研究では、マーケティングおよび消費者行動研究者である Assael が提唱したアサエルの購買行動類型 [4] を活用する。この類型では、購買行動を情報処理型、不協和解消型、バラエティ・シーキング型、習慣型の 4 つに分類している。各分類に共通した「商品を求める瞬間」の予兆が存在するという仮説を立て、分析を行う。

## 3 アサエルの購買行動類型

本章では、本研究の基盤となる「アサエルの購買行動類型」の特徴と、それぞれの購買行動を促すトリガーモーメントについて説明する。

消費者の購買行動に関する理論は多岐にわたるが、Assael によるアサエルの購買行動類型 (Assael's Model of Consumer Buying Behavior) は、消費者行動を体系的に分析するための代表的なフレームワークのひとつである。アサエルの購買行動類型を表 1 に示す。この理論では、消費者が製品やサービスを選択する際に直面する「関与度」と「ブランド間の知覚差異」を基準に購買行動を 4 つに分類している。「関与度」とは、消費者が製品やサービスの購入に際してもつ「関心」や「重要性の認識」の度合いを指し、「ブランド間の知覚差異」とは、消費者が製品カテゴリー内で異なるブランドに対して感じる「違い」や「独自性」の程度を意味する。以下では、この 4 つの購買行動の特徴と、それぞれの購買行動を促すトリガーモーメントについて述べる。

表 1: アサエルの購買行動類型

		関与度	
		高	低
ブランド間 の知覚差異	大	情報処理型	バラエティ・ シーキング型
	小	不協和解消型	習慣型

出所： H. Assael, Consumer Behavior and Marketing Action, Boston: Kent Publishing

情報処理型は、購入しようとしている商品に対する関与度が高く、ブランド間の知覚差異が大きい場合の消費者行動である。消費者は、時間をかけて情報を収

集し、製品の特徴を比較した上で購買の意思決定を行う。この行動は、特に高価な耐久消費財の購入に多く見られ、例として、自動車や住宅が挙げられる。

情報処理型におけるトリガーモーメントとは、消費者が製品に関心をもつききっかけや購買を後押しする要因を指す。消費者行動研究者 Bettman が提唱した Bettman モデル [11] によれば、情報探索のプロセスでは、初期段階で自身の記憶や過去の経験に基づく内部情報を利用し、次第に能動的な外部情報の収集に移行するとされる。そのため、広告や口コミ、製品体験などで消費者の記憶に残る情報が、後に情報探索を開始する際のトリガーモーメントとして機能する。例えば、消費者がある自動車メーカーを広告で認知し記憶に残ると、購入を検討し始めた際に、そのメーカーを起点に情報を探す。

不協和解消型は、購入しようとしている商品に対する関与度が高く、ブランド間の知覚差異が小さい場合に該当する。このような消費者行動は、比較的短時間で情報収集を終えて購入を決定するものの、購入後に「この選択が正しかったのか」と不安を感じることが多い。この不安を軽減するため、購入後の広告や他者の評価が重要な役割を果たす。具体例として、洗濯機や USB ケーブルの購入が挙げられる。

不協和解消型において、直接的なトリガーモーメントは少ないとされる。しかし、購入後に接触する広告や肯定的な評価は、消費者の不安を軽減し、次回の購入時にポジティブな影響を与える可能性がある。例えば、家電購入後に高評価レビューや高い満足度の広告を見ると、選択は正しかったと感じることがある。ブランドへの信頼が生まれ、次回の購入で選択肢として優先されやすくなる。このように、過去の購入経験がトリガーモーメントとして作用しやすい特徴がある [12]。

バラエティ・シーキング型は、購入しようとしている商品に対する関与度が低く、ブランド間の知覚差異が大きい場合に見られる消費者行動である。消費者は新しい体験を求め、ブランドを切り替える傾向が強い。この行動は、加工食品や出張時の経費支出に見られることが多い。

バラエティ・シーキング型におけるトリガーモーメントは、消費者の「飽き」や「退屈」などの内的要因、および広告やパッケージデザインによる衝動買いの誘発などの外的要因に着目する必要がある。また、促進焦点と予防焦点の双方を考慮することが重要である。先述した心理学者 Higgins の制御焦点理論 [7] によれば、消費者は「促進焦点（成長や達成への関心）」または「予防焦点（失敗やリスク回避への関心）」に基づいて行動するとされる。例えば、種類が豊富である即席カップめんを選ぶ際、「期間限定で特別価格」という広告は、促進焦点の消費者に行動することで利益を得られるという認識を促し、「在庫僅少」や「売り切れ次第終了」

という広告は、予防焦点の消費者に行動しないと後悔する可能性を意識させる。これにより、いつもと異なる商品への購買意図が強化される。このため、広告戦略において両方の焦点に訴求することが効果的である。

習慣型は、購入しようとしている商品に対する関与度が低く、ブランド間の知覚差異が小さい場合に該当する。このタイプの消費者は、情報探索に労力をかけることなく、慣性で特定のブランドを選択する傾向が強い。日常的に使用する消耗品の購入がこれに該当し、具体例として水や日用雑貨が挙げられる。

習慣型の購買行動におけるトリガーモーメントとしては、広告の「ノイズ作用」が重要である。ノイズ作用とは、広告に繰り返し接触することで、消費者に潜在的にブランドを記憶させる効果を指す [13]。例えば、日常的に繰り返し表示される洗剤の広告は、消費者の記憶に残りやすく、消費者は購入時に無意識にそのブランドを選ぶ傾向がある。このように、ノイズ作用で形成された記憶は無意識のブランド選択に影響し、購入のトリガーモーメントとなる。

## 4 実データの分析

本研究は、実データを基に異なる購買行動類型のユーザが商品に関心を持つタイミングやクリックパターンや行動が、コンバージョン（以下、CV と表記）に繋がるプロセスを観察することで、各購買行動類型に共通する「商品を求める瞬間」の予兆を特定し、購買行動における特徴的な傾向を体系的に明らかにすることを目的とする。

分析には、広告企業 A が保有するデータを用いて、特にユーザのクリック行動系列データを活用した。このデータは、広告企業 A が収集した一定期間内のログデータを正規化したもの（以下、正規化ログと表記）であり、ユーザのインターネット上のクリック情報が含まれ、購買行動の各段階における傾向を把握するため有用である。例えば、情報探索を目的として特定の商品ページを何度も閲覧する行動や、クリック頻度が増加するパターンが CV に至る兆候として観測されることが期待される。

これらの観測を通じて、ユーザ行動の定量的な特徴を抽出し、購買行動類型ごとの具体的な予兆を明らかにすることを試みる。分析するにあたり、正規化ログの 18,227 ユーザ、全 1,437,130 件の URL の中で、少なくとも 1 件の CV に関連する URL を閲覧したユーザに絞り込み、最終的に 189 ユーザ、全 32,335 件の URL を分析対象とした。あるユーザのクリック行動系列の例を表 2 に示す。

表 2: あるユーザのクリック系列

日時	サイト名
2024-06-10 10:18:10	音楽関連ニュースサイト (1)
2024-06-10 10:59:07	ファイル転送サービス
2024-06-10 12:23:20	音楽・ライブレポート
2024-06-10 13:52:18	音楽関連ニュースサイト (2)
2024-06-10 22:11:11	宿泊予約サイト (エリア A のホテル・旅館)
2024-06-10 22:11:44	宿泊予約サイト (エリア B のホテル・旅館)
2024-06-10 22:12:23	宿泊予約サイト (エリア A のホテル・旅館)
2024-06-10 22:12:36	宿泊予約サイト (エリア C のホテル・旅館)
2024-06-11 00:30:06	ファイル転送サービス
2024-06-11 09:43:59	音楽関連ニュースサイト (1)
2024-06-11 09:50:14	ファイル転送サービス
2024-06-11 13:14:23	音楽関連ニュースサイト (1)
2024-06-11 13:15:26	音楽関連ニュースサイト (1)
2024-06-11 13:17:35	音楽関連ニュースサイト (1)
2024-06-11 16:34:33	ファイル転送サービス
2024-06-11 18:45:47	音楽関連ニュースサイト (1)
2024-06-11 22:18:13	音楽関連ニュースサイト (1)
2024-06-12 13:59:22	音楽関連ニュースサイト (2)
2024-06-13 16:03:01	音楽関連ニュースサイト (1)
2024-06-13 16:03:37	音楽関連ニュースサイト (1)
2024-06-13 20:53:50	書籍・メディア購入サイト
2024-06-13 21:28:27	宿泊予約サイト
2024-06-14 14:30:41	音楽関連ニュースサイト (1)
2024-06-14 15:24:13	宿泊予約サイト (エリア D のホテル・旅館)
2024-06-14 15:29:36	宿泊予約サイト (エリア D)
2024-06-14 15:29:41	宿泊予約サイト (ホテル・旅館 A)
2024-06-14 15:31:02	宿泊予約サイト (エリア D のホテル・旅館)
2024-06-14 15:34:23	宿泊予約サイト (予約詳細)
2024-06-14 16:41:11	音楽関連ニュースサイト (2)
2024-06-14 16:42:14	アーティスト A の公式サイト
2024-06-14 16:43:10	音楽関連ニュースサイト (2)
2024-06-14 16:47:28	書籍・メディア購入サイト
2024-06-14 17:43:26	音楽関連ニュースサイト (2)
2024-06-14 18:03:24	歌詞検索サイト
2024-06-14 18:11:25	ギターコード譜サイト
2024-06-14 20:13:54	音楽関連ニュースサイト (2)
2024-06-14 20:31:50	ファンション誌サイト

#### 4.1 コンバージョンデータの抽出

CV データを抽出する処理を行う。広告企業 A では、商品購入を目的とする広告よりも、商品の認知や、リアル店舗への誘導を目的とする広告が多いため、保有する CV データの量に限界がある。また、CV 設定の定義が一貫しておらず、実際に製品を購入した際に CV とする場合もあれば、商品詳細ページを閲覧しただけで CV とみなす場合も存在する。そのため、正規化ログのクリックデータから CV データを抽出する方法を採用する。

本研究では、売上やリード（見込み客）の獲得に直接関連する行動を CV と定義し、購入完了ページや資料請求完了ページを CV として抽出する。実際に、正規化ログにおいて、自動車の試乗予約ページや宅配ピザの顧客情報入力ページが該当した。

#### 4.2 アサエルの購買行動類型の付与

クリックデータに対して、アサエルの購買行動類型をユーザ単位で付与する方法について説明する。クリックデータの系列を入力し、その系列に含まれる各 URL に対して、ユーザ単位でアサエルの購買行動類型の 4 類型のいずれかに分類する。この過程では、ChatGPT を活用し、ユーザの行動データから適切な購買行動類型を自動的に判別する。具体的には、上記のクリック系列を ChatGPT に入力し、アサエルの購買行動類型のどれに該当するかを出力させる。以下に、ChatGPT のプロンプトおよび回答例を示す。ただし、以下のプロンプト例では、サイト名を加工して表示しているが、実際には、具体的な商品名やサービス名などを ChatGPT に入力する。

(ChatGPT プロンプト例)

以下は、あるユーザが閲覧したサイトを時系列順に表示している。各閲覧サイトの最後の括弧に「CV」がついている場合は、そのユーザがコンバージョンしたサイトであることを示す。以上を踏まえ、このユーザの購買行動をアサエルの購買行動類型の「情報処理型」「不協和解消型」「バラエティ・シーキング型」「習慣型」のいずれに分類するべきか教えてください。

クリック系列（表 2 の通り）：

1. 音楽関連ニュースサイト (1)
2. ファイル転送サービス
- （略）
28. 宿泊予約サイト 予約詳細 (CV)

(略)  
37. ファッション誌サイト

(ChatGPT 回答例)

このユーザの購買行動を分析し、アサエルの購買行動類型に分類すると、「情報処理型」に該当すると考えられます。その理由は以下の通りです。

### 1. ユーザ行動の特徴

継続的な情報収集：ユーザは「宿泊予約サイト」関連のサイトを何度も閲覧し、具体的な宿泊施設の情報を繰り返し調べています。特定の地域（エリア A-D）や施設の詳細を比較しており、これは購入（宿泊予約）の意思決定に際して多くの情報を収集し、慎重に検討していることを示しています。複数の選択肢の比較と評価：同じ宿泊施設を複数回閲覧していることから、料金や立地、設備などの比較を繰り返していると推測されます。この行動は、意思決定前に多くの選択肢を検討する「情報処理型」の典型的な行動です。

購入（CV）への計画的アプローチ：CV 前には複数の「宿泊予約サイト」の宿泊施設ページを訪問しており、最終的に特定の施設を予約しています。この一連の流れは、熟慮の上で意思決定が行われたことを示唆しています。

### 2. 他の分類でない理由

不協和解消型でない理由：購入後に不安を解消するための追加情報収集やレビュー確認などの行動が見られません。むしろ購入前に十分な情報収集を行っており、不安を感じる余地が少ない行動です。

バラエティ・シーキング型でない理由：ユーザは新しい体験を求めてランダムに選択肢を切り替えているわけではありません。同じ地域や宿泊施設を複数回閲覧し、慎重に選択肢を絞り込む行動が観察されます。

習慣型でない理由：情報探索に労力をかけることなく過去の経験や慣性で選択している様子は見られません。むしろ意思決定に多くの時間を費やしています。

### まとめ

このユーザは、購入前に多くの時間を費やして情報を収集し、選択肢を比較・評価したうえで慎重に購入を決定する「情報処理型」の購買行動に該当します。

## 5 商品ごとに見られる購買行動型と考察

分析の結果、各購買行動類型において予兆として捉えられる行動パターンが異なることが分かった。以下の節では、各購買行動類型の分析結果と考察について述べる。

### 5.1 情報処理型の商品購入における予兆

表 2 で示したユーザ購買行動は、情報処理型に分類された。このユーザは、宿泊施設の予約という特定の目的のために、購入前に十分な時間をかけて宿泊予約サイトを何度も訪問し、同じ地域の異なる施設や、同一施設を複数回閲覧する行動が見られ、価格、立地、設備、レビューなど多くの要因を慎重に比較していることが確認できる。

比較検討の過程で、ある有名な宿泊予約サイトに限定した閲覧が見られ、ユーザの過去の経験や記憶に基づき、宿泊予約の際にはこの宿泊予約サイトに頼るという潜在的な考えをもっている可能性がある。したがって、宿泊予約のような目的ができる前にユーザ認知を図ることが、目的が生じた際にトリガーモーメントとして効果的であると考えられる。また、ユーザに対して、価格、設備、レビューなどの明示や「今だけ」などの限定情報を提示することで、最終的な意思決定を後押しすることが考えられる。

### 5.2 不協和解消型の商品購入における予兆

インターネット上で CV を終えた後で、その CV に関連する別のサイトを閲覧したユーザが不協和解消型に分類される場合がいくつか確認された。例えば、家事代行サービスで CV をしたユーザが不協和解消型に分類された。このユーザは、CV 前に「家事代行サービス A」に関連した別のサイトでの比較検討が一切見られず、「家事代行サービス A」のお見積り・ご予約フォームで CV を発生して、サービス選択を終了させた後に、類似のサービスを扱う「清掃サービス B」のサイトを閲覧していた。「家事代行サービス A」での情報収集が限定的だったことが、他社サービスの閲覧を促した可能性がある。また、インターネット上で関連商品やサービスの広告が表示されることが、選択の正当性への疑念を助長したことや、サービスの結果がまだ得られていない状態では、選択の正しさに対する不安が生じやすいことが「清掃サービス B」を閲覧するきっかけになったことが考えられる。

不協和解消型は、トリガーモーメントが起こりにくい類型ではあるが、購入後の不安感を解消するための

FAQ, レビュー, 満足度調査結果などツールを提供することが, 購入した商品・サービスのリピートに繋がることが考えられる。

### 5.3 バラエティ・シーキング型の商品購入における予兆

正規化ログにはバラエティ・シーキング型に分類されるユーザが存在しないため, 分析は不可能であった. 対象の正規化ログを数日間に限定していることから, CV 前の検索数や閲覧数に一定の限界があり, 通常の購買行動と新しい体験を求めてブランドを切り替える行動との違いを判断することが困難であった。

しかしながら, バラエティ・シーキング型の特徴を考慮すると, ユーザの購買行動に関係なく, 情報をちりばめることがトリガーモーメントとして機能する可能性が考えられる。

### 5.4 習慣型の商品購入における予兆

CV 前後で, 他の選択肢や競合を比較検討するような行動が見られることなく CV に至るよう, 購入までの意思決定がルーチン化したユーザの購買行動は習慣型に分類されやすい傾向にある. 特に, 競馬サイトで CV が見られるユーザが習慣型に多く分類された。

慣れたサイトを繰り返し利用し, 効率的に CV に到達するという特徴をもつ習慣型では, 特定サイトでの購買を習慣化させることが重要である. また, 習慣化しているユーザに対して, ユーザの既存の行動パターンを邪魔しないように配慮しながら, 自然な流れで購買を発生することが習慣的なトリガーモーメントの引き出し方であると考えられる. さらに, 新しい機能や関連サービスを提示することで, 無理のない形で行動範囲を広げる可能性がある。

## 6 おわりに

本研究では, アサエルの購買行動類型を基盤に, 広告企業 A が保有する正規化ログを活用し, インターネット上の消費者購買行動における「商品を求める瞬間」の予兆を推測する方法について検討した. 特に, クリック系列データを詳細に分析することで, 消費者が購買行動を起こすきっかけやそのプロセスを明らかにし, 各購買行動類型における特徴的な傾向を示した. 今後も, 正規化ログの分析と考察を引き続き行っていく予定である。

今後の展望としては, クリック系列データを基にリアルタイムでアサエルの購買行動類型に基づくユーザ

分類を行い, その分類結果に応じた広告配信を自動化する仕組みを構築することが挙げられる. このような仕組みにより, 消費者の購買行動に即した効率的な広告戦略の実現が期待される. しかしながら, 近年進展する Cookie の廃止やプライバシー保護の強化といった技術的・法的な課題が, データの利用に制約を及ぼす可能性がある点も留意が必要である。

本研究で得られた成果をさらに一般化して, 購買の予兆検出を定式化できればと考えている。

## 参考文献

- [1] 総務省: 令和 6 年版情報通信白書, 総務省 情報通信政策研究所, 第 2 部, 第 1 章, 第 3 節, 第 2 項 (2023)
- [2] Jim Lecinski: ZMOT: Winning the Zero Moment of Truth, *Google Inc.* (2011)
- [3] J. W. Brehm: A Theory of Psychological Reactance, *Academic Press* (1966)
- [4] H. Assael: Consumer Behavior and Marketing Action, *International Thomson Publishing* (1987)
- [5] 電通: なぜ、電通が CX に挑戦するのか。右脳力と左脳力の両方でつくりあげる New CX へ, 電通公式ウェブサイト (2023)
- [6] 秋山隆平, 杉山恒太郎: ホリスティック・コミュニケーション: アクティブ・コンシューマーの出現で進化する広告と販促の境界, 宣伝会議 (2004)
- [7] E. Tory Higgins: Promotion and Prevention: Regulatory Focus as a Motivational Principle, *Advances in Experimental Social Psychology*, Vol. 30, pp. 1–46 (1998)
- [8] Richard E. Petty, John T. Cacioppo: The Elaboration Likelihood Model of Persuasion, *Advances in Experimental Social Psychology*, Vol. 19, pp. 123–205 (1986)
- [9] Daniel Kahneman, Amos Tversky: Choices, Values, and Frames, *American Psychologist*, Vol. 39, No. 4, pp. 341–350 (1984)
- [10] Russell W. Belk: Situational Variables and Consumer Behavior, *Journal of Consumer Research*, Vol. 2, No. 3, pp. 157–164 (1975)
- [11] James R. Bettman: An Information Processing Theory of Consumer Choice, Chapter 2, *Addison-Wesley* (1979)
- [12] Morgan, R.M., Hunt, S.D.: The Commitment-Trust theory of relationship marketing, *The Journal of Marketing*, Vol. 58, No. 3, pp. 20–38 (1994)
- [13] 電通: ノイズ作用がブランド選択に影響を与え購入のトリガーモーメントとなる, 電通報 (2023)

# 子育て世代が暮らしやすい町の実現を目指した 兵庫県加古郡播磨町における3歳児健診データ分析の試み

## Analyzing 3-Year-Old Child Health Examination Data in Harima-town to Realize a Safe and Comfortable Life for the Child-Rearing Generation

後藤快斗<sup>1\*</sup> 西尾理津子<sup>2</sup> 前川啓子<sup>2</sup>  
Kaito Goto<sup>1</sup> Ritsuko Nishio<sup>2</sup> Keiko Maekawa<sup>2</sup>  
工藤美子<sup>1</sup> 能町しのぶ<sup>1</sup> 笹嶋宗彦<sup>1</sup>  
Yoshiko Kudo<sup>1</sup> Shinobu Nomachi<sup>1</sup> Munehiko Sasajima<sup>1</sup>

兵庫県立大学<sup>1</sup>

<sup>1</sup> University of Hyogo

播磨町<sup>2</sup>

<sup>2</sup> Harima-town

**Abstract:** 本研究が対象とする兵庫県加古郡播磨町は「安心して子育てができる町」を目指しているが、乳幼児健診の結果に応じた支援すべき児の判断基準は組織内で明文化されておらず、ベテランの保健師や医師の判断に頼っているのが現状である。経験の少ない職員には、その判断を補助する仕組みが必要であるため、本研究は、過去の事例から、支援が必要な児に適切なタイミングと方法で支援を進める仕組みづくりの実現を目指している。本稿では、3歳児健診とフォローランチ帳を用いたデータ分析の試みについて紹介する。

### 1 はじめに

厚生労働省[1]は、2021年度の1歳6か月児健診の受診率が95.2%，3歳児健診の受診率が94.6%と発表しており、乳幼児健診は受診率の高い健診である。国立研究開発法人 国立成育医療研究センター[2]は、乳幼児健診で取り扱う健康課題は重層化していることが特徴であり、健康課題のスクリーニングの視点だけではなく、支援（サポート）の視点が必要となっている。これに呼応して医師・歯科医師、保健師、看護師、助産師、歯科衛生士、管理栄養士・栄養士、心理職、保育士などの多くの職種がかかわりをもつようになつたと述べており、乳幼児健診は複雑化、多様化していることが分かる。

本研究が対象とする兵庫県加古郡播磨町は「安心して子育てができる町」を目指しているが、乳幼児健診の結果に応じた支援すべき児の判断をベテランの保健師や医師に頼っており、判断基準が組織内で明文化されていないという課題を抱えている。一般に、ベテランの保健師は、受診者や保護者に安心感を与える対応ができる。子育てにおいては、保護者にとっても初めてのことが多く起きたため、心配ぎみになってしまつ

たり、逆に、楽観的になつたりすることがある。心配し過ぎる人に対しては、安心させつつ診察結果を伝える必要があり、楽観的過ぎる人に対しては、不安にさせないようにしつつも、事実を正しく伝える必要がある。ベテランの保健師に担当されたことで、安心した経験を持つ人は多いと考えられる。

こうしたベテランと同等な対応を、より多くの保健師が出来るようになれば、「安心して子育てができる町」に一步近づくのではないかと筆者らは考えている。後述するように、乳幼児健診では多くの項目についてデータを得るのであるが、ベテランの保健師は、それらデータから、目の前の児について判断するだけではなく、その事実をどのように伝えるのが適切なのかを読み取れているのではないかと考えている。この目標を達成するためには、乳幼児健診のデータから、ベテランの保健師と同等な判断ができるように補助する仕組みが必要である。新人保健師が不安を抱え、支援する仕組みが必要であることは、複数の研究者[3][4][5][6]により指摘されている。

本研究は、過去の事例から、支援が必要な児に適切なタイミングと方法で支援を進める仕組みづくりの実現を目指す。そのための具体的な方法として、過去の乳幼児健診の結果と、その受診者に対して行った支援の有無、その支援に至った経緯、実際の支援内容についてのデータを分析することによって、「適切な支援」が、

\*連絡先：兵庫県立大学社会情報科学部社会情報科学科  
〒 651-2197 兵庫県神戸市西区学園西町 8 丁目 2-1  
E-mail:fa21e037@guh.u-hyogo.ac.jp

「どのような判断」によって行われたかを分析し、明文化する。そのうえで、乳幼児健診を担当している保健師を含めた本研究看護職メンバーと、明文化された知識を分析し、小児発達支援についてベテランの保健師の判断の仕方を明らかにする。明文化した知識を参照することにより、経験の少ない保健師らもベテランと同等な判断をすることができるようになり、乳幼児健診と支援事業全体の質を底上げできると考えている。

本研究では、3歳児健診で支援対象となる児には、運動の発達と、言語の発達に特徴があると分かった。また、保護者の不安にも起因していることが明らかになりつつある。本稿は、「安心して子育てができる町」の実現へ向けた最初のステップとなっている。

## 2 データの概要

播磨町では現在、乳児健診、10か月児健診、1歳6か月児健診、3歳児健診の4つの乳幼児健診が実施されている。そして、それら4つの乳幼児健診のいずれかの健診で支援が必要だと判断された乳幼児は、乳幼児フォロー台帳に記載される。これら5つのデータと新生児訪問、出生児届出状況データを合わせた計7データ、過去5年分(2018年4月から2023年3月)が今回の分析対象データである。データの量については表1に示す。本研究では、分析対象データの一部である3歳児健診と乳幼児フォロー台帳データを使用して分析した。

次に、どのようなデータ項目が存在するのかについて記す。データ項目は大きく4つに分かれる。1つ目は基本情報であり、氏名や生年月日、母親宛名番号、受診日、月齢<sup>1</sup>、カウブ指数<sup>2</sup>等が含まれる。2つ目は発達検査の項目であり、健診時の年齢と標準発達とに差異がないかを確認するための項目からなっている。3つ目は医師の診察項目であり、医師が記入する。ここに含まれる項目には、特定の病気の疑いがないか等が記さ

表 1: データの概要

データ名	行(件数) × 列(データ項目)
新生児訪問	698 × 10
出生児届出状況	965 × 14
乳児健診	1223 × 151
10か月児健診	1229 × 136
1歳6か月児健診	1319 × 190
3歳児健診	1511 × 262
乳幼児フォロー台帳	585 × 23

<sup>1</sup>健診受診日の年齢を表しており、3歳児健診であれば3歳以上4歳未満の範囲内となっている。

2 体重と身長の関係から計算できる数値で、体重 (kg)  $\div$  身長 (m)<sup>2</sup> で求められる。主に肥満度を測るために使われる。

れている。4つ目は支援の判断についての項目であり、健診後に支援が必要な場合に記載され、記載された経緯も記入される。

最後に本研究での支援について定義する。本研究での支援は継続した多職種による支援のことを指し、乳幼児発達相談、親子相談、教室と療育、の3つに分けられる。乳幼児発達相談は集団健康診査、相談等で把握した発達上問題があると思われる乳幼児に対して、心と身体の総合的な発達相談と指導を行うものであり、医師、心理士、保健師による相談事業である。親子相談は心理士による子育て相談であり、心身に障がいを有する者又は心身の発達上何らかの心配のある者の生活上の問題について早期に相談支援を実施し、医療機関、療育、親子教室への紹介や生活全般のアドバイスを行う。教室は集団で実施する親子教室である。療育は個別で実施するもので言語療法(ST)、作業療法(OT)、理学療法(PT)、心理療育がある。以上を総称して本研究では支援と呼ぶ。

### 3 3歳児健診と乳幼児フォローアップ帳の分析

本章では、播磨町での3歳児健診と乳幼児フォローハンモックを掛け合わせたデータを分析する。目的は小児発達支援においてベテラン保健師の乳幼児健診における判断基準を明らかにすることである。分析対象は3歳児健診の受診者、1511名である。特に保健師が関わることの多い発達検査の項目に着目した分析となっている。

### 3.1 乳幼児フォローアップ帳について

本節では、乳幼児フォローランクについて述べる。乳幼児フォローランクとは、支援を行った乳幼児の支援内容や支援の開始日、終了日等を記入したものである。

### 3.2 3歳児の分類と分析

本節では、3歳児健診データを対象に、3歳児を分類して分析した。その結果を示す。なお、分析にはPythonを使用した。

### 3.2.1 前处理

事前に定めたルールに従って前処理した結果、1511件×279データ項目から1311件×50データ項目となった。内訳は前処理前は、支援ありが239名、支援なしが1272名、支援率は15.8%であり、前処理後は、支援ありが202名、支援なしが1109名、支援率は15.4%となった。前処理のルールを以下に示す。

1. 本研究では支援の時期や内容に関わらず支援を受けた3歳児を「支援あり」に統一して分析する。そのた

め、「支援あり」を1、「支援なし」を0とするフォロー列を新たに追加し、乳幼児フォローライフ帳に含まれる列はすべて削除する。: 1511 × 263

2. 本研究では保健師の判断に着目しているため、医師の診察による病気の疑いの有無が記載された列は本分析の対象外とし、削除する。: 1511 × 63

3. 欠損値<sup>3</sup>が100個未満のデータ列に欠損値をもつ3歳児についてはデータの信頼性を担保するため、分析対象外とする。欠損値が100未満の列に欠損値のある3歳児については行ごと削除する。: 1311 × 63

4. 欠損値が1000以上のデータ列についてはデータの信頼性を担保するため、そのデータ列を分析対象外とする。欠損値が1000以上の列は削除する。: 1311 × 50

5. 欠損値が100以上1000未満のデータ列について欠損値のある3歳児は行を削除してしまうと分析対象が大きく減ってしまうため、欠損値を補う。補完するセルが含まれる列は値の散らばり方が正規分布に従っているため、その列の中央値で補完する。: 1311 × 50

6. 「首のすわり、おすわり、人見知り、後追い、一人歩き、初語、2語文」についてはいずれも後天的な発達である。そのため、これらの列に0カ月目に発達がみられたと解釈される0が記入されることは不適切である。また、3歳児健診データは3歳児の間に受けた健診の結果であり、48カ月目以後に発達がみられたと解釈される48以上が記入されているセルも不適切である。不適切なセルは、そのセルが含まれる列の値の散らばり方がすべて正規分布に従っていたため、列の中央値で補完する。: 1311 × 50

7. 体重と身長の列で明らかな桁の間違いがみられる場合は、適切な値に修正する。また、これら2つのセルから計算して求めている肥満度のセルも併せて再計算し、修正する。: 1311 × 50

これら7つのルールに従って前処理を実行した。

### 3.2.2 クラスタリング

本研究では、クラスタリングを適用するデータが1311件と比較的多いため、非階層型クラスタリングであるk-means法を採用した。実行する際に使用した項目は3歳児健診の中で「肥満度、首のすわり、おすわり、人見知り、後追い、一人歩き、初語、2語文」の計8項目である。

「肥満度」の項目は(測定体重 - 標準体重) ÷ 標準体重 × 100で表され、数値が小さいと痩せぎみ、大きいと太りぎみである。その他の7項目はその発達がはじめて見られた月齢であり、数値が小さいと発達の早さが早め、大きいと発達の早さが遅めという意味をもつ。k-means法により、4分類した結果を表2に示す。

「肥満度」は、性別の身長体重曲線から係数を求めて計算した結果であり、-15~+15%が「ふつう」の範

表2: 3歳児の特徴別4分類(k-means法)

クラスター	人数 (人)	支援率 (%)	特徴
0	8	75.0	肥満度 小 首のすわり、おすわり、一人歩き 大
1	242	25.6	初語、2語文 大
2	695	11.5	標準的
3	366	14.8	人見知り 大
全体	1311	15.4	

囲である。この項目の分析対象データ平均は2.91%であり、「ふつう」の範囲内である。「首のすわり」は、支えなしに頭を安定させられる状態のことで月齢を表しており、一般的に生後3~4か月ごろが目安とされている。この項目の分析対象データ平均は3.29カ月であり、目安の範囲内である。「おすわり」は、支えなしに座った状態をキープできることで月齢を表しており、一般的に生後6~7カ月ごろが目安とされている。この項目の分析対象データ平均は6.73カ月であり、目安の範囲内である。「人見知り」は、見知らぬ人に対する恐怖や不安で、人の目を意識しすぎて恥ずかしがったり嫌ったりすることで月齢を表しており、一般的に生後6~12カ月ごろが目安とされている。この項目の分析対象データ平均は7.42カ月であり、目安の範囲内である。「後追い」は、保護者の姿が見えなくなると不安になって追いかけてくる行動で月齢を表しており、一般的に生後7~8か月ごろからはじまり、10~18カ月ごろがピークとされている。この項目の分析対象データ平均は8.73カ月であり、目安の範囲内である。「一人歩き」は、物につかまらずに自力で2、3歩交互に足を出せる状態になることで月齢を表しており、一般的に生後13カ月過ぎが目安とされている。この項目の分析対象データ平均は12.49カ月であり、目安の範囲内である。「初語」は、子どもが最初に発する意味のある1語分で月齢を表しており、一般的に生後12カ月前後が目安とされている。この項目の分析対象データ平均は13.20カ月であり、目安の範囲内である。「2語文」は、2つの単語からなる発語で生後14~24カ月ごろが目安とされている。この項目の分析対象データ平均は22.31カ月であり、目安の範囲内である。

4クラスターそれぞれの特徴について詳しく記す。クラスター0は「肥満度、首のすわり、おすわり、一人歩き」の平均はそれぞれ-2.64%, 7.38カ月, 14.75カ月, 26.38カ月であり、全体の平均である2.91%, 3.29カ月, 6.73カ月, 12.49カ月と比較すると、「肥満度」は小さく、「首のすわり、おすわり、一人歩き」は遅めであるため、痩せぎみで首のすわり、おすわり、一人歩きの発達が遅めのクラスターである。クラスター1は「初語、2語文」の平均がそれぞれ18.54カ月, 26.40カ月であり、全体の平均である13.20カ月, 22.31カ月と比較すると遅めであるため、初語、2語文が遅めのク

<sup>3</sup>データ内の変数において存在しない値のこと。

ラスターである。クラスター2はどの項目も平均付近であるため、標準的なクラスターである。クラスター3は「人見知り」の平均が9.43カ月であり、全体の平均である7.42カ月と比較すると遅めであるため、人見知りが遅めのクラスターである。

各クラスターに所属する児で、支援が必要と判断された者の割合を支援率とする。支援率は降順にクラスター0、クラスター1、クラスター3、クラスター2となっている。支援率が全体の平均を超えているのは、クラスター0とクラスター1の2つである。

### 3.2.3 主成分分析

「肥満度、首のすわり、おすわり、人見知り、後追い、一人歩き、初語、2語文」に対して標準化を行い、主成分分析を適用して、以上の8つの成分を5つの成分に次元削減した。第4主成分までで累積寄与率は0.71となった。結果を以下の表3にまとめる。

表3: 主成分分析(PCA)

	固有値	寄与率	累積寄与率
PC1	2.23	0.28	0.28
PC2	1.30	0.16	0.44
PC3	1.23	0.15	0.59
PC4	0.99	0.12	0.71

主成分の意味づけは、固有ベクトルの数値の大きさと正負の符号から推し量るため、各成分について固有ベクトルを求めて解釈した。各成分は以下のようにPC1は「総合」、PC2は「運動」、PC3は「言語」、PC4は「体質」と名付けた。この後の分析ではPC2とPC3を使用するため、この2成分を例として図1に示し、解釈の仕方を記す。

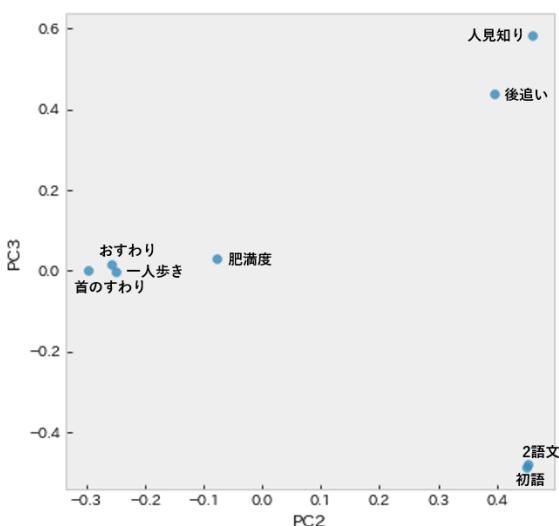


図1: 成分の解釈(PC2とPC3)

本分析では、固有ベクトルの負の値に着目した。PC2は「首のすわり、おすわり、一人歩き」が-0.30, -0.26, -0.25と特徴的であったので、「運動」と名付けた。PC3は「初語、2語文」が-0.49, -0.48と特徴的であったので、「言語」と名付けた。以上のように固有ベクトルの特徴的な項目を用いて、各成分を名付けた。

### 3.2.4 線形回帰

主成分分析で導いた「運動」成分であるPC2をx軸、「言語」成分であるPC3をy軸として分析対象の3歳児1311名を図2に散布図として図示した。オレンジの点が「支援あり」、青の点が「支援なし」となっている。併せて95%信頼区間で回帰直線も図示した。

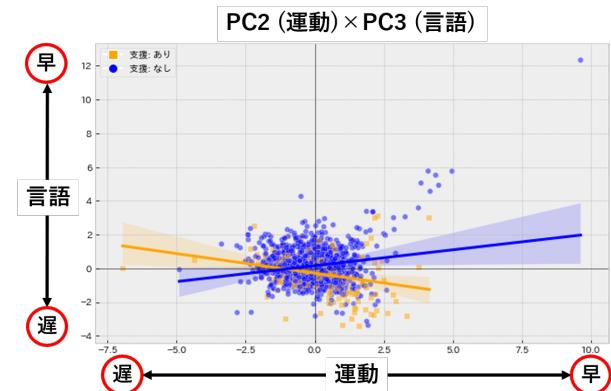


図2: 発達検査の重要項目発見(散布図: 運動×言語)

回帰分析の結果、支援あり3歳児の回帰直線は  
 $y = -0.23x - 0.29$   
 支援なし3歳児の回帰直線は  
 $y = 0.19x + 0.16$  となった。

支援あり3歳児の回帰直線は傾きが負、支援なし3歳児の回帰直線は傾きが正であった。このことから、運動と言語の発達の早さに差異のある3歳児が多く支援されていると読み取れる。

### 3.3 3歳児健診データを用いた神経発症児の疑いの判定について

本節では、神経発達症児<sup>4</sup>の疑いをベテラン保健師が健診データから判断する基準について分析する。

#### 3.3.1 3歳児健診の保護者の回答から判定した神経発達症児の疑い

3歳児健診では神経発達症児の疑いの有無に関わる保護者への質問項目が4つある。「じっとしていることが苦手」、「危険な行動がみられる」、「特にひどいくせがある」、「強いこだわりがある」の4つである。これら4つの質問項目にいくつ該当するのかに応じて分析対象の3歳児1311名を分類した。分類結果を表4に示す。

<sup>4</sup>神経発達症児は、知的障害(知的能力障害)、コミュニケーション障害、自閉スペクトラム症(ASD)、ADHD(注意欠如・多動症)、学習障害(限局性学習症、LD)、発達性協調運動障害、チック症の

表 4: 3歳児の4該当項目分類(累積割合(計100%))

該当項目数	支援あり(%)	支援なし(%)	差(%)
4つ	1.0(2名)	0.5(6名)	+0.5
3つ以上	14.9(30名)	3.6(40名)	+11.3
2つ以上	34.7(70名)	15.1(168名)	+19.6
1つ以上	61.4(124名)	42.7(473名)	+18.7
全体	100.0(202名)	100.0(1109名)	

該当項目数1つ以上から4つまでのすべてで「支援あり」が「支援なし」の該当割合を上回っている。以上のことから、保護者への4つの質問項目が支援の有無に影響を与える項目であると示された。

### 3.3.2 発達の早さと神経発達症の疑いとの関係

発達の早さに着目して神経発達症児の疑いを分析した結果を示す。前述の保護者への4つの質問項目のうち、該当項目数が2つ以上の3歳児を保護者の回答から神経発達症児の疑いありと判定された3歳児と定義する。また、保健師が3歳児健診の健診項目のうち、神経発達症児の疑いに関わる項目に要医療、要精密、要観察などと記入した3歳児を保健師が神経発達症児の疑いありと判定した3歳児と定義する。上述の定義に従い、3歳児健診受診者1311名をA、B、Cの3つに分類した。Aは保護者の回答から神経発達症児の疑いありと判定された3歳児であり、238名、全体の18.2%が属していた。Bは保健師が神経発達症児の疑いありと判定した3歳児であり、225名、全体の17.2%が属していた。Cは保護者の回答による判定と保健師の判定がともに神経発達症児の疑いなしとなった3歳児であり、941名、全体の71.8%が属していた。内訳を図3に示す。

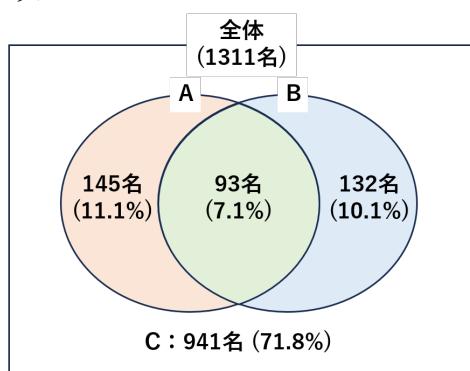


図3: 3歳児の内訳(1311名)

まず、 $A - (A \cap B)$ は145名、全体の11.1%となり、保護者の回答から神経発達症児の疑いありと判定されたが、保健師は神経発達症児の疑いなしと判定した3歳児なので、心配ぎみと名付けてオレンジ色で示した。 $A \cap B$ は93名、全体の7.1%となり、保護者の回答から神経発達症児の疑いありと判定され、保健師も神経発達症児の疑いありと判定した3歳児なので、該当

7つに分けられる。[7]

と名付けて緑色で示した。 $B - (A \cap B)$ は132名、全体の10.1%となり、保護者の回答からは神経発達症児の疑いなしと判定され、保健師は神経発達症児の疑いありと判定した3歳児なので、楽観的と名付けて青色で示した。

次に、これらの3歳児の分布の仕方を可視化した。この分析は標準的な発達の3歳児は支援なし3歳児の回帰直線上であると仮定している。そのため、支援なし3歳児の回帰直線の上側は言語の発達が運動の発達より相対的に早い領域でa、下側は言語の発達が運動の発達より相対的に遅い領域でbと名付けた。赤枠の1を $-1 \leq x < 1, -0.84 \leq y < 1.16$ 、緑枠の2を $-2 \leq x < -1, 1 \leq x < 2, -1.84 \leq y < -0.84, 1.16 \leq y < 2.16$ 、青枠の3をそれら以外とした。つまり、1から3の順に標準的な発達の早さからの遅れ度合いが大きくなっている。これらを使用して、1a、2a、3a、1b、2b、3bの6つの領域に分割した。分割結果を図4に示す。

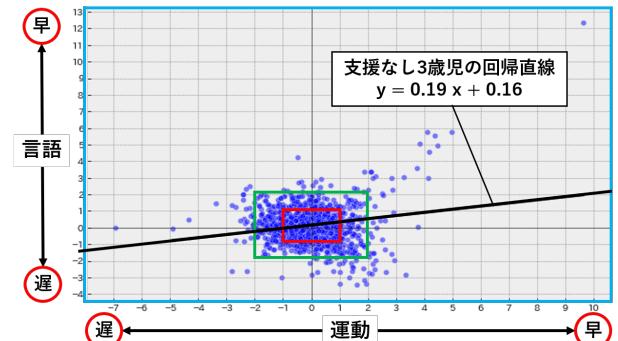


図4: 領域分析(運動×言語)

上記の領域を使用して3歳児を分類した結果を表5と表6に示す。()内は該当する人数を示している。

表5: 3歳児健診受診者の分析(運動  $\leq$  言語  $\rightarrow$  561人)

心配ぎみ	非該当	該当	楽観的
1a (29/262)	77.1% (202/262)	3.8% (10/262)	8.0% (21/262)
2a (24/238)	78.6% (187/238)	3.4% (8/238)	8.0% (19/238)
3a (8/61)	70.5% (43/61)	6.6% (4/61)	9.8% (6/61)
a (61/561)	77.0% (432/561)	3.9% (22/561)	8.2% (46/561)

表6: 3歳児健診受診者の分析(運動  $>$  言語  $\rightarrow$  750人)

心配ぎみ	非該当	該当	楽観的
1b (53/450)	72.9% (328/450)	7.3% (33/450)	8.0% (36/450)
2b (29/247)	65.2% (161/247)	8.5% (21/247)	14.6% (36/247)
3b (2/53)	37.7% (20/53)	32.1% (17/53)	26.4% (14/53)
b (84/750)	67.9% (509/750)	9.5% (71/750)	11.5% (86/750)

a と b の全体を比較すると該当では、3.9 %から 9.5 %に楽観的では、8.2 %から 11.5 %へとどちらも b が a を上回っていることが読み取れる。また、該当と楽観的の合計値は、1 から 3 の領域へと移動するに従って増えていく傾向にあった。心配ぎみと楽観的に着目すると、1b から 3b の心配ぎみを除いて 1 から 3 にかけて概ね割合が増えていく傾向が読み取れた。心配ぎみで最も割合が高かった領域は 3a で 13.1 %、楽観的で最も割合が高かった領域は 3b で 26.4 %であった。

## 4 結果と考察

まず、表 2 のクラスタリング結果を考察する。支援率が 75.0 %と最も高いクラスター 0 に着目する。このクラスターに属する 8 名は痩せぎみで首のすわり、おすわり、一人歩きの発達が遅めであることから、未熟児が支援されていると推察される。続いて、支援率が 25.6 %と 2 番目に高いクラスター 1 に着目する。このクラスターに属する 242 名は初語、2 語文といった言語の発達が遅めであることから、神経発達症児が支援されている可能性が高いと考えられる。

次に、図 2 から、運動と言語の発達の早さに差異のある 3 歳児が多く支援されていることから、発達の早さと支援の有無には関係があると分かった。

そして、表 4 に示す保護者の回答から、該当項目数 1 つ以上から 4 つまでのすべてで「支援あり」が「支援なし」の該当割合を上回っており、3 歳児の保護者の心配や考えが支援の有無に一定の影響を与えていていると考えられる。

最後に、発達の早さに着目した神経発達症児の疑いについて考察する。表 5 と表 6 から、a と b の全体を比較すると保健師が神経発達症児の疑いありと判定している該当と、楽観的のどちらも b が a を上回っていることから、言語発達の早さが運動発達の早さよりも 3 歳児の神経発達症児の疑いの判定について大きな影響を及ぼしているといえる。1 から 3 にかけて該当と楽観的は概ね増加傾向にあることから、標準発達からの遅れの程度も 3 歳児の神経発達症児の疑いの判定に影響を及ぼしているといえる。保護者の回答から神経発達症児の疑いがあると判定された 3 歳児である心配ぎみと該当の合計を a と b で比較すると 1 から 3 のすべての領域で b が a を上回った。東谷ら [8] は、発達障害児を持つ保護者を対象に行われた質問紙調査において、79.0 %の保護者が指摘や診断を受ける前に子どもの発達・行動を気にしていたこと、その理由として、同年代の子どもと比較して少し遅い・違うと思っていたためと答えた保護者が多かったことを報告している。この 2 つの事実から、保護者は言語発達の遅れから子どもの発達の遅れに気づき、子どもが神経発達症児であるのかもしれないという疑念を持つ場合があること

が示唆された。

本研究では、3 歳児健診では「未熟児」、「神経発達症児の疑い」という 2 つの観点が支援の決定に影響を与えることが示された。3 歳児健診では言語の発達が進んできているため、発達の早さによっては保護者が心配になったり、あるいは保護者の気づかない症状が見つかることもあると考えられる。今後は個人の特徴に合わせた適切なアプローチができるように、具体的な支援策も含めて分析していきたい。

## 5 おわりに

本研究では、「安心して子育てができる町」の実現のため、最初のステップとして小児発達支援においてベテランの保健師の乳幼児健診における判断基準を明らかにすることを試みた。乳幼児健診の実データを使用して 3 歳児の分類を行った点、支援される 3 歳児の特徴を発見した点について意義があると考える。特に神経発達症児の疑いについては、保護者と保健師の認識が異なる要因やその合わせ方も含めて、検討の余地があるだろう。

今後は、引き続きベテランの保健師の乳幼児健診における判断基準を明らかにすることに取り組むとともに、支援内容についても踏み込みたい。そして、それらを構造化して支援に必要な乳幼児に適切なタイミングと方法で支援を進めることができるような保健師をサポートするツールの作成を目指す。

## 参考文献

- [1] 厚生労働省, 令和 3 年度地域保健・健康増進事業報告の概況, 2023
- [2] 国立研究開発法人 国立成育医療研究センター, 乳幼児健康診査事業実践ガイド, pp87-193, 2018
- [3] 山内京香, 門田加奈, 乳幼児健診における発達支援が必要な子どもの保護者への対応, 藍野大学紀要, 31 卷, pp.55-61, 2019
- [4] 都筑千景, 援助の必要性を見極める, 日本看護科学会誌, 24 卷, 2 号, pp.3-12, 2004
- [5] 川端泰子, 千田みゆき, 行政で働く新任保健師の困難に関する文献検討, 埼玉医科大学看護学科紀要, 13 卷, 1 号, pp.41-47, 2020
- [6] 香西真由美, 石山美香, 朝倉理映, 森寿々子, 池内明子, 越田美穂, 産後・育児休暇中の先輩保健師の支援による新人保健師技術習得研修 先輩母子をモデルとした沐浴・乳房マッサージ実習の効果, 保健師ジャーナル, 70 卷, 5 号, pp.406-412, 2014
- [7] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, Text Revision Dsm-5-tr, Amer Psychiatric Pub Inc, 2022
- [8] 東谷敏子, 林隆, 木戸久美子, 発達障害児を持つ保護者のわが子の発達に対する認識についての検討, 小児保健研究, 69 卷, 1 号, pp.38-46, 2010

# 理学療法士間の知識共有に向けた 臨床推論テキストの ICF/ICD への分類

## Classifying Clinical Reasoning Texts Using ICF/ICD Frameworks for Facilitating Knowledge Sharing among Physiotherapists

櫟 力輔<sup>1</sup> \*  
Rikisuke Ichii<sup>1</sup>      松下 光範<sup>1</sup>  
Mitsunori Matsushita<sup>1</sup>      堀 寛史<sup>2</sup>  
Hirofumi Hori<sup>2</sup>

<sup>1</sup> 関西大学  
<sup>1</sup> Kansai University  
<sup>2</sup> 甲南女子大学  
<sup>2</sup> Konan Women's University

**Abstract:** 理学療法における臨床推論は、患者の問題を特定し管理について判断を下すための推論プロセスであり、重要な技能である。診療録などの分析や活用が臨床推論教育に有用だと考えられるが、記述内容には理学療法士の思考過程が反映される一方で、自由記述も含まれ、表現の差異が知識共有の障害となる。本研究の目的は、そのような非構造データに医療分野の標準的枠組みである国際生活機能分類 (ICF) や国際疾病分類 (ICD) を取り入れて構造化し、理学療法士の共通理解のもと計算機処理を可能にすることである。本稿ではその端緒として、模擬症例における推論文章の各文が ICF, ICD に該当するか判別する分類器を BERT を用いて作成した。

### 1 はじめに

理学療法は「身体に障害のある者に対し、主としてその基本的動作能力の回復を図るために、治療体操その他の運動を行なわせ、及び電気刺激、マッサージ、温熱などの物理的手段を加えることをいう」と定義されており<sup>1</sup>、理学療法士はそれら医学的なりハビリテーションを提供する専門職である。昨今の高齢化問題を背景に、社会保障の観点から健康寿命を延伸させる必要性が説かれている<sup>2</sup>。健康寿命とは「健康上の問題で日常生活が制限されることなく生活できる期間」を指し、疾病や障害による身体機能の回復や維持、予防を行うことで日常における活動や社会参加の維持・向上を支援する理学療法士は健康寿命の延伸に寄与する重要な役割を担っており、その育成が重要視されている。

理学療法士が治療立案をするにあたって重要な要素の一つに臨床推論が挙げられる。臨床推論は世界理学療法連盟によって「理学療法士やその他医療従事者が、データを収集・評価し、患者の問題特定や管理に関する判断を下すために用いる推論のプロセス」と定義されている<sup>3</sup>。効果的な臨床推論は患者の転帰を良好にするための不可欠なスキルであるとも言われており<sup>[5]</sup>、

臨床推論教育は理学療法教育において重要な位置を占めている。芳野ら<sup>[22]</sup>の研究では職員指導経験がある理学療法士 15 名に対して、指導している理学療法士がどのような能力を獲得したときに理学療法士として自立したと感じるかをインタビューした結果、臨床思考能力が 24.0% と最も多く記録されたと報告されている。しかし、2019 年に日本理学療法士協会が 3942 人の理学療法士を対象に行ったアンケート調査<sup>[17]</sup>によれば、「臨床推論を知っていますか?」という問い合わせて「知っていて実践している」と回答したのは 47% に留まり、36% が「知っているがあまり実践できていない」、16% が「知らない」と回答し、アンケートに回答した理学療法士の半数以上が普段から臨床推論を実践できていないことが明らかとなった。以上のことから、臨床推論が理学療法士にとって不可欠なスキルである一方で、その教育や実践には依然として課題が残っているといえる。臨床推論は、単純な知識の習得で身につくものではなく、患者の個別性や多様な状況に応じて変化する思考プロセスであり、その教育は容易では

<sup>1</sup>理学療法士及び作業療法士法: <https://laws.e-gov.go.jp/law/340AC0000000137> (2024/12/14 確認)。

<sup>2</sup>令和 4 年版厚生労働白書—社会保障を支える人材の確保—: <https://www.mhlw.go.jp/wp/hakusyo/kousei/21/d1/zentai.pdf> (2024/12/14 確認)

<sup>3</sup>World Physiotherapy: <https://world.physio/resources/glossary> (2024/12/14 確認)

\*連絡先：関西大学総合情報学部  
〒 569-1095 大阪府高槻市靈仙寺町 2-1-1  
E-mail: k244081@kansai-u.ac.jp

構成要素		第1部：生活機能と障害		第2部：背景因子	
領域	心身機能・ 身体構造	活動・参加	環境因子	個人因子	
構成概念	心身機能の変化 (生理的)	能力 標準的環境における課題の遂行	物的環境や社会的環境、人々の社会的な態度による環境の特徴がもつ促進的あるいは阻害的な影響力	個人的な特徴の影響力	
	身体構造の変化 (解剖学的)	実行状況 現在の環境における課題の遂行			
肯定的側面	機能的・構造的 統合性	活動 参加	促進因子	非該当	
	生活機能				
否定的側面	機能障害 (構造障害を含む)	活動制限 参加制約	阻害因子	非該当	
	障害				

図 1: ICF の概観 (文献 [16] より引用)

ない。特に、経験の浅い理学療法士や学生にとっては、患者情報の収集や解釈、治療計画の立案といった一連のプロセスにおける適切な考え方を学ぶことは困難であり、理学療法士育成の大きな課題となっている。

臨床推論教育を進める上では、診療録や症例レポートといった実際の記録を分析することが有用であると考えられている [20]。上記に挙げた文書には、患者の主訴、患者の現在の状態や推移、今後の治療方針、それらに対する理学療法士の思考（アセスメント）などが記録されており、これを活用することで、実際に行われている臨床推論の具体的なプロセスを教育に取り入れることが期待される。このときアセスメントにおける複雑な判断の背景などを記載する際には、自由記述形式が採用されていることが多い。しかし、同様の意味を表す記述であっても個人間での表現に差異が生まれ、理学療法士間での知識の共有が困難である。診療録や症例レポートなどを効果的に利用するためには、内容を統一的な形式に整理し、その上で共通の理解を得るための工夫が必要である。

本研究では、診療録や症例レポートに含まれる自由記述文のような非構造データを効率的かつ効果的に活用するために、国際生活機能分類（ICF）や国際疾病分類（ICD）といった国際的に医療分野で広く用いられている分類体系を利用し、これらのデータを構造化することを目指す。ICF や ICD に基づいて整理することが可能となれば、理学療法士間での共通理解が深まるだけでなく、そこに含まれる情報の分析や比較が容易になり、類似症例の検索や症例内容の概観などを迅速に行なうことが可能となる。これにより、教育現場では、理学療法士が症例を効率的に探索・参照することが期

待される。また、研究分野では、構造化データによる大規模解析や傾向分析を通して新たな知見の発見につながる。

本稿では、大腿骨頸部骨折の模擬症例を対象として記述された臨床推論テキストの各文が ICF の心身機能、身体構造、活動・参加、また ICD に該当するかを判別する分類器を BERT[7] を用いて作成した。

## 2 医療標準化

WHO（世界保健機関）が定める ICF 及び ICD は、医療の標準化を推進し、医療の質向上や効率的な組織運営、さらには病態や治療方針に関する説明の透明性向上など、公平な医療提供を実現するための基盤としても機能している。その意義は多岐にわたるが、特筆すべきは疾患や健康状態に関する統一的な診断・記録の枠組みを提供し、国際的な情報共有や比較を可能にする点である。これらの分類を活用した研究を進めることで疾患の実態把握が進むとともに、医療教育においても世界的に標準化された知識体系を構築することにつながる。また、WHO によると ICD と ICF は補完関係にあり、ICD は、疾病・障害・傷害などの健康状態の病因論的枠組みを提供する一方で、これらの健康状態に関連する機能や障害は ICF で分類されるとしており、ユーザにこの 2 つの分類を併用することを推奨している [13]。そのため、本研究においてもこれら 2 つを用いて、診療録や症例レポートに含まれている臨床推論に関する自由記述文の構造化を試みる。

### 2.1 国際生活機能分類（ICF）

国際生活機能分類（International Classification of Functioning, Disability and Health）とは、人間のあらゆる健康状態に関係した生活機能状態から、その人をとりまく社会制度や社会資源までをアルファベットと数字を組み合わせた方式で分類し、記述・表現をしようとするものである [16]。図 1 に示すように、ICF には 2 つの部門があり、それぞれは 2 つの構成要素からなる。以下に本稿で使用する「心身機能」、「身体構造」「活動」「参加」の 4 つの定義を記す。

「心身機能」とは、身体系の生理的機能（心理的機能を含む）であると定義されており、例えば、手足の動き、精神の働き、視覚・聴覚、内臓の働きが含まれる<sup>4</sup>。「身体構造」とは、器官・肢体とその構成部分などの、身体の解剖学的部分であると定義されており、手

<sup>4</sup>厚生労働省：ICF（国際生活機能分類）—「生きることの全体像」についての「共通言語」— <https://www.mhlw.go.jp/stf/shingi/2r985200002ksqi-att/2r985200002kswh.pdf> (2024/12/14 確認)

足の一部、心臓の一部（弁など）などの、体の部分のことを指す。「活動」とは、課題や行為の個人による遂行のことであると定義されており、実用歩行やその他のADL（日常生活行為）だけでなく、調理・掃除などの家事行為・職業上の行為・余暇活動（趣味やスポーツなど）に必要な行為・趣味・社会生活上必要な行為を指す。「参加」とは、生活・人生場面（life situation）への関わりのことであると定義されており、社会参加だけではなく、家庭内役割であるとか、働くこと、職場での役割、あるいは趣味にしても趣味の会に参加する、スポーツに参加する、地域組織のなかで役割を果す、文化的・政治的・宗教的などの集まりに参加する、などの広い範囲のものが含まれる。

ICFは分類コード、分離点、評価点で構成される。具体例として「s7402.181」を挙げる。分類コードは「s7402」にあたり、分離点はピリオドで表される。分離点以降の数字列「181」が評価点にあたる。構成要素それぞれは分類コード内のアルファベットに対応しており、心身機能は「b」、身体構造は「s」、活動・参加は「d」で表される。分類コード内の数字は構成要素の分類を示している。この分類は階層化されており、数字の位が階層の深さに対応している。具体例に示した「s7402」は「s：身体構造」、「7：運動に関連した構造」、「40：骨盤部の構造」、「2：筋肉」を表し、全てをまとめると「骨盤部の筋肉」を表すコードとなる。評価点が示すものはそれぞれの構成要素毎に異なる。主に機能障害の程度や性質、障害部位などを表すために用いられる。

## 2.2 国際疾病分類（ICD）

国際疾病分類（International Statistical Classification of Diseases and Related Health Problems）とは、正式には「疾病及び関連保健問題の国際統計分類」とい、異なる国や地域から、異なる時点で集計された死因や疾病のデータの記録、分析、比較を行うために国際的に統一した基準で設けられた分類であり[15]、ICD-11では疾病・障害・死因が27章構成で分類されている。ICDもICFと同様にアルファベットと数字を組み合わせることで表現され、コードによって疾病や障害部位、原因などを表すことができる。例えば、「S72.00」というコードは「大腿骨頸部骨折、閉鎖性」を表し、「S72.01」というコードは「大腿骨頸部骨折、開放性」を示す。この例では、S72が大腿骨頸部骨折を表し、その後の00・01で閉鎖骨折か開放骨折かを分類することができる。

## 3 関連研究

臨床推論の過程やその思考方法を明らかにしようとする試みやそれらを支援しようとする試みは盛んに行

われている。Wainwright[11]らは、初学者と経験豊富な理学療法士の間で、臨床意思決定能力とそのプロセスにどのような違いがあるかを明らかにし、臨床意思決定に影響を与える要因のモデルを構築すること目的とした。脳血管障害（CVA）患者を治療する初心者理学療法士と経験豊富な理学療法士1名ずつのペア3組を対象に、評価・治療の過程をビデオ録画及び半構造化インタビューを通じてデータを収集し、分析を行った。その結果、臨床意思決定に影響を与える要因は、情報的因子と指示的因子の2つに分類され、初学者は情報的要因に頼る傾向が強く、一方で経験豊富な理学療法士は指示的要因に依存する傾向が見られた。

また、以下のようにICFの枠組みを取り入れ支援を試みている例も存在する。Hageら[6]は筋骨格理学療法における臨床推論の学習を支援する新しいカードゲーム「Hypothesis Family Game」を提案している。ゲームを通して、患者の症状に関連する情報を分析し、仮説を構築する過程を楽しみながら学ぶことができ、学生の知識とスキルを統合的に高めることが期待されている、ここでは、ICFに基づいた仮説カテゴリーの枠組みがゲームの構造に組み込まれており、学生が患者の状態を分析する際に、ICFの視点を適切に反映させることを促すような仕組みになっている。

Atkinsonら[1]は、臨床推論スキルの向上を促すための手段として、振り返りと議論が重要な指摘されているものの、それらの重要なニーズを満たすためのツールが提供されていないことに目を向け、あらゆるスキルレベルの理学療法士が臨床推論スキルを育成するための振り返りと議論を深められるツールを開発した。ツールには「理学療法士実践ガイド」と「ICF」といった既存の概念的枠組みが統合されている。

上記に述べた研究では、様々な方法で臨床推論過程の解明や推論支援が行われているが、定性的な分析も多く計算機による分析を行った例は多くない。また、臨床推論に関する研究は小規模なデータを用いた例が多く、大規模データを分析することによって行われた例は少ない。このような状況の中で、近年の電子カルテの普及に伴い、紙媒体では困難であった大規模な医療情報の利活用が期待されており、医療分野における情報抽出技術を評価・洗練させる目的でMedical Records Track[10]などの共有タスクが国際的に取り組まれている。このような取り組みは医療上の意思決定支援を行うツールの開発を促進することにもつながる。また、Medical Records Trackが英語の医療文書を対象としたタスクであるのに対して、日本語を対象とした取り組みも行われており、情報アクセス技術向上のためのプロジェクトNTCIRのコアタスクの一つとしてMedNLPDoc[3]が取り組まれている。MedNLPDocのタスクに取り組んでいる具体例として、Sakishitaら[9]がルールベースによるICDコードの推定を行っている。この例では、

(1) 診断を含む文の選別, (2) 「ライフサイエンス辞書」を用いた英語医療用語の日本語変換, (3) アルツハイマー型認知症やアルツハイマー病といった表記ゆれへの対応, (4) 病名と部位の組み合わせを検出しコードを推定, (5) XML タグ情報を活用し, 特定のタグ推定という 5 つのルールベース手法を組み合わせたコード推定が行われており, MedNLPDoc タスクにおいて最も厳しい評価基準である「SURE」で最高の F 値を記録し, 他チームを上回る性能を示した。

臨床推論のプロセスや教育支援に関する様々な試みが行われ, ICF の枠組みや振り返り手法などが有用であることが示されてきた. 臨床推論教育を支援していく上では, これまでのようないンタビューなどを通した定性的な研究で詳細を把握していくと同時に, 大規模データを活用して統計的な知見を導き出すなど大きな視点での分析やそれに基づいた支援が必要となると考える. こうした状況を整えるためには, 医療文書の構造化や情報抽出を進め, 理学療法文書に計算機が適用可能な状況を整える必要がある. MedNLPDoc タスクのように日本語の医療文書においてもそのような取り組みは進められており, これらを理学療法分野に取り入れることで新たな知見の発見やそれらを取り入れた支援システムの開発につながると考える. そのため, 本研究では理学療法の推論テキストを ICF や ICD を取り入れて構造化し, 理学療法士の共通理解のもと計算機処理を可能にすることを目指す.

## 4 データセット

臨床推論テキストの各文に対して, ICF の「心身機能」, 「身体構造」, 「活動・参加」, ICD に該当する記述が含まれているかを 0,1 のラベルによってアノテーションを行うことでデータセットを作成した.

### 4.1 臨床推論テキスト

臨床推論テキストには, 吉田ら [21] 及び畠山ら [18] が行った先行研究で収集されたテキストを使用している. 以下にそれらの研究概要及びデータ取得について簡潔に述べる.

吉田らは理学療法初学者が必要な医療情報や理学療法検査の実測値を見落とさないことを目的として情報収集の段階で患者の重要な医療情報や理学療法検査の着目点を提示するシステムを構築した. システムの有用性を評価するために, 急性期病院に所属する 1 年目から 5 年目の理学療法士計 20 名に対し, 模擬症例(大腿骨頸部骨折症例)に対する理学療法プロセスの論理的文章記述と模擬症例の問題点を箇条書きで抽出する課題を与えた. 課題は 2 回与えられ, 1 回目は 20 名全

員が同様条件下で課題を行い, 2 回目は被験者を 10 名ずつに分け片群は吉田らの提案システム使用下で同一課題を行い, 片群はシステムを使用せずに同一課題を行った.

畠山らは理学療法分野のテキスト解析を促進するため, 理学療法分野のテキストにおけるアノテーション基準の構築とその妥当性を検証することを目的とした. 事前に収集した大腿骨頸部骨折術後の模擬症例に対する臨床推論文章 15 件に, アノテーションの付与作業を行い, 作成したアノテーション基準の妥当性を検証するため, 1~8 年目の理学療法士 5 名に, 作成したアノテーション基準を提示した上で, 実際にアノテーションの付与作業を行い, 付与されたアノテーションの一貫性を検証した.

吉田らの論理的文章記述課題において取得された文章計 40 件と畠山らが取得した臨床推論文章 15 件を組み合わせて本稿で取り扱う計 55 件の臨床推論テキストとした. 臨床推論テキストは全て大腿骨頸部骨折患者の模擬症例を基に作成されている.

### 4.2 アノテーション

臨床推論テキストを句点を基に 1 文に分け, 計 816 文をアノテーションの対象とした. 表 1 に示すように各文が ICF の心身機能 (b), 身体構造 (s), 活動・参加 (d) または ICD に該当する記述を含んでいない場合に 0 を含んでいる場合に 1 をそれぞれラベル付けした.

表 1 を用いて, ラベル付与について詳しく説明する. ID1 のテキストでは, 「圧痛」「伸長時痛」「収縮時痛」といった記述が心身機能にあたるため, b の項目に 1 が付与される. 加えて, 「右大腿近位外側部」といった記述が身体構造にあたるため, s の項目に 1 が付与される. その他の項目には該当する記述が存在しないため, 0 が付与される. ID2 のテキストでは, 「短距離であれば杖歩行自立」や「長距離歩行では歩行器を使用し自立レベル」といった記述が活動・参加にあたるため, d の項目に 1 が付与される. ID3 のテキストでは, 「歩行時の荷重痛」といった記述が心身機能にあたるため, b の項目に 1 が付与され, 加えて, 「歩行困難を来している」といった記述が活動・参加にあたるため, d の項目に 1 が付与される. ID4 のテキストでは, 「転倒」や「大腿骨頸部骨折」といった記述が ICD 項目にあたるため, 1 が付与される. ID1 や ID3 の様に記述内容によっては複数のラベルに 1 が付与される場合も存在する.

816 文のうち 1 が付与されたテキスト数は心身機能 (b) が 413 文 (50.6%), 身体構造 (s) が 274 文 (33.6%), 活動・参加 (d) が 186 文 (22.8%), ICD が 77 文 (9.4%) であった. ラベル付けデータ作成後に理学療法士 2 名の精査を通じて, 妥当性の担保を行った.

表 1: アノテーションの例

ID	臨床推論テキスト	b	s	d	ICD
		1	1	0	0
1	本症例では著名な疼痛は認めないものの、右大腿近位外側部に圧痛、伸長時痛、収縮時痛を認める。	0	0	1	0
2	現在術後 14 日目であり、トイレまでの短距離であれば杖歩行自立、長距離歩行では歩行器使用し自立レベルとなっている。	1	0	1	0
3	歩行について見てみると、まず主訴として歩行時の荷重痛を訴えており、それによる歩行困難を来していると考える。	0	0	0	1
4	本症例は転倒契機に大腿骨頸部骨折 (Garden IV) を受傷し、人工骨頭置換術 (後方アプローチ) を施行された症例である。				

表 2: 評価結果 (データ拡張前)

分類項目	正解率	適合率	再現率	F1-score
心身機能 (b)	0.867	0.837	0.923	0.875
身体構造 (s)	0.934	0.850	0.976	0.908
活動・参加 (d)	0.887	0.811	0.665	0.727
ICD	0.937	0.838	0.690	0.757

表 3: 評価結果 (データ拡張前)

評価指標	相関係数	p 値
正解率	-0.663	0.337
適合率	0.236	0.764
再現率	0.777	0.223
F1-score	0.739	0.261

## 5 評価・考察

入力された 1 文に ICF の心身機能 (b), 身体構造 (s), 活動・参加 (d), ICD に該当する記述が含まれているか否かをそれぞれ 2 値で判定する分類器を BERT によって作成し, 計 4 つの分類器を作成した. BERT の事前学習済モデルには東北大学乾研究室が公開している日本語 BERT モデル<sup>5</sup>を用いた. 事前学習モデルに対して, 4 章で構築したデータセットを用いて, ファインチューニングを行うことにより, 分類器を作成した. ファインチューニングにおけるバッチサイズは 8, エポック数は 1 とした. 学習及び評価は 4 章で構築したデータセットの規模を考慮し, 5 分割交差検証を行い, 4 つの分類器それぞれの正解率 (Accuracy), 適合率 (Precision), 再現率 (Recall), F1-score を算出した.

表 2 に評価結果を示す. 心身機能 (b) 並びに身体構造 (s) の分類器については心身機能の適合率を除く全ての項目で 0.85 以上を記録し, 良好な結果となった. 一方で, 活動・参加 (d), ICD の分類器については, 正

解率を除く全ての項目で 0.85 を下回る結果となり, 高い精度での分類には至らなかった. 分類精度低下につながった要因の一つとして, データセット内に含まれる該当記述数が少ないと考えられる. 4.2 節で示した通り, 分類精度が高かった心身機能 (b) や身体構造 (s) に関してはデータセット内に 1 のラベルが付与された該当記述数がそれぞれ心身機能: 413 文, 身体構造: 274 文と多く存在する. しかし, 分類精度が低かった活動・参加 (d), ICD に関しては活動・参加 (d) : 186 文, ICD : 77 文と分類精度が高かったものと比べると該当記述数が少ない傾向が見られた. そのため, それぞれの分類器作成における該当記述数を独立変数, 各評価指標を従属変数として相関係数を算出した. 結果を表 3 に示す. 再現率・F1-score に関しては, p 値が低く統計的有意性は確認できなかったものの, 中程度の正の相関が見られ, 評価結果を増やすことで分類精度が改善する可能性が示唆された.

## 6 データ拡張による分類精度向上

評価結果を踏まえた考察からデータセットの規模を拡大することで分類精度が向上する可能性が示唆された. そのため, 本章では既存のデータセットに対して, データ拡張を実施することで分類精度向上を図ることとした.

### 6.1 データ拡張 (Data Augmentation)

データ拡張 (Data Augmentation) は, 機械学習においてモデルの性能向上を目的として, 既存のデータセットを加工・変換することで新たな学習データを生成することを指し, 自然言語処理のみならず, 画像認識や音声認識などの多様な分野で活用されている. 特に収集可能なデータが限られている状況においてこの手法は効果的であり, 小規模なデータセットであっても追加のデータを作成することで, モデルの過学習を防ぎ, 汎化性能を向上させることが可能である. 本稿

<sup>5</sup>tohoku-nlp/bert-base-japanese-v2: <https://huggingface.co/tohoku-nlp/bert-base-japanese-v2> (2024/12/14 確認)

表 4: データ拡張により生成されたテキストの例

	テキスト例
原文	また、FRT が 15cm とカットオフ値を下回っており股関節伸展制限の影響を受けている。
同義語置換後	また、FRT が 15cm とカットオフ値を下回っており股関節伸展制限の影響を聴き入れる御出でなされる。
ChatGPT 修正後	また、FRT(Functional Reach Test) が 15cm でカットオフ値を下回っており、股関節伸展制限の影響が認められます。

で扱っているような医療ドメインのテキストデータはプライバシー保護などの観点から入手が困難なこともあります。データ量が不足してしまうことが今回のようにモデル性能の低下を招く大きな要因となることもある。また、医療などの専門性が高いドメインに対するアノテーションは多大な労力を要し、加えて、その妥当性を担保することも必要となる。データ拡張によって生成されたデータには拡張前と同じラベルを付与することが可能なため、こうした観点からもデータ拡張を行う意義は大きいと考える。

## 6.2 手法

本節ではデータ拡張に用いた具体的な手法を述べる。自然言語処理分野におけるデータ拡張手法として Wei ら [12] による EDA (Easy Data Augmentation) が提案されている。この手法では、同義語置換 (Synonym Replacement), ランダムな単語の挿入 (Random Insertion), ランダムな単語の入れ替え (Random Swap), ランダムな単語の削除 (Random Deletion) の4つを組み合わせて拡張テキストを生成する。本稿では、EDA 内の同義語置換と ChatGPT を用いたテキスト修正を行うことでデータ拡張を実施した。

まず、同義語置換について述べる。EDA における同義語置換は文からストップワード以外の n 個の単語をランダムに選び、これらの単語をランダムに選ばれた同義語で置き換えるというものである。本稿では、MeCab<sup>6</sup> によりテキストを形態素に分割した後、置き換える単語数 n を 2 に設定し、日本語 WordNet[4] により同義語を検索・置換した。ストップワードには、ライブラリ SlothLib[14] に搭載されているストップワードリストを用いた。この際、テキストを特徴づける医療専門用語が置き換わることを防止するために、置き換える単語の品詞を動詞に限定した。生成されたテキストの具体例を表 4 の同義語置換後に示す。原文の「受け」の部分が「聴き入れる」に置換され、「いる」の部分が「御出でなされる」に置換されていることが見て取れる。一般に用いられる表現とはかけ離れた表現のテキストが同義語置換により生成されたため、診療録などで用い

られていても遜色のないテキストに修正する必要があると考えた。そのため、同義語置換の工程を経た後に、ChatGPT によってテキストを修正した。GPT のバージョンには GPT-4 Omni を用いた。プロンプトには、「以下のテキストを正しい日本語に修正してください。」と入力し、そのプロンプトの下部に同義語置換後のテキストを記述した。生成されたテキストの具体例を表 4 の ChatGPT 修正後に示す。修正後は「聴き入れる」や「御出でなされる」といった表現が「認められます」といったものに変化し、一般に用いられる表現に近づいたことが見て取れる。また、FRT という略称の記述に対して「Functional Reach Test」などの正式名称が補足されるといった副次的な効果も見てとれた。

上記に述べた同義語置換並びに ChatGPT による修正の工程を生成元のテキスト 1 文につき 1 度実施し、新たに 816 文を生成した。新たに生成された 816 文の意味が生成元のテキストと大きく異なることを理学療法士 2 名に確認してもらった上で、生成元のテキストと同じラベルを付与した。ラベルを付与した 816 文を新たにデータセットに加えることでデータ拡張を実施し、データ数を 2 倍の 1632 文とした。

## 6.3 データ拡張後の評価・考察

拡張したデータセットを用いて、新たに 4 つの分類器を作成し、拡張前と同様に 5 分割交差検証を行った。このとき、学習パラメータなどはデータ拡張前と同様である。評価結果を表 5 に示す。全 16 項目のうち、12 項目で精度の向上が確認された。最も大きく上がったものは、活動・参加における再現率で 0.665 から 0.759 となり、0.094 (14.1%) の精度向上が認められた。加えて、活動・参加の F1-score や ICD の再現率に関しても、およそ 0.05 (活動参加/F1-score: 8.1%, ICD/再現率: 7.3%) の精度向上が認められ、前回の考察の通り、該当記述数が少ないものに対して、データ拡張によってデータ量を補うことで一定程度の精度向上が可能であった。一方で、最も精度が低下したのは身体構造の再現率で 0.976 から 0.942 となり、0.034 (3.4%) の精度低下が認められた。

<sup>6</sup><https://taku910.github.io/mecab/> (2024/12/14 確認)

表 5: 評価結果 (データ拡張後)

分類項目	正解率	適合率	再現率	F1-score
心身機能 (b)	0.875	0.840	0.934	0.882
身体構造 (s)	0.927	0.859	0.942	0.897
活動・参加 (d)	0.906	0.818	0.759	0.786
ICD	0.962	0.821	0.741	0.779

## 7 課題・展望

本稿で作成した分類器によって臨床推論テキスト内に ICF の心身機能 (b), 身体構造 (s), 活動・参加 (d) ICD に該当する記述が含まれているかを判別する事が可能となった。しかしながら、今回の分類では 2 章に示した ICF や ICD の非常に浅い階層への分類を達成したのみであり、より深い階層への分類や評価点の推定には至っていない。ICF のコードは評価点があつてはじめて完全なものとなる [16], とも述べられており、臨床推論テキストデータを活用していくためにはより深い階層のコード特定や評価点推定を実現する必要がある。その際、深い階層へ進むにつれ分類項目が増加するため、今回のようにそれぞれの項目に対して 2 値分類手法を用いるとモデル数が膨大になるという問題が生じる。そのため、今後のモデル構築にはマルチラベル分類手法などの導入を検討する必要がある。また、評価値推定に関しては、テキスト内に含まれる検査内容や検査結果、加えてそれらの経時的変化などを捉えることで算出が可能であると考えられるが、現状の 1 文単位という処理単位の粒度はそれに適さない可能性もある。これに対しては、Ando ら [2] や宮本ら [19] が行っているように取り扱うテキストにおける構造化の最適な粒度を設定し、必要に応じてより詳細な処理を実現する必要があると考える。また、本稿で取り扱った分類器や今後のモデル構築においては、特定ドメインのテキストで事前学習されたモデルを用いることで分類精度が向上する可能性がある。日本語の医療テキストに関するものでは、Kawazoe ら [8] によって UTH-BERT モデルが公開されており、このようなモデルを用いて推定を行うことで更なる精度向上を図ることができる。

本研究では、テキスト内に含まれる情報を構造化することで臨床推論への活用を試みている。しかし、判断に使用している非常に細かな情報や個人が持つ独自の解釈などは理学療法士によっては記載しない可能性がある。そのため、実際に理学療法士が診療録などに記載しない情報をインタビューなどを通して明らかにし、そのような暗黙的に用いられる知見を引き揚げる仕組み作りが必要であると考える。

## 8 おわりに

本研究では、診療録や症例レポートに含まれる自由記述文のような非構造データを効率的かつ効果的に活用するために、国際生活機能分類 (ICF) や国際疾病分類 (ICD) といった医療分野で広く用いられている分類体系を利用し、これらのデータを構造化することで理学療法士の共通理解のもと計算機処理を可能にすることを目指す。本稿では、その端緒として、臨床推論テキスト 55 件の 816 文を対象として、1 文に ICF の心身機能 (b), 身体構造 (b), 活動・参加 (d), ICD に該当する記述が含まれているか否かをそれぞれ 2 値で判定する分類器を BERT によって作成し、計 4 つの分類器を作成した。5 分割交差検証による分類精度評価の結果、それぞれの F1-score は心身機能 (b) : 0.875, 身体構造 (s) : 0.908, 活動・参加 (d) : 0.727, ICD: 0.757 となった。その後、分類器の精度向上を企図して、同義語置換と ChatGPT によるテキスト修正を組み合わせたデータ拡張を行い、データ量を倍増した。データ拡張実施後、5 分割交差検証による再評価を行い、評価項目全 16 項目のうち、12 項目で精度向上を達成した。今後はより深い階層への分類や評価点推定を目指す。

## 謝辞

本研究の実施にあたり、吉田龍洋氏（岸和田徳洲会病院）、畠山駿弥氏（兵庫県立尼崎総合医療センター）の協力を得た。記して謝意を表す。

## 参考文献

- [1] Atkinson, H. L., Nixon-Cave, K.: A tool for clinical reasoning and reflection using the international classification of functioning, disability and health (ICF) framework and patient management model, *Physical Therapy*, Vol. 91, No. 3, pp. 416–430 (2011)
- [2] Ando, K., Okumura, T., Komachi, M., Horiguchi, H., Matsumoto, Y.: Exploring optimal granularity for extractive summarization of unstructured health records: Analysis of the largest multi-institutional archive of health records in Japan, *PLOS Digital Health*, Vol. 1, No. 9, pp. 1–19 (2022)
- [3] Aramaki, E., Kano, Y., Ohkuma, T., Morita, M.: MedNLPDoc: Japanese Shared Task for Clinical NLP, *Proceedings of the Clinical Natural Language Processing Workshop*, pp. 13–16 (2016)

- [4] Bond, F., Baldwin, T., Fothergill, R., Uchimoto, K.: Japanese SemCor: A Sense-tagged Corpus of Japanese, *Proceedings of the 6th global WordNet conference*, pp. 56–63 (2012)
- [5] Elvén, M., Boersma, K., Overmeer, T.: Assessing clinical reasoning in physical therapy: discriminative validity of the Reasoning 4 Change instrument, *Physiotherapy*, Vol. 117, pp. 8–15 (2022)
- [6] Hage, R., Fourré, A., Ramonfosse, L., Leteneur, S., Jones, M., Dierick, F.: Description and rules of a new card game to learn clinical reasoning in musculoskeletal physiotherapy, *Journal of Manual Manipulative Therapy*, Vol. 31, No. 4, pp. 287–296 (2023)
- [7] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp. 4171–4186 (2019)
- [8] Kawazoe, Y., Shibata, D., Shinohara, E., Aramaki, E., Ohe, K.: A clinical specific BERT developed using a huge Japanese clinical text corpus, *PLOS ONE*, Vol. 16, No. 11, pp. 1–11 (2021)
- [9] Sakishita, M., Kano, Y.: Inference of ICD Codes by Rule-Based Method from Medical Record in NTCIR-12 MedNLPDoc, *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies*, pp. 80–84 (2016)
- [10] Voorhees, E. M., Hersh, W.: Overview of the TREC 2012 Medical Records Track, *21st Text REtrieval Conference, TREC 2012* (2013)
- [11] Wainwright, S. F., Shepard, F. K., Harman, B. L., Stephens J.: Factors that influence the clinical decision making of novice and experienced physical therapists, *Physical Therapy*, Vol. 91, No. 1, pp. 87–101 (2011)
- [12] Wei, J., Zou, K.: EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 6382–6388 (2019)
- [13] World Health Organization: How to Use the ICF: A Practical Manual for Using the International Classification of Functioning, Disability and Health (ICF) : Exposure Draft for Comment, *WHO*, (2013)
- [14] 大島裕明, 中村聰史, 田中克己: Slothlib: web サーチ研究のためのプログラミングライブラリ, *DBSJ letters*, Vol. 6, No. 1, pp. 113–116 (2007)
- [15] 厚生労働省大臣官房統計情報部編: 疾病、傷害および死因統計分類提要 ICD-10 (2003 年版) 準拠 第 1 卷 総論, 財団法人厚生統計協会 (2006)
- [16] 障害者福祉研究会: 國際生活機能分類 (ICF) —国際障害分類改定版—, 中央法規出版, (2002)
- [17] 日本理学療法士協会: 日本理学療法士協会ニュース (JPTA) , Vol. 323 (2020)
- [18] 畠山 駿弥, 中村 圭介, 松下 光範: 理学療法分野におけるテキスト解析のためのアノテーション基準構築と妥当性検証, 第 35 回兵庫県理学療法学術大会, No. P-022, pp. 40 (2024)
- [19] 宮本 誠人, 松下 光範, 高岡 良行, 堀 寛史: 理学療法初学者の支援を目的とした動作分析テキストの構造の可視化, 2022 年度人工知能学会全国大会論文集, 1I1-OS-6-04 (2022)
- [20] 村瀬 雅敏: 症例報告の重要性を考える, 理学療法の歩み, Vol. 16, No. 1, pp. 5–13 (2005)
- [21] 吉田 龍洋, 畠山 駿弥, 杉本 明文, 堀 寛史, 佐々木 恭志郎, 高岡 良行, 松下 光範: 医療情報における着目点提示による理学療法初学者の臨床推論支援, 電子情報通信学会 HCG シンポジウム 2023 論文集, No. B-2-2 (2023)
- [22] 芳野 純, 二渡 玉江, 大谷 健, 白田 滋: 自立した理学療法士が獲得すべき能力に関する質的研究, 理学療法学, Vol. 37, No. 6, pp. 410–416 (2010)

# 聴解テストから得られたラベルセットによるアニメシーン会話の属性のゼロショット分類の検証

## Verifying Anime Scene Dialogues' Attribute Zero-shot Classification via Extracted Label Set in Listening Test

Yangdi NI<sup>1\*</sup> Junjie SHAN<sup>2</sup> 西原 陽子<sup>2</sup>  
Yangdi NI<sup>1</sup> Junjie SHAN<sup>2</sup> Yoko NISHIHARA<sup>2</sup>

<sup>1</sup> 立命館大学大学院情報理工学研究科

<sup>1</sup> Graduate School of Information Science and Engineering, Ritsumeikan University

<sup>2</sup> 立命館大学情報理工学部

<sup>2</sup> College of Information Science and Engineering, Ritsumeikan University

**Abstract:** 従来のテキスト分類手法では各分類タスクにより、分類モデルを訓練する必要がある。そこで本研究では、ゼロショット分類モデルの分類ラベルを調整することにより、分類精度を向上する手法を提案する。日本語の聴解テストに付与した分類ラベルと日本アニメの会話シーンに提案手法を用いて付与されたラベルを用いて属性分類を行い、分類精度を検証した。会話相手の人間関係、会話の場所、会話の形式の三つの属性を設け、聴解テストの頻出パターンに限定し、アニメシーンでの分類精度を属性ごとに評価した。結果としては、3種類の会話属性の中に1種類以上が正解である割合は88.5%となった。

### 1 はじめに

異なるデータセット中のテキストを属性ごとにマッチングするタスクにおいて、機械学習の分類の方法を用いることができる。テキストに属性を示すラベルを付与して学習データを作成し、学習データを用いて分類モデルを作成することで、テキストの自動分類とマッチングが行える。一般的にテキストにラベルを付与する作業は時間がかかる。なお、ラベルは分類タスクが変化するごとに付け直す必要がある。さらには分類モデルの再作成にも時間がかかる。

本研究では、学習データへのラベリングと分類モデルの再作成の作業負担を減らすため、ゼロショット分類を用いたテキストのマッチング手法を提案する。ゼロショット分類では、入力するテキストが同じであってもラベルセットが変わると分類結果も変化する。そこで本研究では、ゼロショット分類に用いるラベルセットとして分類精度の高いものを抽出し、抽出されたラベルセットを用いてテキストのマッチングを行う。

本論文では、1つのデータセットから得られたラベルセットを用いて、もう1つのデータセットの属性の分類が行えるかを確認し、マッチングの精度評価は今

後の課題とする。

### 2 関連研究

本研究に関する既存研究を示し、本研究の位置付けを行う。

#### 2.1 属性を用いたテキスト分類

テキスト分類を扱った研究は多く、それぞれの研究において分類の属性が用意されている。Hanazono ら [1] の研究は、テキストに感情を表す属性ラベルを付与し、テキストを分類した。Elberrichi [2] らの研究では、単語間の関係を分析し、構文属性と文法属性により、テキストを分類した。また、対話文のテキストを分類するため、会話の形式、会話の場所、会話の活動、話者間の関係性など8種類のメタ属性を取り入れた研究もある [3]。また、Pazzani ら [4] の研究では、ユーザが投稿した商品説明に記載された属性を分類することで、ユーザの興味に合致するアイテムをピンポイントで特定する。テキストの内容からテキストの場所や人物間の関係を自動的に属性で分類する研究もある [5, 6]。

既存研究ではテキストを分類するため、手作業でラベルを付与し、学習データを用意することが多い。本

\*連絡先：立命館大学情報理工学研究科  
〒 567-8570 大阪府茨木市岩倉町 2-150

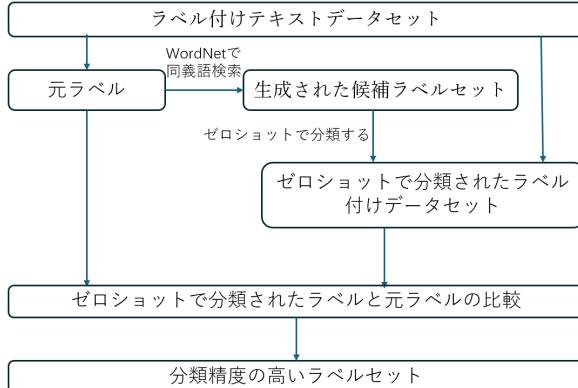


図 1: 分類精度の高いラベルセットの抽出の流れ.

研究では、ゼロショット分類を用いて異なるテキストの属性のマッチングを行う手法を提案する。

## 2.2 ゼロショット分類

ゼロショット分類を用いた様々な研究が提案されている。例えば、銀行が公表する文書をゼロショットで分類することにより、銀行の政策スタンスを評価する手法 [7] や、声質変換をゼロショットにより行う手法 [8] などがある。

既存研究でもゼロショット分類の精度の改善が行われているが、ラベルセットの調整を試みたものは少ない。本研究では、分類精度の高いラベルセットを自動的に抽出することでゼロショット分類を行う手法を提案する。

## 3 提案手法

図 1 と図 2 に提案手法の流れを示す。提案手法は主に 2 つの部分で構成され、1 つ目は属性分類の精度の高いラベルセットを特定する部分であり、2 つ目は分類された属性によりテキストをマッチングする部分である。

図 1 のラベルセットの抽出では、先にラベル付けテキストデータセットを用意する。そして、ラベル付けテキストデータセットの元ラベルを WordNet を利用し、同義語を探し、異なる候補ラベルセットを生成する。生成されたラベルセットでラベル付けテキストデータセットをゼロショットで分類し、分類の結果と元ラベルを比較し、分類精度の高いラベルセットを抽出する。

図 2 の属性マッチングでは、抽出されたラベルセットでラベルなしのデータセットとラベル付けデータセットをゼロショットで分類し、同じ属性ラベルを持つデータ同士をマッチングする。

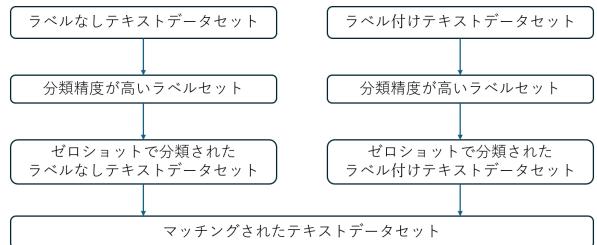


図 2: 属性マッチングの流れ.

### 3.1 ゼロショット分類のためのラベルセットの抽出

分類精度の高いラベルセットを抽出するため、先に WordNet で大量のラベルセットを生成する。そして、それぞれ生成されたラベルセットでテキスト属性を分類し、正解ラベルと比べて、分類精度を算出する。

本研究は WordNet を利用し、(1) 同義語置換、(2) 反義語と組合せ、(3) フレーズを追加の 3 つの手法でラベルセットを生成する。図 3 は (1) 同義語置換の手法を示している。ラベルセット「雑談、相談」の中のラベル「雑談」に対し「ゴシップ」が同義語として示され、もう一方のラベル「相談」に対し「協議」が同義語として示されている。これらの同義語を組合せることにより、ラベルセットの候補「ゴシップ、協議」を生成する。図 4 は (2) 反義語と組合せの手法を示している(2 分類の場合)。ラベル「雑談」に対し「ゴシップ」が同義語として示され、「ではない」を追加し、「ゴシップ、ゴシップではない」を生成する。図 5 は (3) フレーズを追加の手法を示している。ラベル「雑談」に対し「ゴシップ」が同義語として示され、もう一方のラベル「相談」に対し「協議」が同義語として示されている。フレーズを追加し、「ゴシップ+フレーズ、協議+フレーズ」を生成する。図 5 ではフレーズの例として「の会話」が使われており、生成されたラベルは「ゴシップの会話、協議の会話」となる。

生成されたラベルセットでラベル付けのデータセットをゼロショットで分類し、分類結果と元付けたラベルを比較し、ラベルセットの分類精度を評価する。そして、分類精度の高いラベルセットを抽出する。

### 3.2 属性マッチング

抽出されたラベルセットを用いて、2 つの異なるデータセット中のテキストをゼロショットで分類する。分類された属性ラベルをテキストに付与し、属性の相違によりテキスト間でマッチングを行う。

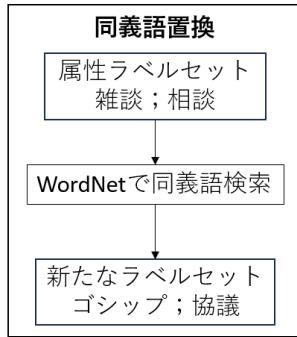


図 3: WordNet でラベルセットの生成（同義語置換）.

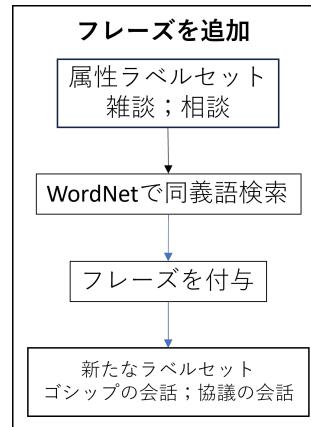


図 5: WordNet でラベルセットの生成（フレーズを追加）.

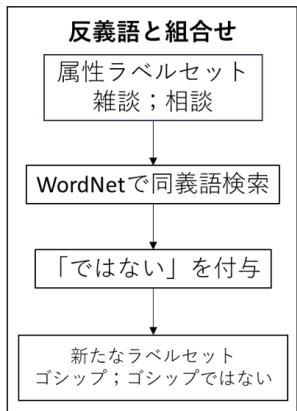


図 4: WordNet でラベルセットの生成（反義語と組み合わせ）.

## 4 予備実験

属性分類に用いるラベルセットを抽出するための予備実験を行った。予備実験と評価実験では、データセットとして日本語能力検定試験の聴解テストの会話テキスト集合と、日本のアニメ会話シーンのテキスト集合を用いた。これは、既存研究において聴解テストと属性が一致するアニメ会話シーンを提供する妥当性が確認されているためである [10].

### 4.1 実験手順

実験は以下の手順に従って行った。

1. 聴解テストの属性ラベルを手作業で付与する。
2. WordNet を利用し、候補ラベルセットを生成する。
3. 候補ラベルセットを用いて聴解テストをゼロショットで分類する。
4. 得られた分類結果と手作業の結果を比較し、分類精度の高いラベルセットを抽出する。

会話テキストに付与する属性は、既存研究 [3] を参考し、「場所」「形式」「関係」の 3 種類とした。それぞれに初期ラベルセットとして、

- **場所 (6 分類)** : 自宅, 職場, 学校, 施設, 屋外, 電話
- **形式 (2 分類)** : 雜談, 用談相談
- **関係 (4 分類)** : 初対面, 知り合い, 友達, 家族

を手作業で付与した。

本研究は 2 つのゼロショット分類方法を利用した。1 つ目はテキストを Embedding してベクトルを作成し、ベクトルのコサイン類似度を評価することで分類する手法である（図 6）。言語モデルを用いてテキストとラベルを Embedding してベクトルを作成し、ベクトルのコサイン類似度を評価して、類似度が高いラベルをテキストの分類結果とする。実験は Open-AI の text-embedding-ada-002 モデル<sup>1</sup>を用いた。

2 つ目はテキストとラベルセットを与えて、事前訓練済み言語モデルを用いてゼロショット分類を行う手法である（図 7）。テキストとラベルセットを事前訓練済み言語モデルに与え、各ラベルの尤度を算出し、尤度が最も高いラベルをテキストの分類結果とする。実験は MoritzLaurer モデル<sup>2</sup>を用いた。

2 つのゼロショット分類方法ごとに分類精度の高いラベルセットを評価した。ラベルセットの分類精度を評価するため、既存研究 [11] を参照し、分類結果の平均二乗平方誤差 (RMSE) を用いて評価を行った。

予備実験に用いた聴解テストの内容を説明する。実験には N1 から N5 まで各レベル 50 件合計 250 件聴解テストを利用し、「場所」、「形式」、「関係」3 種類の属性

<sup>1</sup><https://platform.openai.com/docs/guides/embeddings>

<sup>2</sup><https://huggingface.co/MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil>

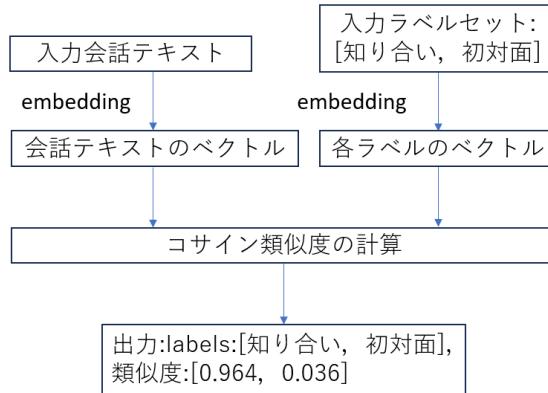


図 6: テキストと分類ラベルの Embedding の類似度によりゼロショット分類を行う例.

表 1: 各ゼロショット分類の属性ごとの RMSE が一番低いラベルセット

ゼロショット分類方法	属性	分類精度の高いラベルセット	RMSE
MoritzLaurer	場所 6 分類	[自宅で話す, 職場で話す, 学校で話す, 施設で話す, 街並みで話す, 電話で話す]	1.75
	形式 2 分類	[世間話, 相談]	3.43
	関係 4 分類	[初対面間の会話, 御近付き間の会話, 身方間の会話, 親類間の会話]	2.39
	Open-AI の embedding	[自宅で話す, 職場で話す, 学校で話す, 施設で話す, 街並みで話す, 電話で話す]	2.54
Open-AI の embedding	形式 2 分類	[雑談, 相談]	3.17
	関係 4 分類	[初対面, 知遇, 駄じみ, 親類]	2.74

のラベルを手作業で付けた. そして, WordNet を利用し, 162 個の候補ラベルセットを生成した.

## 4.2 実験結果

RMSE が最も低いラベルセットを表 1 に示す. 2 つのゼロショット分類方法で異なるラベルセットが得られた. MoritzLaurer モデルで抽出したラベルセットの RMSE の平均値は 2.54, Open-AI の embedding モデルでの場合の平均値は 2.81 であった. 平均値より, MoritzLaurer モデルの方が分類精度の高いラベルセットが抽出されたことがわかった.

一方で, 「形式」の分類においては, Open-AI の embedding モデルを用いた方が RMSE の値が低いラベルセットが抽出された. そこで, 次章の評価実験においては表 1 のラベルセットのうち, (1) MoritzLaurer で得られたラベルセットを用いた分類と (2) MoritzLaurer で得られた「場所」「関係」のラベルセットと Open-AI の embedding モデルで得られた「形式」のラベルセットを用いた分類の 2 種類の評価を行う (表 2).

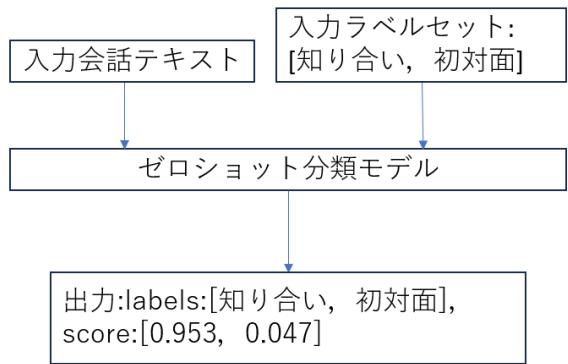


図 7: 事前訓練モデルによりゼロショット分類を行う例.

表 2: 評価実験で用いるゼロショット分類方法とラベルセット

ゼロショット分類方法	属性	分類精度の高いラベルセット
MoritzLaurer	場所 6 分類 (MoritzLaurer)	[自宅で話す, 職場で話す, 学校で話す, 施設で話す, 街並みで話す, 電話で話す]
	形式 2 分類 (MoritzLaurer)	[世間話, 相談]
	関係 4 分類 (MoritzLaurer)	[初対面間の会話, 御近付き間の会話, 身方間の会話, 親類間の会話]
Open-AI の embedding と MoritzLaurer	場所 6 分類 (MoritzLaurer)	[自宅で話す, 職場で話す, 学校で話す, 施設で話す, 街並みで話す, 電話で話す]
	形式 2 分類 (Open-AI の embedding)	[雑談, 相談]
	関係 4 分類 (MoritzLaurer)	[初対面間の会話, 御近付き間の会話, 身方間の会話, 親類間の会話]

## 5 評価実験：アニメーションの会話の属性分類の評価

評価実験では, 聴解テストで抽出したラベルセットがアニメ会話シーンの属性分類に有効であるかを評価した.

### 5.1 実験手順

実験は以下の手順に従って行った.

1. 聴解テストとアニメーションの会話のテキストデータを用意する.
2. 表 2 に示すラベルセットを用いてゼロショットで分類する.
3. 聴解テストに手作業で付与された 3 種類の属性の組合せの数をカウントし, 上位 10 個を頻出の組合せとする (表 3).
4. 表 3 の属性の組合せごとに分類されたアニメーションの会話テキストを 10 個ずつサンプリングする.
5. サンプリングされたアニメ会話シーンの分類が正解か否かを手作業で確認する.

表 3: 聴解テストの各頻出属性の組合せ.

属性の組合せ
用談相談, 職場, 知り合い
用談相談, 施設, 初対面
用談相談, 学校, 知り合い
雑談, 職場, 友達
用談相談, 学校, 友達
雑談, 学校, 友達
用談相談, 職場, 友達
雑談, 自宅, 友達
用談相談, 自宅, 家族
用談相談, 電話, 初対面

初めに聴解テストとアニメーションの会話のテキストデータを用意した。聴解テストは予備実験で使用したものと同じであった。アニメーションの会話は、複数のアニメから既存研究の手法 [9] を用いてシーンごとにカットされたものを用意し、合計 314,930 個であった。

実験では、4 章で説明したように、(1) MoritzLaurer と (2) Open-AI の embedding と MoritzLaurer の 2 つの手法でアニメーションをゼロショット分類した。

## 5.2 実験結果

表 4 に、MoritzLaurer モデルを用いて分類された結果を示す。10 種類の属性の組合せに対し、形式、場所、関係のいずれかが一致した数、3 種類の属性、2 種類の属性、1 種類の属性が一致した数、および一致した属性がなかった数を示す。表 4 では、100 個のサンプリングされたアニメーションの中、形式が一致したのは 81 個、場所が一致したのは 53 個、関係が一致したのは 58 個であった。3 種類の属性が全部一致したアニメーションは 31 個、2 種類が一致したのは 36 個、1 種類が一致したのは 27 個、全部一致しなかったのは 6 個であった。

表 5 は Open-AI の embedding と MoritzLaurer を用いて分類された結果を示す。100 個のサンプリングのアニメーションの中、形式が一致したのは 75 個、場所が一致したのは 54 個、関係が一致したのは 51 個であった。3 種類の属性が全部一致したのは 33 個、2 種類が一致したのは 29 個、1 種類が一致したのは 21 個、全部一致しなかったのは 17 個であった。

## 6 考察

100 個のサンプリングされたアニメーションの分類結果として、1 種類以上の属性で正解であったものに注目する。MoritzLaurer モデルを用いた場合に 94 個 (31+36+27)

一致し、Open-AI の embedding と MoritzLaurer モデルを用いた場合に 83 個 (33+29+21) 一致した。両者の平均は 88.5 となった。この結果から、データセットから得られたラベルセットを用いて、異なるデータセットの分類が行えることが確認された。

属性ごとの分類の結果に注目する。いずれのモデルにおいても「形式」の一致数が最も多く、それぞれ 81 個と 75 個であった。このことから「形式」の分類精度の高いことが確認できた。

## 7 おわりに

本研究では、ゼロショット分類を利用し、テキストを分類属性によりマッチングする手法を提案した。提案手法は、良い分類結果が得られるラベルセットを抽出し、ラベルセットを用いてテキストのゼロショット分類を行う。同一ラベルが付与されたテキスト同士をマッチングする。評価実験では提案手法で得られた 1 つのデータセットから得られたラベルセットを用いて、別のデータセットの属性の分類が行えるかを評価した。

実験では、日本語能力試験の聴解テストの会話テキストから得られたラベルセットを用いて、アニメ会話シーンのテキストを分類する実験を行った。ゼロショットの分類モデルとして、2 種類のゼロショット分類方法を用い、(1) MoritzLaurer モデル 3 種類属性のラベルセットと、(2) MoritzLaurer モデルの「場所」、「関係」のラベルセットと Open-AI の embedding の「形式」のラベルセットの組合せを用いた。100 個のアニメーションに対し、分類された聴解テストの正解数を評価した。実験の結果、1 種類以上の属性が一致した分類結果は 100 個中 94 個 (MoritzLaurer) と 87 個 (Open-AI の embedding と MoritzLaurer) で、平均が 88.5 個となり、1 つのデータセットから得られたラベルセットが他のデータセットの属性分類に適用できることを確認した。

今後、提案手法のマッチング精度を評価する実験を行う。

## 参考文献

- [1] Masaya Hanazono, Takashi Nishiyama, Keiichi Tokuda: An optimization method of constructing corpora for hmm-based emotional speech synthesis, *Proceedings of the Annual Conference of the Institute of Systems, Control and Information Engineers*, Vol. SCI09, No. 0, pp. 303–303(2009)

表 4: MoritzLaurer モデルを用いてアニメーションを分類した結果. 各属性の組合せに一致した属性・属性種類数の統計.

属性の組合せ	形式	場所	関係	3 種類の属性	2 種類の属性	1 種類の属性	なし
用談相談, 職場, 知り合い	6	7	3	2	4	2	2
用談相談, 施設, 初対面	8	3	3	1	2	7	0
用談相談, 学校, 知り合い	10	9	5	4	6	0	0
雑談, 職場, 友達	8	1	8	1	6	2	1
用談相談, 学校, 友達	8	9	8	7	1	2	0
雑談, 学校, 友達	10	7	9	7	2	1	0
用談相談, 職場, 友達	8	6	7	4	3	3	0
雑談, 自宅, 友達	8	4	7	3	4	2	1
用談相談, 自宅, 家族	8	6	3	2	4	3	1
用談相談, 電話, 初対面	7	1	5	0	4	5	1
合計	81	53	58	<b>31</b>	<b>36</b>	<b>27</b>	6

表 5: Open-AI の embedding と MoritzLaurer モデルを用いてアニメーションを分類した結果. 各属性の組合せに一致した属性・属性種類数の統計.

属性の組合せ	形式	場所	関係	3 種類の属性	2 種類の属性	1 種類の属性	なし
用談相談, 職場, 知り合い	10	5	3	3	3	3	1
用談相談, 電話, 初対面	7	1	5	1	2	5	2
用談相談, 学校, 知り合い	8	8	7	4	5	1	0
雑談, 職場, 友達	7	2	5	2	3	2	3
用談相談, 学校, 友達	7	8	4	4	2	3	1
雑談, 学校, 友達	8	7	9	7	1	1	1
用談相談, 職場, 友達	8	6	7	5	2	2	1
雑談, 自宅, 友達	6	4	5	3	3	0	4
用談相談, 自宅, 家族	7	5	3	1	5	2	2
用談相談, 施設, 初対面	6	8	3	3	3	2	2
合計	75	54	51	<b>33</b>	<b>29</b>	<b>21</b>	17

- [2] Zakaria Elberrichi, Abdelattif Rahmoun, Mohamed Amine Bentaalah: Using wordnet for text categorization, *International Arab Journal of Information Technology*, Vol. 5, No. 1(2008)
- [3] 小磯花絵, 天谷晴香, 居友里子, 白田泰如, 柏野和佳子, 川端良子, 田中弥生, 伝康晴, 西川賢哉: 『日本語日常会話コーパス』モニター版の設計・評価・予備的分析, 国立国語研究所論集, pp. 17–33(2020)
- [4] Michael J Pazzani, Daniel Billsus: Content-based recommendation systems, *The adaptive web: methods and strategies of web personalization*, pp. 325–341(2007)
- [5] 篠原宜道, 渡部広一, 河岡司: 常識判断システムを用いた会話意味理解方式~会話文からの場所推定~, 情報処理学会研究報告自然言語処理 (NL), Vol.2003, No.4(2002-NL-153), pp. 89–96(2003)
- [6] 小井出慎, 古宮嘉那子, 佐々木稔, 新納浩幸: 物語における登場人物の親しさ推定, 言語処理学会第22回年次大会発表論文集, pp.985–988(2016)
- [7] 澤木智史, 仲山泰: ゼロショットテキスト分類を活用した含意判定モデルによるFRB金融政策コミュニケーションの読み解き, 人工知能学会第二種研究会資料, Vol. 2023, No. FIN-030, pp. 72-77(2023)
- [8] 鈴木大志, 鷹合大輔, 中沢実: AutoVCを用いたゼロショットトライアルタイム声質変換手法の実装と評価, 情報処理学会論文誌, Vol. 65, No. 2, pp. 529-237(2024)
- [9] Junjie Shan, 西原陽子, 山西良典, 福本淳一: 日本語の単語と文法のレベルを用いたアニメのセリフ

の難易度の推定手法, 人工知能学会全国大会論文  
集, Vol. JSAI2017, pp. 4F14 (2017)

- [10] NI Yangdi, SHAN Junjie, 韓毅弘, 西原陽子: 会  
話属性により分類されたアニメ会話を用いた日本  
語聴解テスト練習支援での妥当性考察, 第 9 回コ  
ミック工学研究会, pp. 60-64(2023)
- [11] Yangdi Ni, Junjie Shan, Yoko Nishihara: Match-  
ing of Japanese Listening Test Dialogues and  
Anime Scene Dialogues based on Zero-shot At-  
tribute Classification *Procedia Computer Sci-  
ence*, Vol. 246, pp. 3820-3829(2024)