

指描きによる疼痛の波形表現に対するオノマトペの推定手法

Estimation Method of Onomatopoeia for Pain represented as Waveform Drawn by a Finger

西原陽子^{1*} 羽鹿諒² 和氣紗也佳¹ 山西良典¹
Yoko Nishihara¹ Ryo Hajika² Sayaka Wake¹ Ryosuke Yamanishi¹

¹ 立命館大学情報理工学部

¹ College of Information Science and Engineering, Ritsumeikan University

² Faculty of Engineering, the University of Auckland

Abstract: This paper proposes an estimation method of onomatopoeia for pain represented as waveform i.e., time series signal information. A pain is represented as a waveform drawn by a finger on a pressure-sensitive touch panel device. The waveform has time-series information in three-dimensions: a finger tip movement on a pressure-sensitive touch panel. The information about the frequency of the waveform is analyzed to obtain features to represent the waveform as a feature vector. Feature vectors are used to make a model of the relation between onomatopoeia and pain. The obtained model can be used for estimation of onomatopoeia. We conducted evaluation experiments and verified that the proposed method could estimate onomatopoeias partially.

1 はじめに

医療機関で治療やリハビリテーションを受ける際には、まず患者は医療従事者による問診を受ける。問診で患者が医療従事者に抱えている問題の内容、例えば疼痛の部位や感じ方、その程度を正確に伝えることができる、効果的な治療やリハビリテーションを受けることができると考えられる。

しかし、疼痛は経験した人にしか分からない主観的な情報である。疼痛の情報は大きく3つに分けられ、それぞれ部位、感じ方、程度がある。部位や程度に比べ、疼痛の感じ方を言葉のみで伝えることは難しく、語彙と表現力が必要となる。語彙が乏しい子供や、母語が異なる人は疼痛の感じ方を適切に伝えることが困難になる。実際に東日本大震災の際には、医療従事者と方言を話す患者さんの間でコミュニケーションが取りづらかった事例も報告されている [1]。したがって、疼痛の感じ方を言葉を介さずとも表現でき、それが言葉として理解できるような仕組みが求められている。

本研究では疼痛の感じ方を計量可能な信号情報に変換し、信号情報から対応するオノマトペを推定する試みについて報告する。具体的には、タッチパネルに指で描くことにより、疼痛の感じ方を信号情報として表現するインタフェースと、信号情報から対応するオノ

マトペを推定する手法を提案する。単純な図形の単純な動きにより対象の感情を表現可能なことは既存研究において示されている [4]。このことから感情と図形やその動きには関連があることになる。本研究では感情の代わりに疼痛の感じ方を扱い、図形の代わりに指描きの3次元の波形を扱う。疼痛の感じ方と3次元の波形の間には何らかの関連があると考えられ、その関連を提案手法を用いてモデル化する。

提案手法では入力された3次元の指描きの波形に対し、疼痛の感じ方に対応するオノマトペをラベリングする。医療者が自らの疼痛の感じ方について、事前にオノマトペと3次元の指描きの波形の対応表を作成できていれば、患者が入力する3次元の指描きの波形について自らのオノマトペとの対応づけが可能となり、疼痛の感じ方が把握しやすくなると考えられる。さらに、他の医療者と患者の情報について共有する際にも、オノマトペを介して行うことができるようになり、引き継ぎが行いやすくなると考えられる。加えて、患者自身も自らの疼痛の感じ方を表現する適切な方法をオノマトペを通じて学んでいくことができる。これらの理由から、本研究では指描きの3次元の波形とオノマトペを対応づける方法を提案する。

既存研究として、子供向けの痛みの程度を表現できる評価ツールが提案されている [6]。球体を基本とし、突起物が付属するツールであり、7種類のパーツで1つのツールとなる。パーツの中から痛みの程度を最も表

*連絡先: 立命館大学情報理工学部
草津市野路東 1-1-1
E-mail: nishihara@fc.ritsumei.ac.jp

現するものを1つ選択することにより痛みの程度を表現する。本研究では疼痛の程度ではなく、疼痛の感じ方を推定する手法を提案する。疼痛の感じ方を推定し、その後、既存手法により痛みの程度を推定すると、疼痛の内容がより伝わりやすくなると考えられる。

オノマトベが与えられると各評価尺度に対し、痛みの質と量を定量的に提示する手法 [5] がある。既存研究ではオノマトベを元にし、痛みの内容を推定するが、本研究では反対に、疼痛の感じ方が信号情報として与えられるとオノマトベを推定する手法を提案する。適切なオノマトベを思いつけなくとも、利用可能な手法になる可能性がある。

2 提案手法

疼痛の感じ方として3次元の指描きの波形が与えられると、疼痛の感じ方に対応するオノマトベを推定する手法を説明する。提案手法は入力のための装置として、感圧式のタッチパネルを用いる。ある疼痛の感じ方について、タッチパネル上に指で描くことにより指描きの3次元の時系列データを取得する。得られた指描きの3次元の時系列データを周波数分析し、特徴量のベクトルを作成する。機械学習を用いて、ベクトルをオノマトベのクラスに分類するモデルを作成する。疼痛の感じ方の新たな指描き表現が得られると、作成したモデルを用い、対応するオノマトベを推定することができる。

2.1 入力と入力装置

入力は、ある疼痛の感じ方に対する3次元の指描きの時系列データとする。3次元の指描きの時系列データとは、タッチパネル上の水平方向(X軸)、垂直方向(Y軸)、鉛直方向(圧力、Z軸)が一定時間ごとに取得されたデータとする。

提案手法では、入力装置として感圧式のタッチパネルを用いる。本稿ではデプスセンサが搭載されたiPhone8を入力装置とした。入力のためのインタフェースをSwiftを用いて作成した。インタフェースの画面を図1に示す。下半分の灰色の部分が指で描き、疼痛の感じ方を入力する部分になる。灰色の部分の横軸がX軸、縦軸がY軸、X-Y軸からなる平面の鉛直方向がZ軸となる。入力された3次元の指描きの時系列データを上半分のグラフ部分に表示する。インタフェースで入力する様子と、得られた指描きの波形(ジンジン、X-Y平面のみ)の例を図2と図3に示す。計測時間は最大10秒、計測周波数は60Hzとした。計測された3次元の指描きの時系列データはcsvファイルとして保存される。

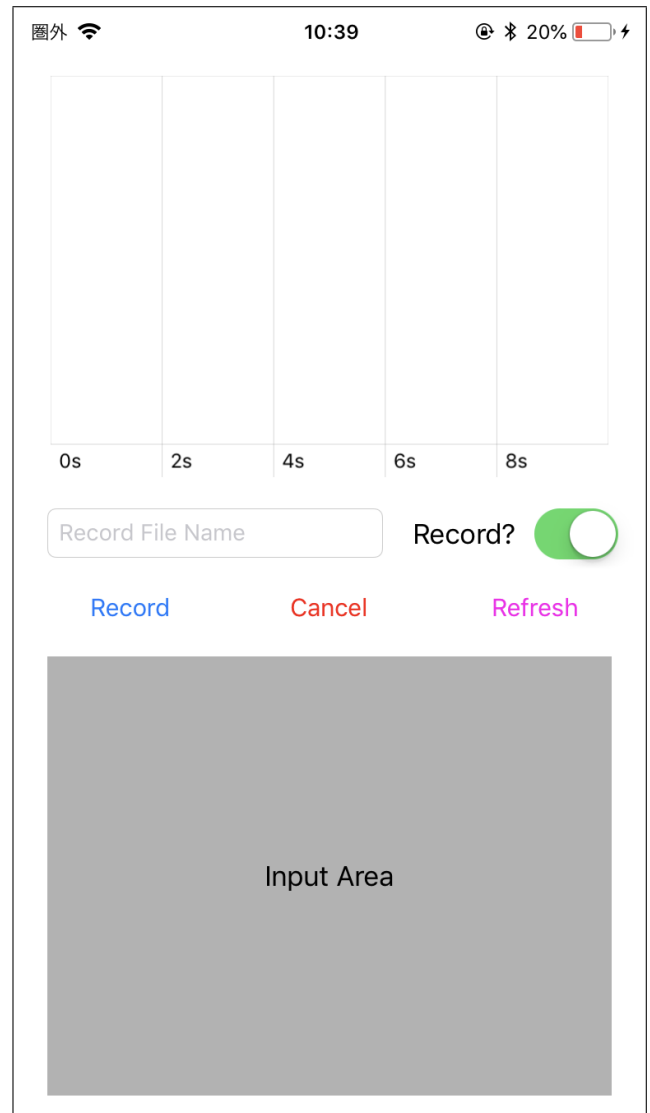


図1: 作成した入力インタフェース。下半分の灰色の部分が指で描いて疼痛の感じ方を入力する部分。入力された3次元の指描きの時系列データを上半分のグラフ部分に表示する。

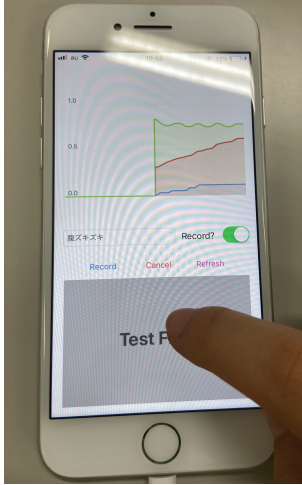


図 2: インタフェースで指描きの波形を入力をする例.

2.2 指描きの時系列データの周波数分析

指描きの 3 次元の時系列データの周波数分析を行い、機械学習に与える特徴量のベクトルを作成する. X 軸, Y 軸, Z 軸の 3 つの軸に対して得られた指描きの時系列データに対し, フーリエ変換を行う. フーリエ変換により得られたパワースペクトルの包絡線に対し, 近似曲線を求め, 近似曲線の係数を特徴量として採用する. フーリエ変換には式 (1) を用いる.

$$F(t) = \sum_{x=0}^{N-1} f(x)e^{-i\frac{2\pi tx}{N}}$$

$$= \sum_{x=0}^{N-1} f(x)\left(\cos\frac{2\pi tx}{N} - i\sin\frac{2\pi tx}{N}\right), \quad (1)$$

ただし, N はデータの個数であり, 本稿では $N = 512$ とした.

続いて, X 軸, Y 軸, Z 軸のパワースペクトルの包絡線に対し, 式 (2) により表現される近似曲線を求める.

$$y = ax^6 + bx^5 + cx^4 + dx^3 + ex^2 + fx + g. \quad (2)$$

式 (2) での x は周波数, y はパワーとする.

式 (2) により各軸に対し, a から g までの 7 つの係数が得られる. 得られる係数を順に並べ, 式 (3) により指描きの波形を表現する特徴ベクトル v を得る.

$$v = [x_a, x_b, x_c, x_d, x_e, x_f, x_g, \\ y_a, y_b, y_c, y_d, y_e, y_f, y_g, \\ z_a, z_b, z_c, z_d, z_e, z_f, z_g], \quad (3)$$

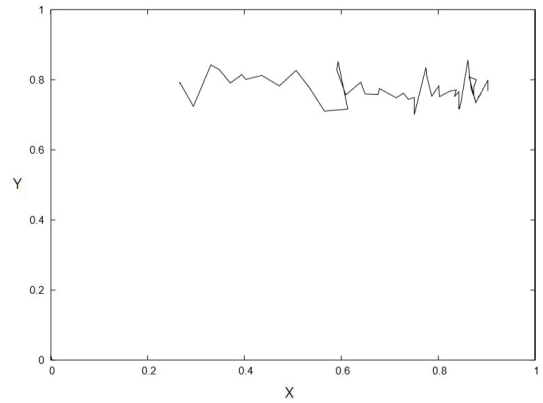


図 3: インタフェースで入力された指描きの波形の例. オノマトペはジンジン (X-Y 平面のみ). 左から右へ描かれた.

ただし, X 軸に対する特徴量を $x_a, x_b, x_c, x_d, x_e, x_f, x_g$, Y 軸に対する特徴量を $y_a, y_b, y_c, y_d, y_e, y_f, y_g$, Z 軸に対する特徴量を $z_a, z_b, z_c, z_d, z_e, z_f, z_g$ とする. 特徴ベクトル v の次元数は 21 次元 ($= 7 \times 3$) となる

2.3 オノマトペの推定モデルの作成

式 (3) により表現される特徴ベクトルにオノマトペの種類を示すラベルを付与し, 合計 22 次元のベクトルを作成する. 機械学習によりベクトルをオノマトペのクラスに分類するモデル, つまりオノマトペの推定モデルを作成する. 本稿では, オノマトペの推定モデルの作成には, 多クラス分類をするサポートベクタマシンを用いた.

3 評価実験

提案手法によるオノマトペの推定精度を評価した.

3.1 実験手順

1. 実験者は被験者に疼痛の部位と疼痛の感じ方としてオノマトペを指定する. 被験者は疼痛を想像しながら, インタフェースで指描きの波形を入力する.
2. 実験者は入力された指描きの 3 次元の波形のデータを用い, オノマトペを分類するモデルを作成する.
3. 実験者はオノマトペの推定モデルの精度を評価する.

表 1: 3 種類のオノマトペに対する推定により得られた混合行列. 上段が提案手法, 下段が比較手法.

提案手法	ジンジン	ズキズキ	チクチク
ジンジン	17	4	6
ズキズキ	19	3	5
チクチク	6	0	21
比較手法	ジンジン	ズキズキ	チクチク
ジンジン	20	2	5
ズキズキ	19	2	6
チクチク	6	1	20

被験者は立命館大学に所属する学生 20 名であった. 年齢は 20 歳から 27 歳, 男性 14 名, 女性 6 名であった.

指定する疼痛の部位とオノマトペは既存研究 [2, 3] を参考にし用意した. 部位ごとに頻繁に用いられるオノマトペは異なるが, 「ジンジン」「ズキズキ」「チクチク」はどの部位であっても頻繁に用いられる. このため, この 3 種類を採用した. 部位については被験者が経験したことが多いと考えられる頭, 腹, 脚とした. 獲得した疼痛の種類は 9 種類 (=3 オノマトペ × 3 部位) になった. 獲得した疼痛のデータ数は 1 種類につき 9 件であった. それぞれ異なる被験者から獲得し, データは合計で 81 件となった.

推定精度を比較するために, 計測周波数を 10Hz に落とした手法を用意した. 以下, この手法を比較手法と呼ぶ.

推定精度を評価するために 2 種類の観点を用意した.

1. 3 種類のオノマトペの分類
3 種類のオノマトペを分類できるかを評価した.
2. 2 種類のオノマトペの分類
ジンジン vs チクチク, ズキズキ vs チクチク, ジンジン vs ズキズキという 2 種類ずつでオノマトペを分類できるかを評価した.

それぞれ混合行列を作成し, 正しく推定された割合を算出し, 精度とした. 精度を提案手法と比較手法の間で比較した. 提案手法, 比較手法でモデルを作成し, テストを行う際は, 1-out-leave で行った.

3.2 実験結果

3 種類のオノマトペの分類における混合行列を表 1 に示す. 3 種類のオノマトペの分類における精度は, 提案手法が 50.6%, 比較手法が 51.9%となった.

2 種類のオノマトペの分類における混合行列を表 2, 表 3, 表 4 に示す. 2 種類のオノマトペの分類における精度は, 提案手法が 65.4%, 比較手法が 68.5%となった.

表 2: ジンジン vs チクチクに対する推定により得られた混合行列. 上段が提案手法, 下段が比較手法.

提案手法	ジンジン	チクチク
ジンジン	21	6
チクチク	6	21
比較手法	ジンジン	チクチク
ジンジン	22	5
チクチク	7	20

表 3: ズキズキ vs チクチクに対する推定により得られた混合行列. 上段が提案手法, 下段が比較手法.

提案手法	ズキズキ	チクチク
ズキズキ	22	5
チクチク	6	21
比較手法	ズキズキ	チクチク
ズキズキ	21	6
チクチク	7	20

3.3 考察

3 種類のオノマトペの分類について考察する. 3 種類のオノマトペの分類における精度は, 提案手法が 50.6%であった. チャンスレベルが 33.3% (= 1/3) であり, チャンスレベルよりも 17.3%高い精度が得られた. このことから, 提案手法によりオノマトペの推定が可能ながことが示された.

3 種類のオノマトペの分類における精度は, 比較手法が 51.9%となった. 比較手法の方が 1.3%高かったが, 2 つの手法の間に推定精度の大きな差はないと考えられる. 提案手法の方が計測周波数が高いため, より詳細な計測が可能となっている. しかし, 精度に大きな差が出なかったため, 学習のさせ方に課題があると考えられる.

2 種類のオノマトペの分類について考察する. ジンジン vs チクチクの分類精度は提案手法と比較手法は同じで, 77.7%であった. ズキズキ vs チクチクの分類精度は提案手法が 79.6%で, 比較手法は 75.9%であった. これに対し, ジンジン vs ズキズキの分類精度は提案手法が 38.8%, 比較手法が 51.8%であった.

ジンジンとズキズキの分類精度が, 他 2 種類の分類 (ジンジン vs チクチク, ズキズキ vs チクチク) よりも低くなった. 27 個のズキズキのデータのうち, 23 個がジンジンと推定された. ジンジンとズキズキの違いは被験者自身もあまり分からないという意見や, 痛みが似ているという意見があった. いずれの疼痛表現も画

表 4: ジンジン vs ズキズキに対する推定により得られた混合行列。上段が提案手法，下段が比較手法。

提案手法	ジンジン	ズキズキ
ジンジン	17	10
ズキズキ	23	4
比較手法	ジンジン	ズキズキ
ジンジン	20	7
ズキズキ	19	8

面に波を描くように入力されることが多かった。ジンジンとズキズキの波形表現が類似していたため，分類の精度が低くなったと考えられる。

一方で，チクチクの入力のされ方は，画面から何度も指を離してタッチするような入力の仕方であり，他の2つのオノマトベとは異なっていた。波形の形に特徴があったために，チクチクと他のオノマトベを分類する精度は高くなったと考えられる。

4 おわりに

本研究では疼痛の感じ方を計量可能な信号情報に変換し，信号情報から対応するオノマトベを推定する手法を提案した。ある疼痛について，感圧式のタッチパネルを利用し，指で描くことにより3次元（X軸，Y軸，圧力）の指描きの時系列データを取得する。これを周波数分析して得られた特徴パラメータをベクトルとして獲得し，波形表現を介し，疼痛の感じ方に対応するオノマトベとの関係をモデル化する。

評価実験では，3種類のオノマトベ（ジンジン，ズキズキ，チクチク）に対する指描きの3次元の時系列データを獲得し，時系列データから疼痛の感じ方に対応するオノマトベを推定する精度を評価した。提案手法の分類精度の平均は50.6%であり，ランダムに推定する場合の33.3%（= 1/3種類）よりも17.3%高かった。評価実験の結果より，提案手法が指描きの波形に対応するオノマトベの推定を一部可能にすることが確認された。

今後の課題は推定精度の向上，およびインタフェースの改善があげられる。推定精度の向上のために，波形を表現するベクトルを改良し，モデル化に用いる手法の再検討を行う。インタフェースの改善のために，ユーザが波形を入力する領域を広くとるように，オブジェクトの配置を見直す。また，疼痛を感じる部分と疼痛の感じ方を関連づけて入力することで，より現実に近いデータが入力される可能性がある。そこで，インタフェース上に体のモデル図や，実際のユーザの体のパー

ツを撮影した画像を示すなどし，疼痛の感じ方の入力のされ方の違いを考察し，インタフェースの改善に役立てる。

参考文献

- [1] 岩城裕之，災害時に見る医療と地域の「問題」－医療者と住民の円滑なコミュニケーションのために，日本ヘルスコミュニケーション学会雑誌，Vol.6，No.1，pp.6–9，2015.
- [2] 楠見孝，中本敬子，子安増生，痛みの比喩表現の身体感覚と認知の構造，心理学研究，Vol.80，No.6，pp.467–475，2010.
- [3] オノマトグラム：メディカルオノマトベ，<http://onomatopelabo.jp/medical/gram/index.html>（2020年2月12日アクセス確認）.
- [4] 富川道彦，尾田政臣，単純な動きを示す対象図形の感情推定，映像情報メディア学会技術報告，Vol.33，No.17，pp.1–4，2009.
- [5] 上田祐也，清水祐一郎，坂口明，坂本真樹，オノマトベで表される痛みの可視化，日本バーチャルリアリティ学会論文誌，Vol.18，No.4，pp.455–463，2013.
- [6] 山下利之，近藤真悟，岡崎章，小児看護におけるプレパレーションのための痛み評価ツールの開発，人間工学，Vol.51，No.6，pp.435–440，2015.

看板画像を用いた検索システムにおける 学習モデル構築手法の検討

Construction Method of Learning Model in Search System with Signboard Images

三溝 俊介* 松下 光範
Shunsuke Mitsumizo Mitsunori Matsushita

関西大学総合情報学部
Faculty of Informatics, Kansai University

Abstract: The aim of the research is to develop a search system for foreigners who explore shops when they visit Japan. The system provides detailed information from shops with signboards written in Japanese characters that they do not understand. To develop the system efficiently, we constructed a learning model for signboard classification with dataset constructed from data that exist on the web. Our proposed system enables users to obtain detailed information by snap signboards with his/her own smartphone. The field experiment that we conducted revealed that the proposed method can determine shops appropriately.

1 はじめに

人々が都市空間において情報を取得する手段のひとつとして看板が挙げられる。看板の目的はその看板が掲示されている場所に存在する店舗や会社の名称を提示してその存在を知らせたり、企業や製品、音楽や映画などを告知したりすることが主であり、その存在が人の目に入ることで、そこに描かれているコンテンツやデザイン（文字等を含む）に何らかの影響を受けて具体的な行動に移すことが企図されている [1]。

一方、近年の携帯端末の普及に伴い、都市空間における情報取得手段が変化してきている。携帯端末では、テキストを用いた検索や GPS 情報を用いた位置検索など多様な検索手法が提供されており、人は得たい情報に応じて様々な検索手法を使い分けることで、任意の場所で必要な情報を取得できるようになってきている。

こうした情報取得手段の多様化は、街中での店舗検索行動を変えつつある。例えば、人々が外出先で飲食店を選択する際には、提供される料理のメニューやその店舗の口コミなどの情報を検索したり、WEB 配信されている割引クーポンの有無を検索したりすることが一般的になってきている。看板は、これまでの役割に加え、こうした店舗検索行動におけるトリガーにもなっている。

このような店舗検索を行う際、ユーザが看板の文字を理解し、携帯端末に店名などのテキストをキーワードとして入力することが前提となっているが、看板に書かれている文字が利用者の母国語でない場合や、崩した文字で書かれているような場合（蕎麦屋や割烹など）は、検索が容易でない。

北村らはこの問題に対し、携帯端末のカメラで撮影した看板を機械学習を用いて分類し、AR を用いて店舗の詳細情報を画面上に重畳表示することで解決を図った [2]。しかし、北村が提案した手法では、看板を分類する機械学習モデルを構築するために手動で看板画像の収集とアノテーションを行っているため、データセットの構築に多大なコストがかかるという問題があり、検索可能な店舗を増やすことは容易ではなかった。そこで、本研究ではこのデータセット構築を自動化することで対象店舗の拡大に要するコストの削減を図る。その端緒として本稿では対象を飲食店に限定し、その看板画像をグルメサイトから収集して、機械学習モデルを構築する手法を提案する。

2 関連研究

本章では本研究で用いるオブジェクトの検出・分類手法について述べる。

オブジェクト検出手法の 1 つである You Only Look Once (YOLO) や YOLO を改良した YOLOv2 では入

*連絡先：関西大学総合情報学部
〒569-1095 大阪府高槻市霊仙寺町 2-1-1
E-mail: mat@res.kutc.kansai-u.ac.jp

力画像を複数のグリッドに分割し、セルごとにどのクラスかを予測し、グリッドとは別に複数のバウンディングボックスとボックス内にオブジェクトが存在する確率を求め、オブジェクトの領域とクラスを出力する[3]。YOLOv2ではリアルタイムなオブジェクト検出が可能だが、セルごとにクラス分類を行うという性質上、1つのセルに複数のオブジェクトが存在する場合に正確な分類が困難である。

オブジェクト分類手法の1つにVGG16モデルがある[4]。VGG16はImageNet[5]のILSVRC-2012データセットに含まれる1000クラスに分類された130万枚の画像を学習させたCNNモデルの1つであり、13層の畳み込み層と3層の全結合層の計16層で構成されている。VGG16では入力された画像を複数回の畳み込みとプーリングを行い、分類したクラスとその確率を出力する。

3 看板を用いた検索手法

本章では提案手法の基礎となる看板を用いた検索手法について述べる。

北村らは、看板画像から店舗情報にアクセスする手法を提案している[2]。人手で看板画像の収集とアノテーションを行ったデータセットを用いて、看板画像の認識・分類を行うモデルを構築し、送信された看板画像から店舗情報を返却するWebAPIサーバの構築を行っている。この手法では、一定時間ごとに携帯端末で撮影された画像をAPIサーバに送信し、APIサーバ上で分類した店舗に紐付いたOpenStreetMap (OSM) [6]に登録されている情報をOverpass API¹で取得している。このシステムではTensorflow1.0[7]で実装したYOLOv2を用いて看板領域の検出を行っており、検出した領域をVGG16で分類している。分類した看板が存在する座標にその店舗の情報を吹き出しによって携帯端末上に重畳表示することで、リアルタイムな認識と情報提示を実現している。

この手法は看板の分類に文字列ではなく看板画像を用いているため、文字の並び方が特殊な場合や崩して書かれた文字であっても分類が可能である。しかし、北村らの提案手法では看板画像の収集及びアノテーションを手動で行う必要があり、対象地域の拡大を行う際に多大なコストがかかる。

4 デザイン指針

本章では、提案手法の実装にあたってのデザイン指針を定める。

3章で述べたように、文字認識を行わず、看板画像を分類することで文字の並び方が特殊な場合や崩した文字であっても分類が可能である。そのため、本研究では北村らの手法を参考に看板画像を分類することで検索を行う。しかし、北村らの提案手法は手動でデータ収集とアノテーションを行っており、対象地域の拡大に要するコストが莫大であった。そのため、本研究ではこれらのコストを削減するために、ウェブ上に存在するデータを活用し、自動でデータの収集とアノテーションを行う。

本研究で実装する検索システムは、外出時にユーザーが目の前にある飲食店に関する詳細な情報を得たいが、文字が読めず検索ができない状況を想定している。そのため、看板から離れた位置で行う検索や、看板と遮蔽物が重なった状態での検索は対象としない。

上述したデザイン指針とシステムが使用される状況から本システムが満たすべき要件を以下に定める。

1. データ収集が自動で行われること
2. アノテーションが自動で行われること
3. 複雑な文字であっても検索が可能であること
4. 実世界において飲食店の検索が可能であること

これら4つの要件を満たすシステムの実現を目指し、実装したシステムでフィールド実験を行うことで、実世界において検索が可能か評価を行う。

5 実装

本研究では、4章で述べた要件を達成できるシステムの実装を行う。要件1を達成するために、ウェブ上のデータからスクレイピングで看板画像の収集を行う。要件2を達成するために、画像の収集は整理されたウェブサイトから行う。要件3を達成するために、文字認識を行わず看板画像の分類を行うことで複雑な文字であるかどうかに関わらず認識を可能にする。要件4が達成されていることを確認するために、実装したシステムを用いてフィールド実験を行う。

本研究では対象とする店舗をグルメサイト食べログ²に掲載されている飲食店とし、BeautifulSoup³を用いて食べログの外観画像ページに存在する店舗の外観画像をスクレイピングによって取得する。スクレイピングは以下の手順で2019年9月26日から9月27日にかけて行った。まず、食べログのランキングページの1ページ目を取得し、取得したページからhtmlのaタグでclassが「list-rst__rst-name-target」の要素に各店

¹<http://overpass-api.de/> (2020/1/31 存在確認)

²<https://tabelog.com/> (2020/1/31 存在確認)

³<https://pypi.org/project/beautifulsoup4/> (2020/1/31 存在確認)

舗情報のページの URL があるため、そこから各店舗情報ページの URL を取得し、リストに格納する。次にリストに格納された各店舗情報ページの URL の末尾に「dtlphotolst/dtlphotolst/4/smp2/D-normal/1/」を加えることで各店舗の外観画像ページの 1 ページ目にアクセスできる。外観画像ページには 1 ページあたりに 40 件の投稿された画像があり、a タグで class が「js-imagebox-trigger js-analytics rstdtl-thumb-list__target」の要素に各画像の URL があるため、その URL から各画像を保存する。ページ内の画像を全て取得した場合は末尾の数字を 1 つ増やすことで次のページにアクセスする。同様の処理を最後の外観画像ページまで行い、取得できる画像がなくなるとリストに保存されている次の店舗で同様の処理を繰り返す。リストの中に保存されている店舗の画像を全て取得した場合はランキングページの次のページで同様の処理を行う。以上の処理を指定回数行うことで外観画像を収集する。

スクレイピングによって得られた画像を次節で述べる YOLOv2 を用いて看板を検出した。検出された信頼度が 80% 以上の看板の領域を切り出し、切り出された画像が 20 枚以上存在する店舗を対象に、画像の 60% をトレーニングデータ、20% をバリデーションデータ、20% をテストデータとして店舗ごとに保存した。看板の検出には YOLOv2 を用いる [3]。YOLOv2 によるオブジェクト検出は 2 章で述べたように、1 つのセルの中に多数のオブジェクトが存在する場合の正確な分類が困難である。しかし、4 章で述べたように、本研究で対象とする状況は目の前に存在する店舗の検索を行う状況であるため、多数の看板が 1 つのセルの中に密集することは稀である。そのため看板の検出には YOLOv2 を採用した。本研究では、北村が学習させたモデルを用いて看板の検出を行う。このモデルは特定の店舗の看板を検出するのではなく様々な店舗の看板を検出することが可能な汎用的なモデルであるため、YOLOv2 で学習させていないスクレイピングで収集した店舗の看板を検出することが可能である。また、北村の行った実験で実世界での看板検出を十分に行えることが確認されているため、看板の検出アルゴリズムとしてその精度が担保されていると考えられる。

看板の分類には VGG16[4] を用いる。ニューラルネットワークのライブラリである Keras で実装された VGG16 を用いて、北村の手法を参考に、全結合層を取り除き、新たに全結合層を追加する転移学習を用いた⁴。そのため、学習に用いる画像が少ない店舗であっても短時間で比較的高精度なモデルの構築が可能である。

上述したモデルを用いて、携帯端末のカメラで撮影された看板の店舗を検索するシステムを実装した。看板認識の処理は高負荷であり、ハイスペックな GPU を

表 1: 看板画像の認識を行うマシンのスペック

要素	スペック
CPU	Intel® Core™ i7-8700K @ 3.70GHz
RAM	16GB
GPU	NVIDIA GeForce GTX1080ti
VRAM	12GB
OS	Ubuntu 18.04

要する。そのため、これらの処理は携帯端末上では行わず、ハイスペックな GPU を搭載したマシン上で看板の認識を行うウェブ API サーバを構築し、携帯端末から 500 ミリ秒ごとにこの API を呼び出すことで分類の処理を行う。看板認識の処理を行うマシンのスペックを表 1 に示す。API の実装には Python3.7.4 を用いて WebAPI フレームワークの Falcon⁵を用いて構築する。この API は携帯端末から送信された画像内の看板領域を Tensorflow1.13.1[7] で実装された YOLOv2 を用いて検出し、検出された看板を VGG16 で分類する。検出した看板領域の左上の座標と右下の座標及び判別した店舗の食べログページの URL を JSON 形式で返却する。

携帯端末で検索する際のインタフェースは HTML5 と JavaScript を用いて実装した。JSON 形式で返却された看板領域の左上と右下の座標からボックスを生成し、その領域以外の明度を下げることで看板領域を図 1 のようにハイライトする。店舗情報の表示には AR を用いず、ハイライトされた領域をユーザがタップした際に、返却された URL に遷移する。また、API に送信された画像内に複数の看板が検出された場合は、ユーザがタップした店舗に対応する URL に遷移する。

6 評価

対象地域拡大によって分類するクラス数が増加するため、クラス数の増加による分類精度への影響を確認する必要がある。そのため、収集した画像から 150 店舗の分類を行うモデルと 30 店舗の分類を行うモデルの 2 種類を構築した。150 店舗の分類を行うモデルは画像サイズを 224×224、バッチサイズを 16、学習率を 10^{-4} 、慣性を 0.9、エポック数を 300 で学習させ、30 店舗の分類を行うモデルはエポック数を 100 で残りの条件は同じ状態で学習させる。最適化アルゴリズムには Stochastic Gradient Descent (SGD)、損失関数には交差エントロピーを用いる。また、学習時に Keras の ImageDataGenerator を用いて学習画像の拡張を行う。

あるクラスについて陽性と予測したクラスが実際に陽性であるものを True Positive (TP)、陽性と予測し

⁴<https://keras.io/> (2020/3/2 存在確認)

⁵<https://falconframework.org/> (2020/1/31 存在確認)



図 1: ハイライトされた看板

たクラスが実際には陰性であったものを False Positive (FP), 陰性と予測したクラスが実際には陽性であったものを False Negative (FN), 陰性と予測したクラスが実際に陰性であったものを True Negative (TN) とした場合に陽性と予測したものの中でそれが実際に正しかった割合である適合率, 実際には陽性であるクラスの中で陽性と予測したものの割合である再現率, 適合率と再現率の調和平均である F 値はそれぞれ以下の式に表される.

$$\text{正解率} = \frac{\text{予測と正解の一致数}}{\text{テストデータ数}} \quad (1)$$

$$\text{適合率} = \frac{TP}{TP + FP} \quad (2)$$

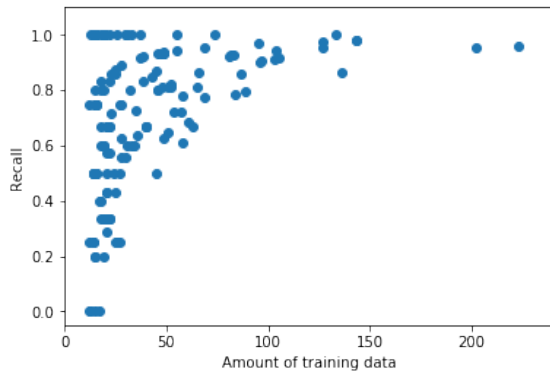
$$\text{再現率} = \frac{TP}{TP + FN} \quad (3)$$

$$F \text{ 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \quad (4)$$

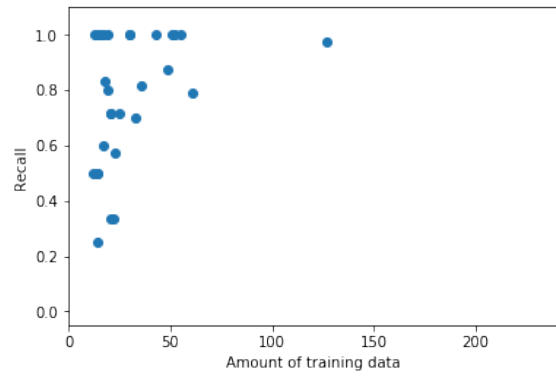
テストデータを用いて 2 つのモデルについて全体の正解率及び適合率, 再現率, F 値に関して全クラスの結果の平均であるマクロ平均を算出した. 150 クラスの分類を行うモデルの全体の正解率は 0.808, 適合率は 0.801, 再現率は 0.710, F 値は 0.726 であった. また, 30 クラスの分類を行うモデルの全体の正解率は 0.848, 適合率は 0.873, 再現率は 0.784, F 値は 0.795 であった. 150 クラスのモデルと 30 クラスのモデルを比較した場合, 全てにおいて 30 クラスのモデルの方が高いことが確認できた. そのため, より数の多いクラスの分類を行う際に分類の精度が低下することが示唆された. また, 150 店舗のモデルにおけるトレーニングデータの数に対する F 値の変化を図 2 (a) に示し, 30 店舗のモデルにおいてのトレーニングデータの数に対する F 値の変化を図 2 (b) に示す. 図に示されるように双方のモデルにおいてトレーニングデータの数が少ない店舗はデータの数が多くの店舗と比較して精度が低いものが多く含まれ, トレーニングデータの数が増えるほど精度の向上が低減することが確認された. しかし, データ数が最も多い 1 店舗については F 値が大きく低下している. これは, 学習に用いたデータ数が店舗によって異なったため, 過度にこの 1 店舗が陽性となるように最適化した可能性がある.

前節で述べたように, 全体的にデータ数が増えるほど全ての評価指標における精度の向上が確認されたが, データが不均衡であったためか最もデータ数が多い 1 店舗の F 値が大きく低下していた. 本節ではトレーニングデータが 20 枚以上であった 101 店舗を対象としてトレーニングデータ数に変更を行わず学習させたモデルと全ての店舗のトレーニングデータ数を 20 枚にアンダーサンプリングして学習させた 2 つのモデルを評価する. 2 つのデータセットを 150 店舗の分類を行うモデルと同様のパラメータで学習させ, 評価には同じテストデータを用いた. アンダーサンプリングを行わなかったモデルにおける全体の正解率は 0.843, 適合率は 0.825, 再現率は 0.772, F 値は 0.778 であった. 一方でアンダーサンプリングを行ったモデルにおける全体の正解率は 0.826, 適合率は 0.807, 再現率は 0.784, F 値は 0.778 であり, アンダーサンプリングによって全体の F 値は改善されなかった.

テストデータだけでなく実世界においても, システムが使用可能であることを確認するために, 5 章で述べたシステムを 150 店舗の分類を行うモデルを用いて実際の看板で使用した. 実装したシステムで 10 店舗の検索を行い, 各店舗ごとに 10 回ずつ複数の角度からの検索を行った. 1 つの店舗に複数の看板がある場合は 2 つの看板で 5 回ずつ検索を行った. 表 2 に検索を行った店舗名, 10 回の内正しく分類された回数を示す. 2 つの看板で検索したものは括弧内に 5 回ずつ行った結果を示す. 表 2 に示されるように, 全体的に正しく分



(a) 150 店舗のモデル



(b) 30 店舗のモデル

f

図 2: トレーニングデータ数と F 値の関係

表 2: 実世界での認識割合

店舗名	分類した割合
板前焼肉 一斗 天下茶屋本店	6 (5,1)
釜揚げうどん一紀	10
無鉄砲 大阪店	8 (3,5)
ラーメン人生 JET	10 (5,5)
まき埜	10
烈志笑魚油 麵香房 三く	8 (5,3)
ROUTE271 梅田本店	5 (0,5)
渡邊カリー 梅田本店	10
鮎処 多田	9
中村商店 高槻本店	8

類される回数が多く、店舗によっては 10 回全て正しく分類するものもあった。2 つの看板で検索したものについてはどちらか一方については 5 回の内 5 回とも正しく分類を行うことができた。「ROUTE271 梅田本店」の一度も正しく分類できなかった看板については、看板の上に時計がついている特殊なものであり、検出を行う際に時計を含めて検出される場合と看板のみが検出される場合があり、看板領域の検出が不安定であった。そのため、正しい分類が行われなかったと考えられる。また、一部の店舗でのれん画像がトレーニングデータに含まれていた。ユーザが実際にシステムを使用する場合にのれんから検索を行うことが想定できる。そのため、実験を行った店舗のうち、のれんがあった「烈志笑魚油 麵香房 三く」の検索を行う際には通常の看板に加え、のれんの検索も行った。のれんの分類は 5 回のうち 3 回は正しく分類できたが、風などによってのれんの形状が変化した場合には精度がさらに低下すると考えられる。

7 課題と展望

本研究では食べログのデータを活用することで、看板画像を用いた検索システムにおいて従来は手動で行われていたデータ収集とアノテーションの自動化を図った。これにより、多数の飲食店の看板が分類可能になった。しかし、食べログから得られるデータは飲食店のみであり、飲食店以外の店舗には拡張できないという問題がある。この問題を解決するには、多種多様なウェブ上のデータを収集し、店舗ごとに整理する必要がある。これを自動化することは容易ではない。

また、実験により、このモデルが実世界での店舗の分類にもある程度用いられることが確認されたが、データ数の少ない店舗や学習に不要な画像が混じっている店舗があった。一般的にトレーニングデータの不足や誤ったデータの混入が精度に影響を与える。このことから開店して間もない店舗や看板と誤認識するオブジェクトが含まれる店舗の分類が困難である。そのため、精度の向上には構築されたデータセットに含まれている不要なデータを取り除く処理が必要であると考えられる。

本研究では、対象とするデータを飲食店の看板としたが、飲食店の看板には矩形の中に文字が存在する典型的な看板から矩形が無い看板や特殊な形をしたものまで様々な看板がある。また、光が看板の状態に影響を与えるため、電飾の有無も考慮する必要がある。これらを厳密に定義し、収集したデータを用いて、モデルを構築することで、不要な画像の検出量を減らすことが可能になると考えられる。

データ収集とアノテーションについては自動化を行い、大幅なコストの削減を図った。しかし、機械学習モデルを構築する際のパラメータをデータの数や種類によって適宜調整しなければならない。本研究においてもこの部分については手動で調整を行っているため、完全な自動化を行うにはこのようなモデル構築時の設定

を自動化する必要がある。

8 終わりに

本研究では、食べログの外観画像ページに存在する画像データをスクレイピングで収集し、看板領域を抽出した画像を用いて構築した機械学習モデルの評価を行った。今後、7章で述べた課題を解決し、より多種多様な店舗を対象としたシステムの実現を目指す。

参考文献

- [1] 小山 雅明, 高橋 由樹, 椎塚 久雄: “そそる看板” デザインの基礎的考察, 日本感性工学会論文誌, Vol. 15, No. 1, pp. 65-73 (2016).
- [2] 北村 茂生, 松下 光範: オンサイト検索:携帯端末を用いた看板画像からの店舗情報アクセス手法, *DEIM Forum 2019*, F6-4 (2019).
- [3] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection, *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788 (2016).
- [4] Liu, S. and Deng, W.: Very deep convolutional neural network based image classification using small training sample size, *2015 3rd IAPR Asian Conference on Pattern Recognition*, pp. 730-734 (2015).
- [5] Deng, J. *et al.*: ImageNet: A large-scale hierarchical image database, *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248-255 (2009).
- [6] Haklay, M. and Weber, P.: OpenStreetMap: User-Generated Street Maps, *IEEE Pervasive Computing*, Vol. 7, No. 4, pp. 12-18 (2008).
- [7] Abadi, M. *et al.*: TensorFlow: A System for Large-Scale Machine Learning, *12th USENIX Symposium on Operating Systems Design and Implementation*, pp. 265-283 (2016).

深層学習を用いた Twitter からの趣味情報の抽出

Extraction of Interests Information from Twitter Using Deep Learning

若宮 悠希^{1*} 砂山 渡² 畑中 裕司² 小郷原 一智²
Yuki WAKAMIYA¹ Wataru SUNAYAMA² Yuji HATANAKA² Kazunori OGOHARA

¹ 滋賀県立大学大学院工学研究科

¹ Graduate School of Engineering, The University of Shiga Prefecture

² 滋賀県立大学工学部

² School of Engineering, The University of Shiga Prefecture

Abstract: In recent years, research has been actively conducted to extract personal information from Twitter. However, since comments posted on Twitter are relatively short, and comments on various topics are often made, comments on specific interests appear only locally. Therefore, in this paper, to the related vocabulary related to the designated interest be learned widely by deep learning starting from the word set related to the particular interest, whether or not each Twitter user has a designated interest determine by the proposing system.

1 はじめに

近年、インターネットの発展に伴い、Twitter¹やfacebookといったSNSなどが広く普及することで、オンライン上でのコミュニケーションの場が広く普及してきた。これにより、近しい周囲の人間にとどまらず、距離に縛られない広い範囲の人間と交流を気軽に行うことができるようになった。

これらのサービスを利用するユーザはそれぞれが違う性格や考え方、趣味、嗜好を持つため、交流を行うときは相手の持つ個性が自分のものと合うか、もしくは受け入れられるかどうか重要となる。そのため、交流相手の個性を十分に理解して受け入れる、もしくは自らが受け入れやすいユーザを選んで交流することが求められる。

交流相手と気が合うかどうかは、同一の話題について興味をもっているかを判断材料とすることができる。Twitterのような不特定多数のユーザが匿名で様々な話題について自由にコメントできるサービスにおいては、検索機能を利用して同じ趣味を持つユーザを選択して交流することもできるが、プロフィールで趣味を明言しているユーザや、サーチワードを含むコメントを投稿したユーザのみしかヒットせず、潜在的に話題に興味を持っているユーザを探すことができないこと

があるため、選択の余地が狭まる。

そのため、Twitterに投稿されたコメント集合から、ユーザが特定の趣味を持つか否かを抽出することができれば、交流相手の個性の理解や、新たな交流相手としてユーザを推薦するなど、ユーザ間の交流を手助けが行えることを期待できる。

Twitterから個人情報を抽出する研究は近年盛んに行われるようになってきているが、Twitterにおいては投稿されるコメントが比較的短く、また様々な話題についてのコメントがなされることが多いため、特定の趣味についてのコメントは局所的にしか現れてこない。

そこで本研究では、特定の趣味に関わる単語集合を起点とした深層学習 (Deep Neural Network) により、指定の趣味に関わる関連語彙を潜在的に幅広く学習させた上で、各 Twitter ユーザの投稿文であるツイートの集合から指定の趣味を持つか否かを自動抽出することを目的とする。

以下本論文では、2章で関連研究について述べる。3章で趣味情報抽出システムについて述べる。4章で提案システムの趣味情報抽出制度の評価について述べ、5章で本論文を締めくくる。

2 関連研究

Twitterに投稿されたコメントから個人情報を抽出する研究は近年盛んに行われてきている。

*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 若宮悠希

〒522-8533 滋賀県彦根市八坂町 2500
E-mail: of23ywakamiya@ec.usp.ac.jp

¹<https://twitter.com>

これまでに、各ユーザの家族構成や所有物、趣味嗜好などを抽出することで個人情報推定する研究 [1][2] が行われている。これらの研究では、投稿されたコメント (ツイート) に含まれる単語により個人の情報を抽出する手法を提案しており、「俺のギター」「私の子ども」など、一人称所有格の後に名詞が現れているツイートから、そのユーザの所有物を抽出し、「ギター」であれば「音楽」など所有物が趣味嗜好と関係のあるものならば、所有物を起点にユーザの趣味を抽出する。

また、この他に抽出対象ユーザのプロフィール文やツイートではなく、相互に交流関係のあるユーザのプロフィール文を元に属性を抽出する研究 [3] が行われている。この研究では、交流関係のある複数のユーザのプロフィール文に頻繁に出現するものを本人に深く関わりのある単語と仮定して取得することで、これらの単語を元にして対象のユーザに関わりがあると考えられる属性を推定する。

本研究では、抽出したい趣味を利用者があらかじめ決定して関連単語を与えることにより、趣味と関連する単語として幅広い語彙を網羅した学習を行うことで、推定対象のユーザのツイート集合のみから、指定の趣味を持つか否かの判断を行う。

3 深層学習を用いた Twitter ユーザからの趣味情報抽出システム

3.1 深層学習を用いた趣味情報抽出システムの構成

本研究で提案する深層学習を用いた趣味情報抽出システムの構成を図 1 に示す。

まず、システム利用者が抽出対象として決定した趣味ひとつひとつに対応する趣味情報抽出ネットワークを構築するため、各趣味の学習用データとして Twitter の投稿文であるツイートを複数収集する。収集したデータを深層学習の一種である Deep Neural Network に入力できる形に整形し、これらを用いて趣味情報抽出ネットワークを学習により構築する。次に、複数の抽出対象ユーザが投稿した各ツイート集合をシステムに入力し、構築した趣味情報抽出ネットワークにより、各趣味に興味を持っているかどうかを 1 ツweet ずつ 2 値で判定する。この判定結果を可視化インタフェースを利用してユーザごとに数値化し、提示することで、システム利用者がどのユーザが指定の趣味に興味を持っているのかを解釈する支援とする。

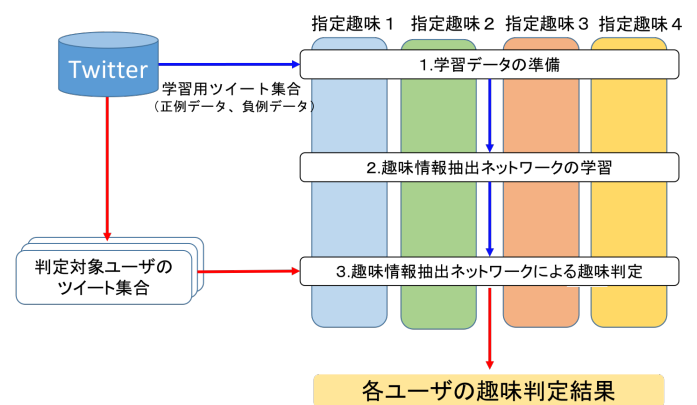


図 1: 趣味情報抽出システムの構成図

3.2 抽出対象の趣味の選定

提案システムは、深層学習により入力データのツイート集合に出現する単語の出現頻度を学習することで趣味情報を抽出するため、抽出対象の趣味として「不変性が高い趣味特有の単語集合を用意できるもの」と定義し、このうち学習用ツイート集合を十分に収集できる趣味を対象として趣味情報抽出システムを構築する。

例えば、「野球」や「将棋」などのように「ルール」「手法」「用語」などユーザによって変化する単語が変化しないもので構成されるものと、「jpop」「バラエティ番組」などのように「タレント」「アーティスト」など流行りや趣向によって大きく変化してしまう単語で構成されるものが存在する。そのため、後者の趣味を深層学習により抽出しようと試みると、一部の単語集合のみが学習されることにより、正しく抽出できるユーザに偏りが発生してしまうことから、本システムでの趣味情報抽出には不向きとなる。

3.3 深層学習に用いる学習用データの準備

3.3.1 深層学習に利用するデータ集合

本研究では、趣味情報抽出の対象を Twitter ユーザに定め、各ユーザの投稿したツイート集合を深層学習によって分析し、指定した趣味が文章中に現れているかを抽出する。Twitter ユーザの趣味情報抽出を行う際は、対象ユーザの投稿した複数のツイートを 1 件ずつ推定していくことで、全体のうち何%に趣味が現れているかを元に最終的な判断をする。

3.3.2 深層学習に利用するデータの選定

深層学習に用いる入力データは、推定対象と同じく、Twitter から収集したツイートを利用する。抽出対象

とする趣味が文章に現れているツイートを正例、対象とする趣味が文章に現れていないツイートを負例として正解ラベルをつける。これらのツイート集合を入力データとして、深層学習により趣味情報抽出ネットワークを構築する。

ここで、ツイートに指定の趣味が現れているか否かは、指定の趣味に関連のある単語を設定し、いずれかの単語が文章中に出現しているか否かを判断する基準とする。例えば、「野球」に関していえば「阪神タイガース(プロ野球チーム名)」や「ホームラン(野球用語)」などが考えられる。この単語集合を本研究では「特有単語」と定義する。

Twitter REST API²を用いてツイートを収集し、指定の趣味ごとに設定した特有単語を含むツイート、含まないツイートを選別して正例、負例を選択する。

3.3.3 特有単語の設定

特有単語は対象の趣味と関連が深いと考えられる単語となる。正しく趣味情報抽出を行うためには正例データに含まれるツイートは、対象の趣味の話題が現れているツイートのみを限定し、関係のないツイートを収集しないようにする必要がある。そのため、特有単語として設定することに望ましい単語として、その趣味特有の単語であることが求められる。

この際考えられるものの1つとして「趣味名」が挙げられるが、抽出対象趣味によっては一般的に広く知られているものとなり、収集されるツイートが趣味の話題が現れているもののみに限られない可能性がある。そのため、本システムでは趣味名を特有単語として設定することはしない。

特有単語は、抽出対象の趣味を決定する際にシステム利用者が設定する必要がある。よって、利用者による特有単語設定を補助するために TF-IDF により特有単語決定手法を提案する。

TF-IDF とは、単語の出現頻度 (Term Frequency, 以下 TF と表記) と、逆文書頻度 (Inverse Document Frequency, 以下 IDF と表記) の積で表される指標となる。これは複数異なる条件の文書集合に出現する単語に対し、各文書に対する重要度を評価する手法となる。TF は対象の単語がある 1 つの文書中にどれだけの頻度で出現したかを表し、IDF は全文書中に対象の単語が出現した文書の割合を逆数で表したものとなる。これらの積を求めることで、特定の文書にのみ高い頻度で出現する単語を抽出することができる。

TF-IDF は以下の式 (1) により求められる。ここで $tf(t, d)$ はある文書 d における単語 t の出現頻度、 $idf(t)$ は全文書中に単語 t を含む文書数、 N は全文書数となる。

$$tf - idf(t, f) = tf(t, d) \left(\log \frac{N}{df(t)} + 1 \right) \quad (1)$$

TF-IDF を用いることにより、他の趣味の話題に出現しにくい対象の趣味の特有単語を抽出し、特有単語設定の助けとする手法について説明する。まず抽出対象に設定したい趣味と、それ以外に考えられる複数の趣味について、趣味名を含むツイートを複数収集する。1 つの趣味について収集したツイートを 1 つの文書とみなし、複数趣味の文書集合に出現する全ての単語を対象に TF-IDF を算出する。そして、システム利用者の手により抽出対象の趣味に対する TF-IDF 上位の単語のうち、その趣味特有の単語だと考えられるもの 1 語を特有単語に設定する。野球を例とすると、TF-IDF 上位となった単語として「投手」「選手」「満塁」「匿名」が出現した場合、このうち「投手」「満塁」が野球特有の単語であると利用者が判断できたもののうちから、より TF-IDF 値の高い単語を特有単語として設定することが考えられる。特有単語を複数設定すると各単語と同時に出現する単語を幅広く学習することが可能となるが、一般的な単語について広く学習する可能性が高まるため、1 語に絞る。

3.3.4 学習用データの整形

深層学習の入力データや推定対象の文章は、それぞれ文章から BoW (Bag of Words) に変換して利用する。BoW とは全文章中に出現する単語を並べ、各文章での単語の出現頻度をベクトルで表現したものである。また、抽出対象の趣味特有の表現を学習したいため、使用する単語は 15 ツイート以上に出現した名詞のみとし、その中でも判定に関わらないと考えられる単語はあらかじめ除去する。除去する単語としては、「リツイート」「リプライ」など、Twitter で趣味関係なく広く使われる言葉や、特有単語の有無に関わらずツイートに出現する単語となる。特有単語を含む正例データと、特有単語を含まない約 220 万のツイート集合との間で、出現した単語全てについてカイ 2 乗検定を行う。ここで有意水準 5 % を下回る単語以外を全て除去する。そのほか、「日」「月」などの 1 文字の単語、「12」「www」などの英数字についても除去している。

3.4 深層学習による趣味情報抽出ネットワークの学習

本研究では、推定対象ユーザが指定の趣味に興味を持っているかを判断するために、深層学習による趣味情報抽出ネットワークを構築、これを元にツイート集合を分類する。

²<https://developer.twitter.com/en/docs>

表 1: 学習時の DNN のパラメータ

中間層ノード数	100
中間層数	3
ミニバッチサイズ	256
エポック数	50
L1 正則化	係数 : 0.01
活性化関数	中間層 : ReLU 出力層 : SoftMax
最適化アルゴリズム	Adam
学習率	0.1

深層学習とは、機械学習の一種で入力と出力の関係をパターンとして学習したネットワークを構築する方法である。構築されたネットワークにより、人間に与えられない細かな分類規則によって新たなデータを分類することができる。

本研究では、深層学習ライブラリの 1 つである Deep Learning for Java(DL4J)³を利用して、Deep Neural Network を扱う。

3.4.1 趣味情報抽出ネットワークの構築

指定した各趣味に対し、正例、負例の入力データを元に深層学習を行い趣味情報抽出ネットワークを構築する。入力として BoW を用いることから、入力層ノード数は学習用データ中に出現する名詞の総数となるため、抽出対象となる趣味ごとに入力層ノード数が異なる。その他の深層学習のパラメータは、全て以下のもの

で統一する。

今回、深層学習を行う上で、趣味情報抽出の結果が後述の評価において最も高い精度となったパラメータを採用した。

3.5 趣味情報抽出ネットワークを用いた Twitter ユーザの趣味判定

構築した趣味推定ネットワークを利用して対象ユーザのツイートから趣味情報抽出を行う。趣味情報抽出を行う際には、推定対象の Twitter ユーザが投稿した複数のツイートの集合を対象とし、1 ツイートを 1 データとして、文章中に趣味が現れているかどうかを 2 値で判定する。次に、趣味が強く現れていると正判定されたツイートが、全体のツイート集合に対しどの程度の割合で現れたかを算出してグラフ出力する。抽出結果を提示するための可視化インターフェースとして、統合開発環境である TETDM[4] を利用する。

³<https://deeplearning4j.org>

この時、最終的にユーザが指定の趣味に興味を持っているかどうかは、あらかじめ定めた閾値を超えているかどうかで判断する。指定する趣味の範囲や、どれだけ一般的かにより、文章からの抽出されやすさが異なるため、閾値は一般的な Twitter ユーザの判定結果よりも正判定ツイートの割合が少し高くなるように設定する。ランダムに収集した 15435 ユーザに対し上記の趣味判定を行い、各ユーザの判定結果から正判定ツイート割合の偏差値を算出し、偏差値 60 となる割合を閾値とする。

3.6 趣味情報抽出システムの使用例

提案システムを用いたユーザ推薦の例を挙げる。

システム利用者として、他のユーザとの交流の輪を広げたいユーザを設定する。この利用者は自分の趣味である「将棋」と「筋トレ」のどちらにも興味のあるユーザを検索することを目的に、この 2 つの趣味を抽出対象として提案システムに設定、無作為に収集した候補ユーザ 19 名に対し趣味情報抽出を行う場面を考える。その結果を図 2 に示す。

システムにはユーザ@A~S という 19 名の候補ユーザそれぞれのツイートが 100 件ずつ入力されており、各ユーザの判定結果が出力されている。ここで、右のパネルを確認すると将棋に興味があると判定されているユーザは 3 人、筋トレに興味があると判定されているユーザは 3 人存在している。その中で、ユーザ@R というユーザは将棋、筋トレのいずれにも興味があると判定されており、条件に当てはまるユーザとして水色のハイライトが表示されている。よって R さんを選択し中央と左のパネルにより詳細を確認する。中央のパネルにより詳細を確認すると、将棋、筋トレについて正判定されたツイートが複数存在しており、実際に将棋、筋トレについてのツイートを投稿していることがわかる。左のパネルにより詳細を確認すると、将棋、筋トレともに正判定ツイート割合が偏差値 60 を上回っている。さらに将棋は 132.57 と、Twitter ユーザの中でも特に強く興味を持っていることがわかる。よって、Twitter 上から収集した無作為な候補ユーザ群より、利用者が設定した複数の条件に当てはまるユーザをシステムにより発見することができた。

4 趣味情報抽出システムの有効性の評価

3 章で述べた提案手法の有効性を評価するため、収集したテストユーザに対して複数の趣味に対し構築した趣味情報抽出ネットワークを構築、実際に趣味情報

表 2: 評価用の各趣味における特有単語とテストユーザ数

趣味名	特有単語	正ラベルユーザ	負ラベルユーザ
ポケモン	メタモン	25	175
コーヒー	タリーズ	25	175
将棋	棒銀	25	175
ゴルフ	キャロウェイ	25	175
乃木坂 46	まいやん	25	175
ラグビー	ノーサイド	25	175
筋トレ	腕立て	25	175
プログラミング	CSS	25	175

抽出を行い、各ネットワークの趣味情報抽出精度を評価した。また、本システムを Twitter 検索機能と比較するために、Twitter 検索で探したユーザに対する趣味情報抽出の評価を行った。



図 2: 趣味情報抽出システムの使用例

4.1 趣味情報抽出精度の評価

提案システムを評価するため、抽出対象の趣味を複数設定し、各趣味に対応して構築された趣味抽出ネットワークの抽出精度をテストユーザを用いて精度指標を示す。今回の評価では、3 章で述べた趣味の定義に従い、「ポケモン」「コーヒー」「将棋」「ゴルフ」「乃木坂 46」「ラグビー」「筋トレ」「プログラミング」と 8 つの趣味を設定した。

設定した 8 つの趣味それぞれに対し、TF-IDF を活用して特有単語を筆者自身が設定、これを元に学習用データを収集して趣味抽出ネットワークの構築を行った。趣味ごとに設定した特有単語と収集したテストユーザの数を表 2 に示す。各趣味に対しプロフィールやツイートの情報から趣味を持っていると思われるユーザを収集し、これを正ラベルユーザとして設定する。一方で趣味について興味を持っていないユーザとして負ラベルユーザを用意することにより、正負を正しく判定できるかを評価する。今回は負ラベルテストユーザとして、「他 7 つの趣味の正ラベルテストユーザ」を用いる。これらのユーザは対象の趣味に興味を持っていないことを人手で確認している。各テストユーザについて最大 100 件の最新ツイートを収集し、これをシステムに入力するツイート集合とする。学習時に使用した全単語数、すなわち入力層ノード数と使用した学習用データ数は表 3 に示す。この他の学習時のパラメータは全ての趣味で 3 章で述べた表 1 の値を用いている。

各テストユーザを提案システムに入力して趣味抽出を行い、正判定に対する Precision, Recall, F 値により各趣味に対応する趣味抽出ネットワークの抽出精度を示す。ここで、本章において正判定とは「対象の趣味についての対象ユーザの正判定ツイート割合が閾値

表 3: 評価用の各趣味における学習用データ数と入力層ノード数

趣味名	学習用データ数	入力層ノード数
ポケモン	18528(正例 9264)	695
コーヒー	11528(正例 5764)	441
将棋	470(正例 235)	18
ゴルフ	284(正例 142)	13
乃木坂 46	17616(正例 8808)	542
ラグビー	9430(正例 4715)	313
筋トレ	10932(正例 5466)	369
プログラミング	11098(正例 5549)	493

である偏差値 60 を超えているという判定」，負判定は反対に「正判定ツイート割合が閾値を下回ったという判定」と定義する．また，正判定に対する Precision はシステムが正判定したユーザのうちの正ラベルユーザの割合，Recall は正ラベルユーザのうちの実際にシステムが正判定できたユーザの割合，F 値は Precision と Recall の調和平均となる．正判定に対するこれらの精度指標は式 2，式 3，式 4 により表すことができる．この時， TP は正ラベルユーザを正判定できた数とされる真陽性ユーザの数， FP は負ラベルユーザを誤って正判定してしまった数とされる偽陽性ユーザの数， FN は正ラベルユーザを誤って負判定してしまった数とされる偽陰性ユーザの数となる．

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F \text{ 値} = \frac{2Precision * Recall}{Precision + Recall} \quad (4)$$

表 4: 評価用の各趣味における趣味抽出ネットワークの評価結果

趣味	Precision	Recall	F 値
ポケモン	0.706	0.960	0.814
コーヒー	0.821	0.920	0.868
将棋	0.690	0.800	0.741
ゴルフ	0.821	0.920	0.868
乃木坂 46	0.864	0.760	0.809
ラグビー	0.686	0.960	0.800
筋トレ	0.545	1.000	0.704
プログラミング	0.455	0.800	0.580

評価用の各趣味に対する趣味抽出ネットワークの評価結果を表 4 に示す．

各趣味について用意したテストユーザの正判定に対する Precision, Recall, F 値を確認すると，全ての趣味で Recall が 0.7 から 0.9 と高い水準を見せていることに對し，Precision は 0.7 から 0.8 と高い数値を見せているものから，0.3 から 0.6 とあまり高いとは言えない結果となった趣味が存在している．高い Recall を見せていることからわかる通り，各趣味は正ラベルユーザの多くを正しく正判定することができている．その一方でほぼ同数以上の負ラベルユーザを正判定してしまい，偽陽性が多く出現している趣味があることがわかる．

提案システムでは，DNN により趣味ごとに判定に関わる重要な単語や，単語の組み合わせを学習し，未知

ツイートが入力された際は学習された単語やその組み合わせが出現しているか否かにより正判定，負判定を行う．そのため，学習時にどの単語が判定に重要であると学習されたのかを理解することで，結果考察の助けとする．安藤らの研究 [5] では，BoW を用いたテキストベースの DNN において分類パターン上での重要度の高い単語，すなわち重要単語の解釈を行える枠組みを開発している．これを用いて各趣味における趣味抽出ネットワークの正判定に対する重要単語を出力し，各趣味の上位 20 単語を表 5 に示す．

表 5 には評価に使用した 8 つの趣味の重要単語がそれぞれ示されており，各趣味に強く関連する単語が上位 20 単語として出現している．

これらの中にも，ポケモンの「個体」，コーヒーの「当方」，将棋の「相手」など，表中でアンダーラインを引いている単語は一般的で広く使われると考えられ，現在の提案システムでも学習データから一般語を除去しきることはできていないことがわかる．これら一般語が学習時に残っていることが偽陽性が発生する要因の 1 つと考えられる．

ラグビーの重要単語として，「紅白歌合戦」や「きのこ」「たけのこ」など，一見して関係がなさそうな単語が含まれている．これらの単語は，ラグビーを題材としたドラマの主題歌が紅白歌合戦で歌われたことや，Twitter にて「#きのこたけのこ国民総選挙ノースサイド」というハッシュタグが流行していた時期に正例ラベルデータを収集したことが関係していると考えられる．また，プログラミングの Precision が他趣味に比べ低い結果となっている．誤って正判定としてしまった負ラベルユーザを確認したところ，プログラミングでは「初心者」「ルール」「レッスン」「筋トレ」などの多趣味に関連する単語が重要単語として学習されていた．そのため，これらを含む将棋，ゴルフ，筋トレの正ラベルユーザを誤って正判定してしまっていた．提案システム単体では趣味抽出の学習の解釈はできず，システム利用者はどのような基準で正判定が行われたかを理解することが難しい．そのため，安藤らのシステムと連携して判定の際にツイートに含まれていた表 5 の重要単語を抽出し，可視化インタフェースで出力することにより解釈の助けとすることが考えられる．

4.2 Twitter 検索により探索したユーザに対する趣味情報抽出の評価

本システムによる趣味情報抽出を用いることで指定趣味に興味を持つユーザ推薦を有効に行えるかを示すため，Twitter 検索によるユーザ探索を比較とした評価を行った．

用意したテストユーザを用いた評価では，一般語や他

表 5: 評価用の各趣味における趣味抽出ネットワークの重要単語上位 20 語

ポケモン 海外 色違い ポケモン剣盾 オシャボ 個体値 レイド 個体 国産 ザシアン 特性 マスボ 図鑑 性格 外国産 臆病 コゴガラ 火炎 王冠 ミュウ	コーヒー コーヒー ドトール ラテ ロイヤルミルクティー 付近 チョコ ミルクティー 当方 缶コーヒー タピオカ ジャーボテドッグ 募集中 珈琲 ドリンク スノーマンラテ ベア 抹茶 ハニーミルクラテ 真斗 ホットドッグ	将棋 相手 将棋 矢倉 斜め 嬉野流 振り 戦法 四間 腰掛け 将棋ウォーズ 三間 動画 石玉 定跡 後手 原始 棋譜 先生	ゴルフ ゴルフ アイアン メンズ ベスト マーベリック クラブ ドライバー セット モデル ボール シャフト 矢野 ゴルフシューズ
乃木坂 46 レーン 乃木坂 シンクロニシティ 飛鳥 紅白 バスラ 白石麻衣 真夏 乃木 与田 みなみ ななみん 白石 桃子 まちゅ ガルル 生駒 募集中 迷惑メール 秋元真夏	ラグビー ユーミン 紅白 松任谷由実 ラグビー きのこ 紅白歌合戦 ゲーム たけのこ きのこの山 麗美 たけのこの里 党首 キャンペーン 戦い ラグビー日本代表 米津玄師 日本代表 名曲 大晦日 総選挙	筋トレ 腹筋 筋トレ プランク 背筋 セット 筋肉 回数 トレ 体幹 体重 積み上げ 逆立ち キロ ブリッジ クランチ アブローラー パカ 大胸筋 ランニング プッシュアップバー	プログラミング コース サイト コーディング エンジニア 積み上げ 初級 中級 レベル フレームワーク 案件 基礎 アニメーション 要素 ドットインストール 初心者 プロパティ レスポンス 上級 ウェブカツ 言語

趣味関連語を除去できていないことにより、Precision が低い結果となってしまった趣味がある。そのうちの 1 つであるコーヒーと、高い Precision を見せたポケモンについて、実際にシステムを活用する際に想定する方法で収集した候補ユーザに対して趣味情報抽出を正しく行えるのかを確かめた。

趣味情報抽出を行う際、提案システムでは抽出対象として対象の趣味名を含むツイートを投稿したユーザを収集し、このうちから趣味に興味を持つユーザを抽出することを想定している。よって「ポケモン」「コーヒー」をそれぞれ含むツイートを投稿したユーザを収集し、正ラベルユーザ、負ラベルユーザに分類したものをテストユーザとした。趣味名で収集したユーザはポケモンはうち 7 割、コーヒーはうち 3 割が正ラベル

となった。ポケモンなど娯楽となりやすい趣味よりも、コーヒーなどのより生活習慣に近い趣味は多くのユーザの話題として出現しやすいためであると考えられる。

趣味情報抽出の結果として正判定に対する Precision は、ポケモンが 0.886、コーヒーが 0.667 となった。ポケモンは高い Precision をみせたが、コーヒーは一般語や他趣味関連語による誤判定が多く見られ、正判定のうち 3 割程度は誤判定となってしまった。一方で、趣味情報抽出の結果の Precision は、趣味名を元に検索し収集したユーザのうちの正ラベルユーザの割合を上回っているため、Twitter ワードサーチによる単純なユーザ探索と比較して、そのうちから提案システムによりユーザを絞り込むことで趣味に興味を持つユーザをより高い割合で見つけることができると考えられる。

表 6: 趣味名を含むツイートを元に収集したテストユーザ数

趣味名	テストユーザ総数	正ラベルユーザ	負ラベルユーザ
ポケモン	145	105	40
コーヒー	146	51	95

表 7: 趣味名を含むツイートを投稿したテストユーザの趣味情報抽出評価結果

ポケモン				コーヒー			
		正解ラベル				正解ラベル	
		正	負			正	負
予測	正	70	9	予測	正	10	5
結果	負	35	31	結果	負	41	90
Precision		0.886		Precision		0.667	
Recall		0.667		Recall		0.200	
F 値		0.760		F 値		0.300	

以上の結果より、提案システムでは多様な単語表現を学習することによって高い Recall を示せたことにより対象の趣味に興味のあるユーザを網羅して抽出することができているが、一方で一般的な単語や他趣味に関連する単語を学習してしまっているため、偽陽性ユーザが発生してしまうことがわかった。趣味名を含むツイートをを行うユーザを収集し、これを元に趣味情報抽出を行った結果、Twitter ワードサーチと比較して交流ユーザ探索により有用であることを示せたが、入力する候補ユーザ群によっては多すぎる偽陽性が出現してしまうと使用例として想定するユーザ推薦にも支障をきたしてしまうことが予想される。そのため、偽陽性の原因となる一般的な単語や他の趣味に関連する語などを TF-IDF 等を利用して抽出、除去し、抽出精度を上げていくことが今後の課題となる。

5 おわりに

本研究では、Twitter ユーザを対象として趣味抽出を行い、システム利用者が設定した任意の趣味についての抽出結果を提示することを目的として行った。任意の趣味についての学習用データを Twitter より収集し、テキストベースの深層学習を用いて抽出対象ユーザの投稿文の集合より対象の趣味に興味を持っているかどうかを判定し、結果を提示するシステムを開発した。

提案システムの趣味抽出精度の評価として、設定した 8 つの趣味について用意したテストユーザに対する Precision, Recall を求めた。その結果、現在のシステムでは高い Recall を示せたことにより対象の趣味に興味

のあるユーザを網羅して抽出することができ、Twitter ワードサーチを用いた単純なユーザ探索に比べより有用であることを示せたが、一方で一般的な単語や他趣味に関連する単語を学習していることにより偽陽性が発生し Precision が低下することがわかった。

今後の課題として、これら判定に悪影響を及ぼす単語を除去し、抽出精度を上げていくことを目標としていきたい。

参考文献

- [1] 馬縹美穂, 徳久良子, 寺寫 立太: ユーザの嗜好と所有物の関係性を用いた属性分析研究報告情報基礎とアクセス技術 2014-IFAT-114, pp.1-6 (2014)
- [2] 那須川哲也, 西山莉紗, 金山博, 吉田一星, 大野正樹: 一人称所有格を用いたプロフィール推定, 言語処理学会第 19 回年次大会発表論文集, pp.952-955 (2013)
- [3] 上里 和也, 浅井 洋樹, 山名 早人: Personalized PageRank を利用した網羅的 Twitter ユーザ属性推定, DEIM 2016 第 8 回データ工学と情報マネジメントに関するフォーラム D2-2 (2016)
- [4] 砂山渡, 高間康史, 徳永秀和, 串間宗夫, 西村和則, 松下光範, 北村侑也: 統合環境 TETDM を用いた社会実践, 人工知能学会論文誌 32 巻 1 号, pp.NFC-A.1-12(2017)
- [5] 安藤 雅行, 河原 吉伸, 砂山 渡, 畑中 裕司: テキストベースの深層学習における分類パターンの解釈支援, 知能と情報 Vol.31, No.4, pp.779-787 (2019)

ベクトル単位での重み共有化と符号化による カプセルネットワーク軽量化の試み

Experimental study on the vector-based compression of Capsule Networks

島田 研太¹ 尾崎 知伸^{1*}
Kenta Shimada¹ Tomonobu Ozaki¹

¹ 日本大学 文理学部

¹ College of Humanities and Sciences, Nihon University

Abstract: In this study, we discuss the compression of the capsule networks (CapsNets), which is one of extensions of the convolutional neural networks (CNNs). Focusing on the fact that the basic unit of CapsNets is a vector, we propose a vector-based quantization and Huffman coding for the compression by extending typical weight reduction methods for CNN. Experimental evaluations using MNIST and CIFAR-10 with various vector sizes and quantization granularities show that the proposed method can compression the networks with a slight loss of prediction accuracy.

1 はじめに

深層学習の代表的な手法の一つである畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) は、特徴抽出を行う畳み込み層と次元圧縮を行うプーリング層とを複数接続させることで、画像認識分野において非常に高い精度を実現している。しかしその一方で、少なくとも2つの側面から問題点が指摘されている。

第一の問題点は、モデル容量に関するものである。一般に CNN は、膨大な数のパラメタを含み、結果としてモデルの容量が巨大になる傾向が認められる。これに対し、これまで様々な方面で改善策が提案されている。例えば文献 [1] では、訓練済みモデルから別の軽量モデルを構築する蒸留と呼ばれる技術が提案されている。また、絶対値が小さな重みをもつ連結辺を削除する枝刈り [2, 3] や、類似する重みをグループ化し代表値に縮約する重み共有とその符号化 [4] に関してもそれぞれ提案が行われており、これらの技術を組み合わせることで、精度を劣化させずに大幅なモデル圧縮が達成されることが報告されている [5]。

CNN の第二の問題点は、精度に関するものであり、これまでプーリング層に起因する問題が指摘されている [6]。プーリング層は、畳み込み層で抽出された特徴を一定範囲で集約・集計することで位置や大きさに対する頑健性を CNN に与えている。しかしその反面、プー

リングの結果として、抽出された特徴の向きや大きさ、特徴間の位置関係等の情報が捨てられることとなり、精度向上に対して負の影響を与えることが考えられる。この問題に対し、プーリング層を排除し、カプセルと呼ばれる新たなニューロングループを一つの単位として利用するカプセルネットワーク (Capsule Networks; CapsNet) [6] と呼ばれる手法が提案されている。詳しくは後述するが、CNN など従来の深層学習モデルでは、各ニューロンはパラメタとして重みベクトルを持ち、一つのスカラー値を出力するが、CapsNet では、各ニューロンはパラメタとして重み行列群を持ち、ベクトル値を出力する。そのため、必然的にモデル (パラメタ数) が大規模化し、大量の演算リソース (計算時間とメモリ) が必要とされることが予想される。

本論文では、CNN の2つの問題点、すなわちモデルの大規模化とプーリング層に起因する精度劣化を同時に解消することを目指し、CapsNet に対する軽量化手法を模索する。具体的には、CNN に対する代表的な軽量化手法である枝刈り、重み共有化、符号化を基本とするが、スカラー出力をベクトル出力に変更したという CapsNet の大きな特徴を考慮し、各技術の適用単位をスカラーではなく、ベクトルとすることを提案する。

本論文の構成は以下の通りである。2章でカプセルネットワークの概要を示す。3章で代表的な軽量化手法について概観し、4章でそれらをカプセルネットワークへと応用する。5章で実験と評価を行い、最後に6章でまとめを行い今後の課題を述べる。

*連絡先: 日本大学文理学部情報科学科
〒156-8550 東京都世田谷区桜上水 3-25-40
E-mail: tozaki@chs.nihon-u.ac.jp

2 カプセルネットワーク

カプセルネットワーク [6] とは、カプセルと呼ばれる、特定タイプの実体に関する（大きさや向きなどの）パラメタを表す活性化ベクトルを持つニューロンのグループを基本単位とするニューラルネットワークである。活性化ベクトルは、その名の通り実数値ベクトルであり、ベクトル長を用いて実体の存在確率を、向きを用いてパラメタ値を表す。カプセルネットワークは、複数のカプセルを連結させたネットワークであり、カプセル間でベクトル表現される情報を伝播することで学習と予測が行われる。すなわち、あるレベルで活性化されたカプセルがより高いレベルのカプセルのパラメタに対して予測を行い、また複数カプセルによる予測が一致すると、より高いレベルでのカプセルが活性化されることになる。非常に大雑把に言えば、カプセルネットワークとは、CNN からプーリング層を排除し、伝播させる情報をスカラー値からベクトルに変更したものと言える。

カプセル j の出力であるベクトル v_j は、その長さをを用いて入力値に対する (j が表す) 実体の存在確率を表す。このため、長さの短いベクトルはその長さを 0 に縮約し、逆に長いベクトルはその長さを 1 に近づけるような非線形変換を行う Squashing 関数と呼ばれる活性化関数が提案されている。

Squashing 関数 squash は、カプセル j への入力 s_j を引数とし、活性化ベクトル v_j を出力する関数であり、以下の様に定義される [6]。

$$v_j = \text{squash}(s_j) = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|}$$

ここで squash への入力である s_j は、下位層のカプセル i の出力ベクトル $\mathbf{u}_i$ と重み行列 $\mathbf{W}_{ij}$ の積として得られる予測ベクトル $\hat{\mathbf{u}}_{j|i}$ の重み付き和であり、以下の様に定義される。

$$s_j = \sum_i c_{ij} \hat{\mathbf{u}}_{j|i}, \quad \hat{\mathbf{u}}_{j|i} = \mathbf{W}_{ij} \mathbf{u}_i$$

また結合係数（結合強度）と呼ばれる重み

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}$$

は、カプセル i から j への（正規化済みの）結合の強さを表し、ルーティングと呼ばれる下記の計算を反復することで学習される。

$$\begin{cases} \mathbf{c}_i \leftarrow \text{softmax}(\mathbf{b}_i) \\ \mathbf{s}_j \leftarrow \sum_i c_{ij} \hat{\mathbf{u}}_{j|i} \\ \mathbf{v}_k \leftarrow \text{squash}(\mathbf{s}_j) \\ b_{ij} \leftarrow b_{ij} + \hat{\mathbf{u}}_{j|i} \cdot \mathbf{v}_j \end{cases}$$

3 ニューラルネットワークの軽量化

ニューラルネットワークに対する汎用的な圧縮・軽量化の一つの手法として、これまでに、重みの枝刈りと共有、ハフマン符号化を用いる手法が提案されている [4]。

枝刈り： ニューラルネットワーク中の各ニューロンは一般に、下位層から連結されているニューロンの出力に対する重み付き和を入力とし、何らかの演算を行う。従って、重み（の絶対値）の大きさがその影響度に相当する。枝刈りは、絶対値が 0 に近い重みを有する接続辺は影響が小さいものとして削除する操作を指す。より形式的には、閾値 ϵ を用いて、重み w の値を

$$w' = \begin{cases} 0 & (|w| \leq \epsilon) \\ w & (\text{otherwise}) \end{cases}$$

と更新し、 $w' = 0$ に相当する辺を削除する。また、枝刈り操作と枝刈り後のネットワークに対する再学習を重みが収束するまで繰り返し、精度劣化を抑制した上での不要接続辺の排除を実現している。

重み共有： 重み共有とは、訓練済みのネットワークの各層において類似する重みをグループ化し、各重みを所属グループの代表値に置き換える操作を指す。この操作により、重みの数（種類数）がグループ数にまで削減されることとなるので、ネットワーク全体の軽量化が期待できる。

具体的には、K-平均法を用いて重みの集合 $W = \{w_1, w_2, \dots, w_n\}$ を

$$\arg \min_C \sum_{c_i \in C} \sum_{w \in c_i} |w - c_i^*|$$

を満たす k 個のクラスター $C = \{c_1, c_2, \dots, c_k\} (k \ll n)$ にグループ化し、各重み w を所属するクラスターの代表値 c_i^* に置き換える、すなわち

$$w' = c_i^* \text{ s.t. } c_i = \arg \min_{c_i \in C} |w - c_i^*|$$

と w を更新する操作を行う。また、クラスター c_i の代表値としては、クラスター内平均 $c_i^* = \frac{1}{|c_i|} \sum_{w \in c_i} w$ や、セントロイド $c_i^* = \arg \min_{w \in c_i} \sum_{w_j \in c_i} |w - w_j|$ などが利用される。

ハフマン符号化： 各重み w をビットで符号化することで、モデル全体の圧縮を行う。具体的には、重み共有後にハフマン符号化を適用する。ハフマン符号化により、頻度の高い情報に短いビットが、頻度の低い情報に長いビットが割り当てられ、モデル全体としての圧縮が期待できる。

4 カプセルネットワークの軽量化

2章で概観した通り、カプセルネットワークでは、2つのカプセル i と j の連結に対し、結合強度を表す c_{ij} に加えて重み行列 $\mathbf{W}_{ij}$ が付与される。そのため一般的なニューラルネットワークと比較し、パラメタ数(重み総数)が膨大となる。特に c_{ij} がスカラーであるのに対し、 $\mathbf{W}_{ij}$ は行列であり、含まれるパラメタ数も大きい。従って、カプセルネットワークの圧縮においては、重み行列の圧縮が重要となる。

本研究では、重み行列に対し、枝刈り処理(絶対値が0に近い重みの値を0にする処理)を施した後、要素、行、列のそれぞれの単位での重み共有化とハフマン符号化を適用することを提案し、各単位での効果を実験的に比較する。

対象とする重み行列の全体集合を $\mathcal{W}$ とし、各重み行列 $\mathbf{W}_{ij} \in \mathcal{W}$ を

$$\mathbf{W}_{ij} = \begin{pmatrix} w_{ij}^{1,1} & \cdots & w_{ij}^{1,n} \\ \vdots & \ddots & \vdots \\ w_{ij}^{m,1} & \cdots & w_{ij}^{m,n} \end{pmatrix}$$

と表記する。このとき、行列に含まれる要素、行ベクトル、列ベクトルの全体集合をそれぞれ

$$\mathcal{W}^{r,c} = \{w_{ij}^{k,l} \mid 1 \leq k \leq m, 1 \leq l \leq n, \mathbf{W}_{ij} \in \mathcal{W}\}$$

$$\mathcal{W}^{r,*} = \{\langle w_{ij}^{k,1} \cdots w_{ij}^{k,n} \rangle \mid 1 \leq k \leq m, \mathbf{W}_{ij} \in \mathcal{W}\}$$

$$\mathcal{W}^{*,c} = \{\langle w_{ij}^{1,l} \cdots w_{ij}^{m,l} \rangle \mid 1 \leq l \leq n, \mathbf{W}_{ij} \in \mathcal{W}\}$$

と表記する。各集合 $\mathcal{W}^{r,c}$ 、 $\mathcal{W}^{r,*}$ 、 $\mathcal{W}^{*,c}$ を重み集合と捉え、それぞれクラスタリングを通じた重み共有とハフマン符号化を適用する。なお、要素を単位とする場合は3章で示した一般的な圧縮方法と等価となる。また、行ベクトルまたは列ベクトルを単位とすることで、複数の重みが一塊として扱われることとなり、より一層の圧縮が期待できる。

5 評価実験

5.1 データセットと実験設定

下記に示す2つのデータセットを対象にモデルを構築し、実験を行った。

MNIST: MNIST データセット¹は、一桁の手書き数字画像に関するデータセット(データ数70,000, クラス数10)である。実験では、60,000事例を訓練例とし、

¹<http://yann.lecun.com/exdb/mnist/>

表 1: 実験結果: モデル容量 (単位は MByte)

k	共有化			符号化		
	$\mathcal{W}^{r,c}$	$\mathcal{W}^{r,*}$	$\mathcal{W}^{*,c}$	$\mathcal{W}^{r,c}$	$\mathcal{W}^{r,*}$	$\mathcal{W}^{*,c}$
MNIST: 枝刈り前 50.78, 枝刈り後 44.63						
32	40.81	41.37	41.55	14.47	11.11	11.10
64	40.90	41.55	41.60	14.55	11.20	11.20
128	40.89	41.63	41.92	14.82	11.22	11.21
CIFAR-10: 枝刈り前 99.89, 枝刈り後 99.20						
32	90.20	92.14	92.22	88.45	87.46	87.45
64	90.50	92.40	92.45	88.56	87.50	87.45
128	90.75	92.88	92.93	88.90	87.58	87.49

10,000事例をテスト例として利用した。また文献[6]と同様に、畳み込み層、畳み込み層+カプセルネットワーク、全結合層+カプセルネットワークの3部でモデルを構成した。なお、カプセル間の重み行列は8行×16列である。

CIFAR-10: CIFAR-10 データセット²は、物体認識を意図した画像データセット(データ数60,000, クラス数10)である。実験では、50,000事例を訓練例とし、10,000事例をテスト例として利用した。またモデルは、VGG-16[7]を基本とし、全結合層3層をカプセルネットワーク2層で置き換えた構成とした。MNIST データの場合と同様、カプセル間の重み行列は8行×16列である。

5.2 実験結果と考察

提案する軽量化手法の効果を確認するため、枝刈りに関する閾値を $\epsilon = 0.25 \times$ 重み標準偏差、重み共有化のためのクラス数を $k \in \{32, 64, 128\}$ として実験を行い、圧縮手法適用前後での識別精度およびモデル容量を計測した。なおモデル容量は、圧縮対象としたカプセルネットワーク中の重み行列($\mathbf{W}_{ij}$)と結合係数(c_{ij})のみを対象に計測を行っている。また各パラメタ設定毎に、訓練例とテスト例を変化させながら10回の試行を行い、その平均を算出した。実験結果を表1および表2に示す。両表において、 $\mathcal{W}^{r,c}$ 、 $\mathcal{W}^{r,*}$ 、 $\mathcal{W}^{*,c}$ はそれぞれ要素、行、列を単位として軽量化手法を適用した場合を表す。

実験結果より、両データセットにおいても、クラス数 k の値が大きいほど、符号化時の圧縮率が高いことが分かる。また条件が同じ場合、 $\mathcal{W}^{r,c}$ と比較して $\mathcal{W}^{r,*}$ と $\mathcal{W}^{*,c}$ の方が圧縮効率が高く、提案した行、列を単位とする軽量化が有効に働いていることが確認できる。一方で、精度に関しては条件が同じ場合に $\mathcal{W}^{r,c}$ が優れ

²<https://www.cs.toronto.edu/~kriz/cifar.html>

表 2: 実験結果：精度（単位は%）

k	クラスタ内平均			セントロイド		
	$\mathcal{W}^{r,c}$	$\mathcal{W}^{r,*}$	$\mathcal{W}^{*,c}$	$\mathcal{W}^{r,c}$	$\mathcal{W}^{r,*}$	$\mathcal{W}^{*,c}$
MNIST：枝刈り前 99.65, 枝刈り後 99.62						
32	82.25	81.25	70.58	99.53	89.45	78.81
64	88.54	89.78	87.86	99.54	97.37	90.31
128	88.45	85.54	84.41	99.53	98.78	96.13
CIFAR-10：枝刈り前 63.26, 枝刈り後 64.36						
32	53.54	53.21	49.54	64.33	55.57	44.71
64	45.54	50.05	42.58	64.41	53.34	49.27
128	56.87	53.85	45.89	64.38	56.61	53.88

ており、精度とモデル圧縮との間にトレードオフがあることが分かる。また、重み共有化における代表値として、クラスタ内平均を用いる場合とセントロイドを用いる場合とを比較すると、すべての場合においてセントロイドを用いた場合が精度が高いことが確認できる。これは、平均値は外れ値に影響を受けやすく、結果として精度が低下したからではないかと考察できる。

実験結果からも、モデルの圧縮率と精度にはトレードオフの関係が認められる。ここで、両者のバランスを考慮した手法の評価を行うため、容量削減率と精度保持率を導入する。枝刈り適用前のモデルサイズを M_{org} 、軽量化手法 x 適用後のモデルサイズを M_x と表記する。一方、枝刈り適用前の精度を A_{org} 、 x 適用後の精度を A_x と表記する。このとき、 x による容量削減率 MRR_x と精度保持率 ARR_x をそれぞれ

$$MRR_x = 1 - M_x/M_{org} \quad ARR_x = A_x/A_{org}$$

と定義する。

表 3 に各手法による容量削減率と精度保持率、両者の調和平均とその順位を示す。また図 1 に、容量削減率と精度保持率とのプロット図を示す。

表 3 より、容量削減率と精度保持率の調和平均に着目すると、提案した行、もしくは列単位での重み共有化と符号化が有効に働いていることが分かる。また図 1 より、提案手法で得られる結果の多くが極大である、すなわち容量削減率と精度保持率の両方が他の手法より劣ることはないということが確認できる。

6 まとめと今後の課題

本研究では、カプセルネットワークを対象とした軽量化の一手法として、ベクトル単位での重み共有化と符号化を提案した。また 2 つの実データを用いた実験により、その効果を確認した。

表 3: 容量削減率と精度保持率（単位は%）

k	MMR	ARR	F1	Rank	
MNIST					
$\mathcal{W}^{r,c}$	32	71.50	99.88	83.34	6
	64	71.35	99.89	83.24	7
	128	70.82	99.88	82.87	8
$\mathcal{W}^{r,*}$	32	78.12	89.76	83.54	5
	64	77.94	97.71	86.72	2
	128	77.90	99.13	87.24	1
$\mathcal{W}^{*,c}$	32	78.14	79.09	78.61	9
	64	77.94	90.63	83.81	4
	128	77.92	96.47	86.21	3

CIFAR-10					
$\mathcal{W}^{r,c}$	32	11.45	101.69	20.59	7
	64	11.34	101.82	20.41	8
	128	11.00	101.77	19.86	9
$\mathcal{W}^{r,*}$	32	12.44	87.84	21.80	1
	64	12.40	84.32	21.63	4
	128	12.32	89.49	21.66	3
$\mathcal{W}^{*,c}$	32	12.45	70.68	21.18	6
	64	12.45	77.88	21.47	5
	128	12.41	85.17	21.67	2

今後の課題としては、より大規模かつ多様なデータを用いた提案手法の評価や、重み共有のためのクラスタリング手法と代表値算出手法の改善に加え、拡張されたカプセルネットワーク（例えば [8, 9, 10]）への対応などが挙げられる。

謝辞： 本研究の一部は、JSPS 科研費 17K00315 の助成を受けたものです。

参考文献

- [1] G. Hinton, O. Vinyals, and J. Dean : Distilling the knowledge in a neural network, *NIPS 2014 Deep Learning Workshop*, pp.1–9, 2014.
- [2] Y. Cheng, F. X. Yu, R. S. Feris, S. Kumar, A. Choudhary, and S. Chang : An Exploration of Parameter Redundancy in Deep Networks with Circulant Projections, *Proc. of the 2015 IEEE International Conference on Computer Vision*, pp. 2857–2865, 2015.
- [3] S. Han, J. Pool, J. Tran, and W. J. Dally : Learning both weights and connections for efficient neural networks, *Proc. of the 28th Inter-*

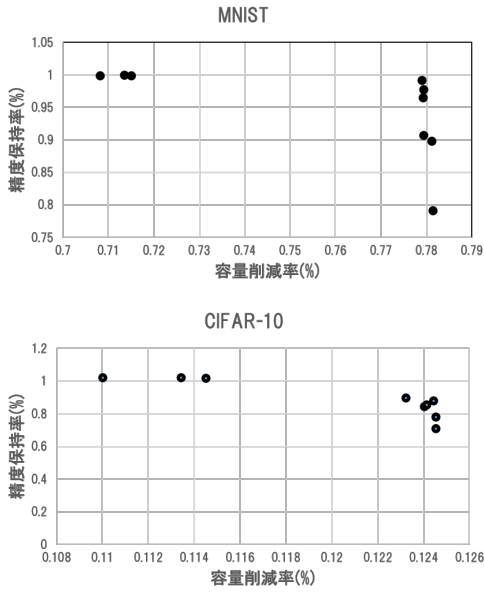


図 1: 削減率と精度のプロット図

national Conference on Neural Information Processing Systems, Vol.1, pp.1135-1143, 2015.

- [4] Y. Gond, L. Liu, M. Yang, and L. Bourdev : Compressing deep convolutional networks using vector quantization, *International Conference on Learning Representations 2015*, 2015.
- [5] S. Han, H. Mao, W. J. Dally : Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding, *International Conference on Learning Representations 2016*, 2016.
- [6] S. Sabour, N. Frosst, G. E. Hinton : Dynamic Routing Between Capsules, *Advances in Neural Information Processing Systems 30*, pp.3856-3866, 2017.
- [7] K. Simonyan and A. Zisserman : Very Deep Convolutional Networks for Large-Scale Image Recognition, *International Conference on Learning Representations 2015*, 2015.
- [8] T. Jeong, Y. Lee, and H. Kim : Ladder Capsule Network, *Proc. of Machine Learning Research*, Vol.97, pp.3071-3079, 2019.
- [9] S. Verma and Z.-L. Zhang : Graph Capsule Convolutional Neural Networks, *Proc. of Joint ICML and IJCAI Workshop on Computational Biology*, 2018.

- [10] Z. Xinyi and L. Chen : Capsule Graph Neural Network, *International Conference on Learning Representations 2019*, 2019.

グラフ構造データを対象とした解釈可能決定集合の拡張

Interpretable Decision Sets for Graph Structured Databases

松山 航太¹ 尾崎 知伸^{1*}
Kota Matsuyama¹ Tomonobu Ozaki¹

¹ 日本大学 文理学部

¹ College of Humanities and Sciences, Nihon University

Abstract: Interpretable Decision Sets (IDS) is one of representative algorithms for deriving classification rules having high interpretability while keeping high accuracy. In this paper, we propose an extension of IDS for handling graph structured data. In the proposal, after converting a graph database into a set of transactions using frequent subgraph mining, classification rules will be extracted based on our developed objective functions on interpretability of subgraph patterns.

1 はじめに

深層学習を中心とした機械学習技術の発展に伴い、近年、非常に高精度な機械学習モデルが構築されている。しかし機械学習モデルの多くは、利用者がその内容を理解・解釈することが困難な大規模かつ複雑なブラックボックスであり、モデルのデバッグ・更新が容易ではなく、また安全面や倫理面からもその応用範囲が限定されてしまうことも考えられる。これらの問題を軽減するため、モデル理解の支援や解釈可能（容易）なモデルの構築など、機械学習の解釈性が着目され、多数の研究が行われている [1, 2]。本研究では、分類規則抽出に関する代表的な解釈性研究の一つである解釈可能決定集合（Interpretable Decision Sets; IDS）[3]に着目する。IDS は、データから導出されるクラス相関ルール集合から、解釈性を考慮した評価関数を最大化する部分集合を抽出することで、分類精度を維持したまま解釈性の高い少数ルール群を特定する。本研究では、IDS のアイディアをグラフ構造データへと拡張する。具体的には、頻出部分グラフ発見技術 [4] を用いてグラフデータベースをトランザクション化した上で、クラス相関ルール [5] を導出する。加えて、クラス相関ルールの構成要素が部分グラフパターンであることに着目し、グラフ形状を考慮した解釈性に関する新たな評価尺度をルール集合選択基準に組み込む。これらにより、グラフ構造データからの解釈容易な分類規則集合の導出を目指す。

本論文の構成は以下の通りである。2 章で関連研究について言及する。3 章で IDS の要点を整理する。4 章で、グラフ構造データからのクラス相関ルールの導出

およびグラフ構造を考慮したルール集合選択基準を導入し、IDS をグラフ構造データへと拡張する。5 章で評価実験を行い、最後に 6 章でまとめと今後の課題を述べる。

2 関連研究

近年、多様な側面から機械学習の解釈性に関する研究が活発に行われている [1, 2]。これらの研究は、入力事例に対する判断根拠を示すことを目的とした局所的説明に関する研究や、モデル全体を別の解釈容易なモデルで近似する大域的説明に関する研究などに大別することができる。本研究で対象とする IDS は、データから直接可読性の高い解釈可能モデルを構築する手法と認識されるが、実際にはクラス相関ルール集合から決定集合を抽出しており、モデルを経由する大域的説明に関する手法と関係が深い。

IDS と関連し、これまで、決定木アンサンブルとして表現されるモデルから少数のルールを抽出する手法が提案されている [6, 7]。なお、決定木中の根から各葉までのパスは一つのルールに対応するので、決定木アンサンブルからのルール抽出は、大規模ルール集合からルール抽出に相当する。例えば inTrees[6] は、決定木アンサンブルをルール集合に変換した後、不要な条件の削除と複数ルールの集約を行い、集合被覆アルゴリズムを用いて少数ルールを選択する。また defragTrees[7] は、分類器アンサンブルを確率的生成モデルとしてモデル化し、EM アルゴリズムやベイズ推定を用いた最適化を通じて少数ルールを導出する。

解釈性研究のグラフ構造データへの応用に関しては、文献 [8] において、グラフ構造データをトランザクショ

*連絡先：日本大学文理学部情報科学科
〒156-8550 東京都世田谷区桜上水 3-25-40
E-mail: tozaki@chs.nihon-u.ac.jp

ン化した上で決定木アンサンブルを導出し, inTrees と defragTrees を用いて解釈容易なルールを抽出する試みが報告されている. 一方, IDS の一つの拡張として, 2 段階構造を持つ決定集合を対象に, 元モデルへの忠実度と明確性, 解釈容易性を同時に満たす少数ルールを選択する手法が開発されている [9].

3 解釈可能決定集合

解釈可能決定集合は, 下記の 2 段階処理を通じ, 分類規則集合を獲得する.

1. クラス相関ルール集合の獲得: アイテムの全体集合を $\mathcal{X}$, クラスラベルの全体集合を $\mathcal{C}$ とし, アイテム集合 $x_i \subseteq \mathcal{X}$ とそのクラスラベル $y_i \in \mathcal{C}$ の対 $t_i = \langle x_i, y_i \rangle$ をトランザクションとするデータベース (訓練例集合) $D = \{t_1, t_2, \dots, t_{|D|}\}$ から, クラス相関ルール $r = (s, c) \in S \subseteq 2^{\mathcal{X}} \times \mathcal{C}$ の集合 S を導出する.

2. 最適部分集合の獲得: D と S を入力とし, (後述する) 各評価関数 $f_i(D, R)$ ($i \in \{1, \dots, 7\}$) の重み λ_i (> 0) を用いた線形和関数

$$F(D, R) = \sum_{i=1}^7 \lambda_i f_i(D, R)$$

を最大化する S の部分集合

$$R^* = \arg \max_{R \subseteq S} F(D, R)$$

を最終的な決定集合 (ルール集合) として出力する. ただし実際には, $F(D, R)$ が非単調な劣モジュラ関数であることに着目し, 平滑化局所探索 (Smooth Local Search; SLS) [10] を用いて R^* の近似解を獲得する.

一方, テスト事例 x の予測クラス $\hat{y}$ は, 決定集合 R^* を用いて以下のように算出される.

$$\hat{y} = \begin{cases} c & (P = \{c\}) \\ \text{By tie-breaking rule} & (|P| > 1) \\ \text{By default rule} & (P = \emptyset) \end{cases}$$

where $P = \{c \mid (s, c) \in R^*, s \subseteq x\}$.

すなわち, R^* 中の各ルールを独立に考え, x に適用できるすべてのルールが同じクラス c を予測した場合 ($P = \{c\}$) はその値 c を採用する. また, 予測クラスに衝突がある場合 ($|P| > 1$) や適用できるルールが存在しない場合 ($P = \emptyset$) は, それぞれ利用者が設定する衝突解消規則とデフォルト規則を用いてクラスを決定する.

以下では, 各評価関数 f_i について説明する.

モデルの簡潔性: 関数 f_1 および f_2 はモデル R の簡潔性を評価する関数であり, それぞれルール数が少ない場合および各ルールが少ない条件数で構築されている場合に高い評価値が与えられる.

$$f_1(D, R) = |S| - |R|$$

$$f_2(D, R) = |S| \cdot \max\{|s'| \mid (s', c') \in S\} - \sum_{(s, c) \in R} |s|$$

モデルの有意義性: 関数 f_3 および f_4 はモデルの有意義性を評価する関数であり, 各ルール間で被覆する訓練例集合の重複が小さいほど高い評価値を与える.

$$f_3(D, R) = |D| \cdot |S|^2 - \sum_{(s_i, c_i), (s_j, c_j) \in R, i < j, c_i = c_j} |\{\langle x, y \rangle \in D \mid s_i \subseteq x, s_j \subseteq x\}|$$

$$f_4(D, R) = |D| \cdot |S|^2 - \sum_{(s_i, c_i), (s_j, c_j) \in R, i < j, c_i \neq c_j} |\{\langle x, y \rangle \in D \mid s_i \subseteq x, s_j \subseteq x\}|$$

また関数 f_5 では, 予測可能なクラス数という面からモデルの有意義性を評価する.

$$f_5(D, R) = |\{c \mid (s, c) \in R\}|$$

モデルの予測精度: 関数 f_6 および f_7 はモデルの精度を評価する関数であり, f_6 は適合性, f_7 は再現性の概念にそれぞれ相当する.

$$f_6(D, R) = |D| \cdot |S| - \sum_{(s, c) \in R} |\{\langle x, y \rangle \in D \mid s \subseteq x, y \neq c\}|$$

$$f_7(D, R) = |\{\langle x, y \rangle \in D \mid \exists (s, c) \in R \text{ s.t. } s \subseteq x, y = c\}|$$

4 グラフ構造データへの拡張

IDS のグラフ構造データへの拡張として, (1) グラフ構造データからのクラス相関ルールの抽出, および (2) グラフ構造に着目した評価関数の導入を行う.

4.1 クラス相関ルールの獲得

グラフ $g = (V_g, E_g, l_g)$ を, 頂点集合 V_g と辺集合 $E_g \subseteq V_g \times V_g$, ラベル関数 $l_g : V_g \cup E_g \rightarrow \mathcal{L}$ ($\mathcal{L}$ はラベルの全体集合) の 3 項組で表現する. グラフ g_i とそのクラスラベル y_i の対 $t_i = \langle g_i, y_i \rangle$ をトランザクションとするデータベース $G = \{t_1, \dots, t_{|G|}\}$ からクラス相関ルールを抽出することを考える.

部分グラフ p の G における支持度を

$$S(p, G) = |\{\langle g, y \rangle \in G \mid p \subseteq g\}| / |G|$$

と定義する. ここで $g_i \subseteq g_j$ は, g_i が g_j の部分グラフであることを表す. 支持度に関する閾値 σ ($0 < \sigma \leq 1$) に

対し, 条件 $S(p, G) \geq \sigma \wedge \forall q \sqsubseteq p (S(q, G) > S(p, G))$ を満たす部分グラフを頻出フリー部分グラフ [11] と呼び, その全体集合を $\mathcal{F}_G^{\sigma}$ と表記する.

以上の準備の下, $\mathcal{F}_G^{\sigma}$ を用いて G から頻出フリー部分グラフ集合をアイテムとするトランザクションデータベース

$$D_G^{\sigma} = \{ \{p \in \mathcal{F}_G^{\sigma} \mid p \sqsubseteq g_i\}, y_i \mid \langle g_i, y_i \rangle \in G \}$$

を構築し, アイテム集合に対する一般的な手法を用いてクラス相関ルール $r = (s, c) \in S_G^{\sigma} \subseteq 2^{\mathcal{F}_G^{\sigma}} \times C$ の集合 S_G^{σ} を導出する. なお, ルール $r = (s, c)$ の条件部が頻出フリー部分グラフの集合である ($s \subseteq \mathcal{F}_G^{\sigma}$) ことに着目し, 条件 $\exists g_i, g_j \in s \text{ s.t. } g_i \sqsubseteq g_j$, すなわち部分グラフの関係にあるグラフ同士を含むルールを S_G^{σ} から排除し, 冗長性を排除する.

以降では簡略化のため, D_G^{σ} を D と, また S_G^{σ} を SS とそれぞれ同一視する.

4.2 グラフ構造に着目した評価関数の導入

ルール条件部が部分グラフパターン集合であることに着目し, 評価関数 $F(D, R)$ を拡張する. 具体的には, 簡潔性の観点から新たに3つの評価関数 $f_i (i = \{8, 9, 10\})$ を導入し, パラメタ $\lambda_i (> 0)$ を伴う非単調な劣モジュール関数

$$F_G(D, R) = F(D, R) + \sum_{i=8}^{10} \lambda_i f_i(D, R)$$

を最適化基準として採用する.

以下, 各評価関数を導入する.

グラフサイズ: 関数 f_8 は, ルールに含まれるグラフサイズ (頂点数+辺数) の総計に基づき値を算出する. なお, サイズが小さいほど高い評価を与えるものとする.

$$f_8(D, R) = \sum_{(s, c) \in R} \left(\max_{(s', c') \in S} \{ \text{size}(s') \} - \text{size}(r) \right) \\ \text{where size}(s) = \sum_{(V, E, l) \in s} |V| + |E|$$

頂点・辺のラベル種: 関数 f_9 は, グラフ中の頂点ラベル, 辺ラベルに着目した評価関数であり, 総種類数が多いほど高い評価を与える.

$$f_9(D, R) = \left| \{ l(o) \mid o \in V \cup E, (V, E, l) \in s, (s, c) \in R \} \right|$$

表 1: データセットに関する統計量

	クラス Y			クラス N		
	$ V $	$ E $	$ \lambda $	$ V $	$ E $	$ \lambda $
平均	25.49	26.63	9.84	24.31	25.57	9.34
分散	3.92	4.56	0.01	6.78	7.55	0.02
最小値	14.00	13.00	8.00	2.00	1.00	1.00
Q1	21.00	21.00	9.00	20.00	20.00	9.00
中央値	24.00	26.00	10.00	23.00	24.00	9.00
Q3	28.50	31.00	10.00	27.00	29.50	10.00
最大値	49.00	52.00	11.00	56.00	59.00	11.00

$|V|$ は頂点数, $|E|$ は辺数, $|\lambda|$ はラベル種数を表す

ルール内非類似度: 関数 f_{10} はグラフの非類似度に着目した評価関数であり, 各ルール条件部が互いに非類似な部分グラフで構成される場合に高い評価を与える. なお今回は, 各ルール $r = (s, c)$ の条件部 s に含まれる部分グラフ g_i, g_j 間の非類似度 $\text{ged}(g_i, g_j)$ をグラフ編集距離を用いて算出し, その平均をルール r における非類似度としている.

$$f_{10}(D, R) = \sum_{(s, c) \in R} \frac{1}{|s| C_2} \sum_{g_i, g_j \in s, i < j} \text{ged}(g_i, g_j)$$

5 評価実験

5.1 データセットと実験設定

提案手法の有効性を確認するため, 株式会社 LIFULL が国立情報学研究所の協力により研究目的で提供している「LIFULL HOME'S データセット¹」を基に構築されたグラフデータセット [8] を用いた実験を行った.

データセットは, 東京 23 区内のマンションの各間取りをグラフ表現したものであり, 専有面積や最寄り駅までの徒歩距離, 近隣相場等を用いた推定賃料と実際の賃料との差から, クラス属性が付与されている. クラス属性は 2 値 (Y, N) であり, データ数は各クラス 240 件, 合計 480 件である. 表 1 にデータセットの概要を示す. なお, 辺ラベルは存在せず, 頂点のラベル種数は 11 である.

頻出部分グラフマイナー gSpan[12] と適切な後処理を用い, 7,616 件の頻出フリー部分グラフを抽出した後, クラス相関ルール (最小支持度 0.3, 最小確信度 0.5) の全体集合 $S^{all} = S^Y \cup S^N$ を導出した. ここで S^C はクラス値が C であるクラス相関ルールの集合であり, それぞれのルール数は, $|S^{all}| = 68,607$, $|S^Y| = 67,706$, $|S^N| = 901$ である. 表 2 に得られたルールに関する統計量を示す. 実験では, S^Y と S^N からそれぞれ 500 件

¹<https://www.nii.ac.jp/dsc/idr/lifull/homes.html>

表 2: クラス相関ルールに関する統計量

	$ s $	sup	conf	size	$ \lambda $	ged
クラス Y						
平均	2.95	53.91	59.59	23.21	6.00	0.52
分散	3.61	0.03	4.38	0.01	5.90	3.41
最小値	1.00	46.71	50.00	2.00	1.00	0.00
Q1	3.00	49.13	57.14	18.00	5.00	0.37
中央値	3.00	51.90	59.87	24.00	6.00	0.50
Q3	3.00	57.09	62.24	28.00	7.00	0.67
最大値	3.00	96.89	65.04	44.00	8.00	1.00
クラス N						
平均	2.87	75.08	51.37	14.01	4.28	0.64
分散	2.00	0.61	0.01	0.16	0.01	1.44
最小値	1.00	58.13	50.00	2.00	1.00	0.00
Q1	3.00	64.70	50.48	12.00	4.00	0.56
中央値	3.00	72.66	51.09	14.00	4.00	0.67
Q3	3.00	83.05	52.14	16.00	5.00	0.75
最大値	3.00	98.62	53.94	26.00	7.00	1.00

$|s|$ はルール本体長, sup は支持度, conf は確信度, size はグラフサイズ, $|\lambda|$ はラベル種数, ged は平均グラフ編集距離を表す。

のルールをランダムに選択することでクラス相関ルール集合 $S_i^{all} = S_i^Y \cup S_i^N$ ($i \in \{1, \dots, 10\}$) を構築し, 評価関数 $F(D, S_i^{all})$ に従い最適部分集合 $R_i^* \subseteq S_i^{all}$ を抽出するという操作を 10 回繰り返した。また, 表 3 に示す各評価基準それぞれに関し, R_i^* ($i \in \{1, \dots, 10\}$) の平均値を求めることで評価を行った。

ここで各評価基準の概要について説明する。#_Rules と Avg.Length は, モデルの簡潔性に関する評価基準であり, それぞれ f_1 と f_2 に対応し, 値が小さい方が優れていると判断する。一方, モデルの有意義性に関しては Frac.Overlap と Frac.Classes を用いて評価する。Frac.Overlap は f_3, f_4 に対応する。値域は 0 ~ 1 であり, 値が小さい方が優れていると判断する。Frac.Classes はクラスに関する網羅性を表し f_5 に対応する。今回は 2 値分類を対象としているので値域は 0.5 ~ 1 となり, 1 に近いほど良いと考える。モデルの精度に関しては, Frac.Uncovered と AUC を用いて評価する。Frac.Uncovered は説明されない事例の割合を表し f_7 に対応する。値域は 0 ~ 1 であり, 小さい方が優れていると考える。

評価基準 Graph.Size, Node.Type, Graph.Dist はグラフ構造に焦点を当てた評価基準であり, それぞれ f_8, f_9, f_{10} に対応する。Graph.Size はグラフの大きさの総計であり, 値が小さい方が優れていると判断する。一方, ノード種とルール内非類似度に相当する Node.Type と Graph.Dist は値が大きい方が優れていると考える。

表 3: 評価基準

#_Rules(R)	=	$ R $
Avg.Length(R)	=	$\frac{1}{ R } \sum_{(s,c) \in R} s $
Frac.Overlap(R)	=	$\frac{ D }{ s \binom{C}{2}} \sum_{(s_i, c_i), (s_j, c_j) \in R, i < j} \{(x, y) \in D \mid s_i \subseteq x, s_j \subseteq x\} $
Frac.Classes(R)	=	$ \{c \mid (s, c) \in R\} / C $
Frac.Uncovered(R)	=	$1 - \cup_{(s,c) \in R} \{(x, y) \in D \mid s \subseteq t\} / D $
Graph.Size(R)	=	$\sum_{(s,c) \in R} \text{size}(s)$
Node.Type(R)	=	$f_9(D, R)$
Graph.Dist(R)	=	$f_{10}(D, R)$

表 4: 実験結果

	F_G	$F_G^{\{8\}}$	$F_G^{\{9\}}$	$F_G^{\{10\}}$	$F_G^{\{8,9,10\}}$
AUC	0.58	0.52	0.58	0.58	0.52
#_Rules	4.20	2.00	4.20	4.20	2.00
Ave.Length	3.00	2.95	3.00	3.00	2.95
Frac.Overlap	0.00	0.35	0.00	0.00	0.35
Frac.Classes	1.00	1.00	1.00	1.00	1.00
Frac.Uncovered	0.11	0.22	0.11	0.11	0.22
Graph.Size	10.94	12.25	10.88	10.94	11.95
Node.Type	9.40	7.90	9.30	9.40	7.80
Graph.Dist	0.50	0.27	0.50	0.50	0.26

5.2 結果と考察

実験結果を表 4 に示す。なお表中の F_G^M は, R^* の選択基準として評価関数

$$F_G^M(D, R) = \sum_{i \in \{1, \dots, 10\} \setminus M} \lambda_i f_i(D, R)$$

すなわち M に含まれる評価関数を取り除いた場合を採用したことを表す。また $F_G^{\{8,9,10\}}(D, R) = F(D, R)$ は IDS オリジナルの評価関数に相当する。

提案した評価関数 F_G と IDS における評価関数 $F_G^{8,9,10}$ の結果を比較すると, 精度 (AUC と Frac.Uncovered) や有意義性 (Frac.Overlap) の面で F_G の方が優れていることが確認できた。一方で, ルール数 (#_Rules) は $F_G^{8,9,10}$ の方が値が小さく, 簡潔さの面では一部劣る結果となったが, 実際の値は #_Rules = 4.20 と十分に小さく, 精度 (AUC) とのバランスを考えると必ずしも大きな問題であるとは言えない。このことから, 新たに導入した評価関数 $f_8 \sim f_{10}$ は概ね有効に働いていることと結論付けることができる。

次に $f_8 \sim f_{10}$ のそれぞれの効果を確認するため, $F_G^8 \sim F_G^{10}$ の結果について考察する。表 4 より, F_G^8 と $F_G^{8,9,10}$ の結果が類似していること, および F_G^9, F_G^{10}

と F_G の結果が類似していることがそれぞれ確認できる。 F_G^8 はグラフサイズに関する評価関数 f_8 を排除した場合であり、 F_G^8 の結果が $F_G^{8,9,10}$ と類似するということから、グラフサイズが精度と有意義性の向上に強く寄与していることが示唆される。一方で、 F_G^9 と F_G^{10} の結果が F_G の結果と類似するということから、 f_9 (頂点・辺のラベル種) と f_{10} (ルール内非類似度) は、共に簡潔さに関する評価関数ではあるが、簡潔さの改善に寄与しないばかりか、精度・有意義性についても大きな影響を持たないことが伺える。以上の結果より、今回利用したデータセットに関して言えば、精度を保ったまま解釈容易なルールを導出するためにはグラフサイズを考慮することが有効であると結論付けることができる。

6 まとめ

本研究では、分類規則抽出に関する代表的な解釈性研究の一つである解釈可能決定集合を対象に、その適用範囲をグラフ構造データへと拡張する枠組みを提案した。また実データを用いた実験により、その可能性を示した。

今後の課題としては、更なる評価関数の開発と、他の実データを用いた評価があげられる。また、解釈性という観点から、人間による主観評価の実施も検討している。

謝辞： 本研究では、株式会社 LIFULL が国立情報学研究所の協力により研究目的で提供している LIFULL HOME'S データセットを利用した。なお本研究の一部は、JSPS 科研費 17K00315 の助成を受けたものです。

参考文献

- [1] A. B. Arrieta et al.: Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI, *Information Fusion*, Vol.58, pp.82–115, 2020.
- [2] 原 聡: 私のブックマーク「機械学習における解釈性」, 人工知能, Vol.33, No.3, pp.366–369, 2018.
- [3] H. Lakkaraju, S. H. Bach, and J. Leskovec: Interpretable Decision Sets: A Joint Framework for Description and Prediction, *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.1675–1684, 2016.
- [4] H. Cheng, X. Yan, and J. Han: Mining Graph Patterns, In C. Aggarwal, and H. Wang (eds) *Managing and Mining Graph Data*, pp.365–392, Springer, 2010.
- [5] B. Liu, W. Hsu, and Y. Ma: Integrating Classification and Association Rule Mining, *Proc. of the 4th International Conference on Knowledge Discovery and Data Mining*, pp.80–86, 1998.
- [6] H. Deng: Interpreting Tree Ensembles with in-Trees, *International Journal of Data Science and Analytics*, Issue 4/2019, 2018.
- [7] S. Hara and K. Hayashi: Making Tree Ensembles Interpretable: A Bayesian Model Selection Approach, *Proc. of the 21st International Conference on Artificial Intelligence and Statistics*, pp.77–85, 2018.
- [8] T. Ozaki: Extraction of Characteristic Subgraph Patterns with Support Threshold from Databases of Floor Plans, *Proc. of the 2019 Seventh International Symposium on Computing and Networking*, pp.197–203, 2019.
- [9] H. Lakkaraju, E. Kamar, R. Caruana, and J. Leskovec: Faithful and Customizable Explanations of Black Box Models, *Proc. of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp.131–138, 2019.
- [10] U. Feige, V. S. Mirrokni, and J. Vondrák: Maximizing non-monotone submodular functions, *SIAM Journal on Computing*, Vol.40, No.4, pp.1133–1153, 2011.
- [11] Z. Zeng, J. Wang, J. Zhang, and L. Zhou: FOGGER: An Algorithm for Graph Generator Discovery, *Proc. of 12th International Conference on Extending Database Technology*, pp.517–528, 2009.
- [12] X. Yan and J. Han: gSpan: Graph-based Substructure Pattern Mining, *Proc. of the 2002 IEEE International Conference on Data Mining*, pp.721–724, 2002.

評価表現辞書の精錬による化粧品レビューの推薦精度向上

Improving recommendation adequacy of cosmetics reviews by refining evaluation expression dictionary

今江 柚実*
Yumi Imae

松下 光範
Mitsunori Matsushita

関西大学 総合情報学部
Faculty of informatics, Kansai University

Abstract: The aim of this study is to improve the recommendation adequacy of the system for providing cosmetic reviews. In recent years, to help a user's exploration of cosmetics, a system has been proposed that recommends cosmetics reviews close to the user's preference. The system, however, sometimes provide reviews with incorrect scores. The reason is that the value of the evaluation expression dictionary used by the system, which was created from the review corpus automatically, was incorrect, hence the system miscalculated the evaluation score. To solve this problem, this paper proposes a method to refine the evaluation expression dictionary; the method categorizes review sentences with the naive Bayes method and extracts sentences that mention the effect of the cosmetics product from the review corpus.

1 はじめに

化粧品の種類は日々増加している。これは、単なるブランド数の増加によるものではなく、肌質や使用感に焦点をあて、各ブランドが商品展開を広げているためである。利用者は、これらの多くの商品を比較し、自身の嗜好や条件に合致するものを選択する。化粧品は人体に直接塗布するという性質上、誤った化粧品を選択すると肌トラブルを引き起こす原因となるため、購入前に情報を収集することが重要である。近年では、利用者がこうした情報を集める手段として、化粧品レビューサイトが参照されている。レビューサイトの利用には様々な人の使用感や感想を取得できるというメリットがある。しかし、ひとつの商品に対し多数のレビュー文が投稿されているうえ、各レビュー文中には使用感や価格など様々な情報が混在しているため、それらのレビューを全て確認して必要な情報を得ることは困難である。これに加え、同じ価格や使用感であってもその捉え方には個人差があるため、利用者自身の嗜好や肌質に合致する化粧品の情報を効率的に取得することは容易ではない。

この問題を解決するため、松波らは利用者の好みや使用感に着目し、価値観の近い投稿者によるレビューを推薦するシステムを提案している [1]。このシステム

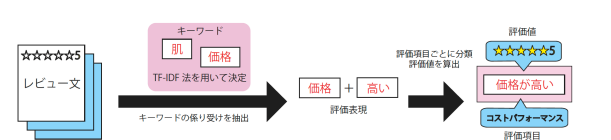


図 1: 評価表現辞書の構築の流れ（文献 [2] より抜粋）

では、化粧品の評価に関するキーワードとそれを修飾する特徴や程度、否定の有無で構成される評価表現に評価値を付与した辞書（以下、評価表現辞書と記す）を用いて化粧品の得点を算出しているが、この辞書の作成は手動で行われており大きなコストがかかっていた。この処理の軽減を目的として、酒井らは評価表現辞書の自動構築手法を提案した [2]。酒井らの辞書構築の流れを図 1 に示す。

酒井らが作成した辞書は、松波が手動で構築した辞書を参考に、@cosme¹のレビュー文や評価値のデータを用いて構築されていた。しかし酒井らの辞書には、辞書の評価表現の核となるキーワードに評価と関係のない単語が含まれていることや、「シミ」+「できる」のようなネガティブと判定すべき評価表現に高い評価値が付与されていたり、「シミ」+「なくなる」のようなポジティブと判定すべき評価表現に低い評価値が付与されていたりするなど、精度の点で問題があった。この問題の原因として、辞書の作成に用いるレビュー文

*連絡先：関西大学総合情報学部
〒569-1095 大阪府高槻市霊仙寺町 2-1-1
E-mail: k943749@kansai-u.ac.jp

¹<https://www.cosme.net/> (2020/3/3 確認)。

に本人の肌質に関する情報など商品の効果に関わらない文が含まれていることが挙げられる。こうした問題を解決するため、本研究では評価表現辞書に使用するレビュー文を商品の効果に関わる文に絞ることで評価表現辞書の精練を試みる。

2 関連研究

レビューサイトを分析することで、利用者の情報取得を支援したり、開発者の手助けとなる情報を抽出したりする研究が様々に行われている。本章では、レビュー情報を用いた研究とナイーブベイズ分類器を用いた研究について述べる。

開発者にとってレビューの情報を得ることはアイデアを得るきっかけになるなど重要であるが、レビュー文中には不必要な情報も多く含まれる。こうした観点から、Walidらは、App storeのレビュー文を分析し、バグの報告、機能へのリクエスト、ユーザの経験、評価への自動分類を可能にした[3]。開発者などにインタビューを行い、手法による精度を比較したところ、メタデータのみを基にした分類器は特に精度が悪く、テキストベースのナイーブベイズ分類器ではより精度の高い分類を確認した。

紀本らは、映画レビュー文から映画の内容や評価に関する情報を抽出し、利用者に提示する手法を提案した[6]。ある物事を評価する語（評価語）とし、これにより評価される語（評価属性）を特徴語として抽出した。紀本らによる特徴語抽出手法とTF-IDF手法により得られた単語と比較するため、評価やレビューの数である「評価数」と、映画の内容である「内容」、映画の中で肯定的か否定的に評価されている点である「評価点」、意見の傾向である「傾向」の4つの項目についてアンケート調査を実施した。その結果、紀本らの手法はこれらの項目のうち、「評価数」「評価点」「傾向」の評価を得る際に有用であることが確認された。

金子らは、購買意図によって商品の検索を行うシステムの階層的知識を自動獲得するため、商品レビューから抽出した購買意図の表現や、商品のカテゴリ情報を用いて購買意図の構造化を行った[5]。レビューをCaboCha²を用いて解析し、それに係る目的語を含む文節を抽出した。この抽出した表現に対しアノテーションを行い購買意図を表現している表現を判断した。この判断した表現が出現するカテゴリの集合の包含関係とcos類似度から購買意図の階層化を行った。構造化の観点からの分析や階層的知識の実用性の調査により「美肌を作る」などの具体的な意図よりも「痩せる」というような抽象的な意図の方が、出現カテゴリ数が多いという仮説が示された。また「セルフネイルをやってみる」

と「ジェルネイルデビューする」のように、意味が似ている表現間で、カテゴリ毎に頻度分布の形も似る傾向にあることが示唆された。

漆原らは、商品販売サイトのレビューにおいて商品の評価を一目で判断できるようにするため、評価表現の印象がポジティブかネガティブかを示す印象極性に色を付けて表示する手法を提案した[4]。この印象極性に対し、「悪く」の直後の「ない」のような否定表現を抽出し、その複合語の印象極性を反転し、文脈上に逆接表現が含まれる場合に直後の評価表現を修正した。人手で作成した正解データと比較し印象極性の推定実験を行い、辞書による判定ができなかったものが判定できるようになったことが確認された。

細田らは、青少年のインターネット・リテラシーレベルの向上のため、twitter³情報を用いて、ナイーブベイズ分類器を導入したリテラシーレベルを評価する問題解決環境を提案した[7]。この問題解決環境とは、ユーザが特別な知識を持たずともそれを評価できる環境のことである。ここではツイート内容が誹謗中傷であるか否かが対象であるため、時間などの評価に関係のない情報や、引用ツイート、URLは評価の対象から取り除いた。形態素解析したツイートに誹謗中傷のキーワードを含んでいるかを分類し、含まれていた場合、再度ナイーブベイズ分類器によりそのツイートが誹謗中傷かを分類した。これにより、従来手法であった誹謗中傷内容の単語によるパターンマッチングでの識別と比較して1.71倍高い結果が得られた。

本研究ではこれらの手法を参考に、評価表現辞書に用いるデータを商品の効果に関わる文に絞り、辞書の精練を行う。

3 評価表現辞書の精練手法

本研究ではレビュー文を商品の効果に関わる文であるかに注目し分類して、辞書構築に用いるレビュー文を商品の効果に関わる文に絞り、自動作成の際の精度の向上を図る。計算機による自動分類を実現するため、本研究では教師あり機械学習手法のひとつであるナイーブベイズ分類器を用いてレビュー文の分類を行い、その分類した文を用いて評価表現辞書を作成した。

3.1 レビュー文の収集

辞書構築に用いるレビューデータは、肌質による違いを評価表現辞書に反映するため、「普通肌」と「混合肌」の2つの肌質について、各々「化粧水」「口紅」「ファンデーション」の3つのカテゴリ（計6種類）を対象とした。この6種類について@cosmeの最新口コミラン

²<https://taku910.github.io/cabocha/> (2020/1/31 確認)

³<https://twitter.com/> (2020/1/31 確認)

キング上位 50 商品を選定し、1 つの商品につき 100 件の評価値とレビュー文を取得した。ここで、辞書の評価表現に対する評価値に、レビュー文に付与されている評価値を使用するため、評価値が「評価しない」と設定されているレビュー文は除去した。レビュー文の取得には、BeautifulSoup⁴を用いた。このうち、10 商品分の各 50 件を分類の学習に用いるデータとした。

3.2 ナイーブベイズ分類器を用いた商品効果文の抽出

前節で収集したレビュー文に対し、句点や、否定後の読点で文を区切り、これを人手で商品の効果に関わるものと関わらないものの 2 つに分類した。これらの文を単語ごとに分割し、その単語数をカウントさせナイーブベイズ分類器で学習した。

句点や、「ですが」のような否定形の後の読点である場合、「乾燥がひどく悩んでいましたが、これを使うとしっかりとりました」のように、読点の前後で記載されている情報が商品の効果に関する文と関わらない文に変化する可能性があるため、句点や否定後の読点で区切った。ナイーブベイズ分類器を用いて、これらの文を商品に関する効果が記載されている文と、レビューする商品の効果に関わらない文の 2 種類に分類した。この分類には機械学習ライブラリ scikit-learn⁵を用いた。

商品の効果に関する文のみを「yes」、本人の肌質や、他の商品の話題、商品の何が効果的であるのか記載されていない文などの、商品の効果に関しない文を「no」として人手で訓練データを作成した。

レビュー文が商品の効果に関する情報であるか否かという文章の分類に対しては肌質による違いはなかったため「化粧水」「口紅」「ファンデーション」のカテゴリごとに、普通肌のレビュー文の学習と混合肌のレビュー文の学習において同じ教師データを用いた。本研究では全ての品詞を対象にストップワードは設定せずに、それぞれ約 3000 文を訓練データとして扱うこととした。学習させるデータのサイズを統一させるため、形態素解析を行い文章を単語ごとに分割し 1 つの文章に含まれている単語数をカウントした。形態素解析には MeCab-0.996⁶を用いた。この際、「ネイル用品」が「ネイル」と「用品」に分割されるように 1 つの単語が複数の単語として認識されることを防ぐため、事前に MeCab に対し化粧品のカテゴリ名やブランド名などに対応したユーザ辞書を登録した。

このように作成したデータの 8 割を訓練データ、2 割をテストデータとして、学習数での精度を測った。訓練

データ数は、化粧水が 3191 文、ファンデーション 3192 文、口紅 3207 文であった。精度はいずれのカテゴリにおいても 78 %程度であった。

これらを学習させ、「yes」と判断された文のみを抽出し、本研究で評価表現辞書に用いるデータとした。

3.3 評価表現辞書の構築

前節で抽出した化粧品の効果に関する文集合を対象として、酒井らの手法を参考に評価表現辞書を構築した。評価表現は「ニキビ」や「乾燥」などのキーワードを修飾する語とした。キーワードの設定には、「化粧水」などの 1 つのカテゴリにおいて、どの商品のレビュー文にも用いることができる辞書を作成するため、化粧品ブランド名や商品名をストップワードとした。これに加え、「いつ」などの使用感を示さない単語がキーワードに入ることを防ぐため SlothLib⁷を用いてストップワードを拡張した。

以上の前処理を行った化粧品のレビュー文の中から MeCab を用いて名詞のみの単語を抽出した。さらに、TF-IDF 法を用いて抽出した単語ごとの重要度を算出した。この算出した重要度が上位 1 % の単語を評価表現辞書の評価表現の核であるキーワードとした。このように設定したキーワード数は、酒井らの手法を用いて作成した辞書では約 280 個、本研究の手法で作成した辞書では約 170 個であった。係り受け解析器 Cabocha を用いて設定したキーワードを修飾する形容詞、副詞、動詞、サ変接続の名詞、助動詞の「ない」を収集し、評価表現である「特徴」「程度」「否定」の抽出を行った。

評価値は、投稿者がその商品に対し付与した@cosme 上のクチコミ評価の星の数を用いた。また、設定した「キーワード」「特徴」「程度」「否定」「評価値」を辞書に登録した。

このようにして登録した辞書に@cosme に効果として付与されている「うるおい」や「アンチエイジング」などの単語を参考に評価項目を設定した。Word2Vec を用いて、評価表現である「キーワード」「特徴」と設定した「評価項目」の類似度を算出し、最も値が高いものをそれぞれの評価表現に対する評価項目として割り当てた。評価項目と評価表現の単語が一致する場合、その単語を、類似度を算出せず評価項目とした。

4 実験

4.1 辞書の精度比較

酒井らの評価表現辞書による精度と、本提案手法により作成した評価表現辞書の精度の比較実験を行った。

⁷<https://ja.osdn.net/projects/slothlib/> (2020/3/2 確認)。

⁴<https://www.crummy.com/software/BeautifulSoup/bs4/doc/> (2020/1/31 確認)

⁵<https://scikit-learn.org/stable/> (2020/1/31 確認)

⁶<https://taku910.github.io/mecab/> (2020/1/31 確認)

実験参加者は20から50代の女性15名で行った。実験には化粧水、口紅、ファンデーションの3カテゴリにおいて、普通肌と混合肌の2つの肌質を対象にした。この対象においてレビュー文をランダムにそれぞれ30件をあらかじめ用意した。評価値は、@cosmeの星の数と同様に設定しており、評価値が高いほど、評価の良いものとした。実験参加者は、レビュー文を読み、評価項目に対し0から7の評価値を付与することが指示された。レビュー文から判断不可能な評価項目がある場合は、その項目の欄に「×」を記入するように指示された。

評価表現辞書を用いた評価値として、レビュー文から取得した評価表現と同様のものを評価表現辞書の中から探し、評価項目に対する評価値を算出した。これを酒井らの手法の評価表現辞書と本手法による評価表現辞書の2つの辞書において行った。精度の比較方法としては、酒井らの提案した評価表現辞書を用いた評価値と、実験参加者に付与してもらった評価値の平均絶対誤差を算出した。これに加え、本研究の提案手法にて作成した辞書を用いた評価値と、実験参加者に付与してもらった評価値の平均絶対誤差を算出し、この2つの平均絶対誤差を比較した。

個人の化粧への関心や化粧品レビューの使用度や違いによって評価値に変動があるかを調査するため、事前アンケートを行った。実験参加者は化粧の経験の有無、化粧にかかる時間、化粧の頻度、購入方法、化粧品のレビューを普段見るか、見る人に対してはどのくらいの頻度であるかが尋ねられた。15人全ての人が化粧経験があり、週に6回以上化粧をしていることが確認できた。購入場所はネットと店頭で半分程度に分かれたが、化粧品のレビューを確認する頻度は、商品購入時に確認する程度であった。

実験の結果を図2、図3、図4に示す。化粧水（普通肌）では5項目に対し平均1.27、化粧水（混合肌）では3項目に対し平均0.65、口紅（普通肌）では7項目に対し平均0.73、口紅（混合肌）では7項目に対し平均0.41、ファンデーション（普通肌）では6項目に対し平均0.44、ファンデーション（混合肌）では7項目に対し平均0.57の平均絶対誤差が減少した。その他の15項目については、平均絶対誤差の減少は確認されなかった。

4.2 評価表現辞書のキーワードの比較

キーワードは評価の対象となる単語であるため、どの商品に対しても扱えるものでなくてはならないが、酒井らの手法による辞書には適当でないキーワードが含まれていた。化粧水におけるキーワードの例を表1に示す。酒井らの手法による辞書では「サンプル」や「キャ

表 1: キーワードの比較

酒井らの辞書にのみ出現	本研究の辞書にのみ出現
コスメ	もちもち肌
肌質	トロツ
年齢	ノリ
評価	皮膚
クチコミ	つけ心地
大丈夫	ローズ
風呂上がり	アロマ
現品購入	テカリ

ンペーン」「大好き」「正直」など商品の効果に関わらないものがキーワードに含まれていたが、本研究の手法の辞書では除去されていた。代わりに「トロツ」や花の名前など効果を表すキーワードが追加されたことが確認できた。しかし、「シミ」や「年齢」のように評価を表すにあたり必要なキーワードにおいて、酒井らの手法により提案された辞書のみでしか確認できないものがあり、辞書に必要なキーワードも除去されていたことがわかった。

5 考察

前節で確認した実験結果を踏まえ、酒井らの提案した評価表現辞書と本研究で提案した評価表現辞書の違いを考察した。これに加え、実験を行った結果、平均絶対誤差の減少が確認されなかった問題点や改善すべき点について述べる。

5.1 評価値について

酒井らは、逆説の接続詞または接続助詞の前後で評価が反転すると指摘されていることから、接続詞の位置によって評価値の変動を与えていた。文中に逆説語を含む場合は前の評価値を、文頭に逆説語がある場合は文字数の少ない方の評価値を、7から与えられている評価値を引いた値としていた。しかし、本研究では1文ずつ否定の表現後の読点で区切って分類を行ったため、酒井らが行っていた否定による評価値の変化を与えることができなかった。否定的な表現に対してもそのままの評価値を当てはめており精度を下げてしまう原因となる。このため今後は1つずつの評価表現に対しポジティブかネガティブの判定を行うなど、他の評価値の算出方法を検討する必要がある。

5.2 辞書の精練について

分類をする上で否定形後の読点で文を区切ったため、読点後の文に主語がなくなる問題があった。例えば、「毛穴が隠れないと聞いて心配していましたが、私は気になりませんでした。」のように1度しか主語が現れない場合において「毛穴が隠れないと聞いて心配していましたが、」と「私は気になりませんでした。」を1文ずつ取得することになるため、1つ目の文は「レビュー文とは関係のない商品についての記述」「効果が具体的に示されていない記述」として、どちらの表現も商品の効果に関しない文と認識されていた。このため「毛穴が気にならない」という評価表現に対する評価値を算出することができなかった。このため今後は文を区切る前に文の主語を抽出する必要がある。

5.3 キーワードの設定について

前節で述べたように、本手法を用いて用いるデータを商品の効果に絞ることで評価表現辞書のキーワード中に「大丈夫」や「キャンペーン」など評価に関わらない単語が除外され、代わりに「トロッ」や花の名前など評価に関わる単語が加わった。しかし、「シミ」や「年齢」という評価に関わる単語も除外されていた。これは、TF-IDF 値の上位 1 % の名詞をキーワードとしたため、キーワード数が酒井らの手法に比べ 100 個ほど少なくなったことが原因であると考えられる。そのため今後は、キーワードの取得数を検討する必要がある。

5.4 肌質による辞書の違いについて

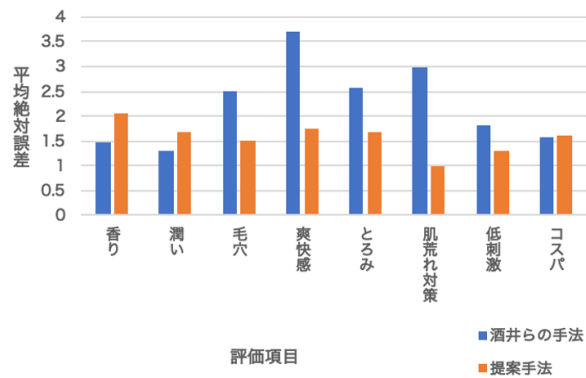
本研究では、肌質による評価値の変化を確認するため 2 つの肌質に対し評価表現辞書の構築を行った。これは、@cosme の「肌質」の項目に用いられている肌質から「普通肌」と「混合肌」を抽出し、それぞれの肌質のレビュー文を収集した。しかし、レビュー文を分類する上で「普通肌」として取得したレビュー文の中に、「私は乾燥肌なので」や「私は混合肌ですが」という表現が現れており「肌質」に設定している肌質とは異なる投稿者が多く確認された。より有用に肌質を考慮した辞書を構築するためには、文章中出现する肌質も考慮する必要がある。今後は、1 つずつのレビュー文の中から「普通肌」「敏感肌」のように肌質名を抽出し、肌質を決定するべきである。ただし、「私は乾燥してしまいましたが脂性肌の人には良いと思います」のように投稿者の肌質でない文が含まれる。これを、投稿者の肌質と設定することは精度を下げる原因となるため、他の人に対して推薦する文や肌質を述べた文が過去形である時、肌質としない設定をする必要がある。

6 おわりに

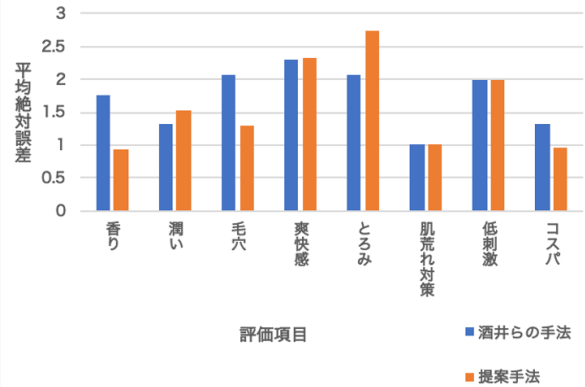
化粧品レビューサイトはレビュー数が多く、また使用感などの情報が混在しているため、利用者が求める情報を取得することは難しい。この問題を解決するため、個々の利用者に対し有用なレビュー文を推薦するシステムが提案されているが、そのシステムを構築する際に用いる評価表現辞書の精度が悪いという問題があった。本研究では化粧品の評価表現辞書に用いるレビュー文を商品の効果に関わる文に絞ることで自動構築における評価表現辞書の精練を行った。先行研究による手法と本提案手法の評価表現辞書の精度を比較した結果、35 項目の評価項目に対し、平均絶対誤差の減少が確認された。今後は、分類後に主語がなくなることへの対策、辞書構築における評価値の算出方法、キーワードの設定数、投稿者の肌質の判定方法、について検討し、より精度を向上させることを目指す。

参考文献

- [1] 松波友稀, 上田真由美, 中島伸介: コスメアイテムに対する評価項目別レビュー自動スコアリング方式の開発, 第 9 回データ工学と情報マネジメントに関するフォーラム, B5-3 (2017).
- [2] 酒井美春, 松下光範, 上田真由美: 化粧品の評価項目別スコア生成のための評価表現辞書の自動構築, 第 11 回データ工学と情報マネジメントに関するフォーラム, B6-2 (2019).
- [3] Walid, M., Zijad, K., Hadeer, N. and Christoph, S.: On the automatic classification of app reviews, *Requirements Engineering*, Vol. 21, No. 3, pp. 311-331 (2016).
- [4] 漆原駿, 町田翔, 延澤志保: 印象極性による商品レビューの視覚化, 情報処理学会第 80 回全国大会講演論文集 (第 2 分冊), pp. 311-312 (2018).
- [5] 金子貴美, 村上浩司, 石野亜耶: 商品レビューからの購買意図の抽出と自動階層化, 人工知能学会全国大会論文集, pp. 2P113in1(2016).
- [6] 紀本雅大, 伊藤淳子, 宗森純: 評価表現に着目した映画レビューからの評価情報抽出, 情報処理学会研究報告, Vol. 2019-GN-106, No. 43, pp. 1-8 (2019).
- [7] 細田尚志, 前田太陽, 井上聡, 石崎博基, 権藤俊彦: ナイーブベイズ分類器を用いたインターネット・リテラシーレベルを評価する問題解決環境の有効性, 日本計算工学会論文集, Vol. 2017, Paper No. 20170003 (2017).

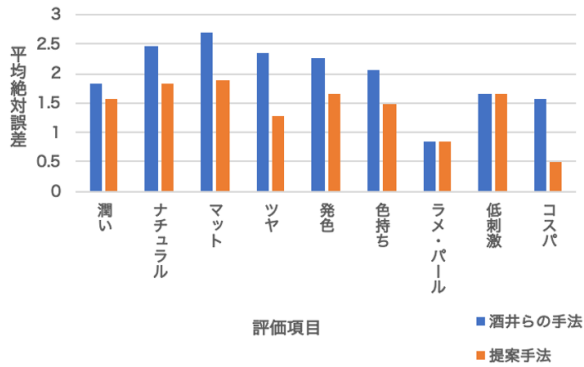


(a) 化粧水（普通肌）の平均絶対誤差

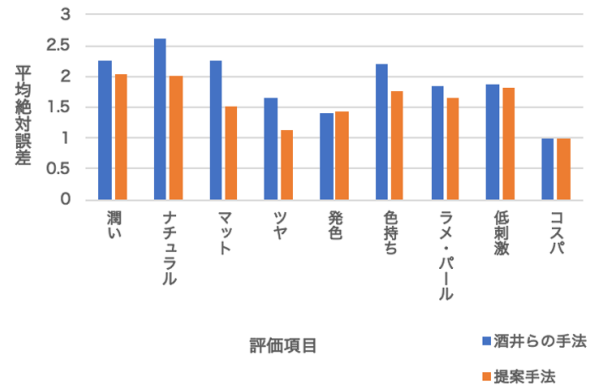


(b) 化粧水（混合肌）の平均絶対誤差

図 2: 化粧水の結果

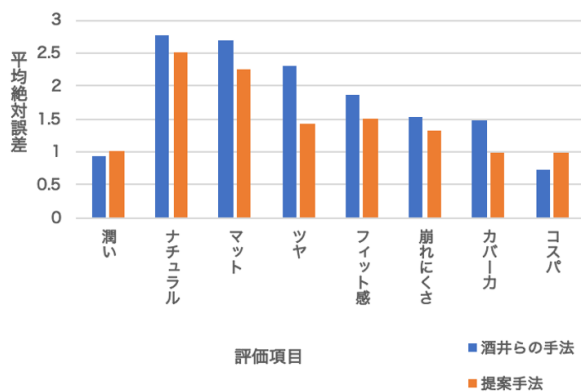


(a) 口紅（普通肌）の平均絶対誤差

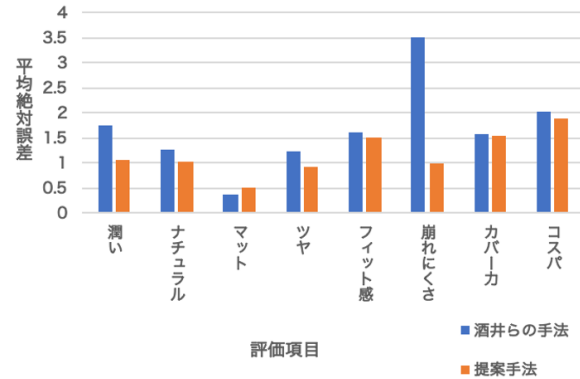


(b) 口紅（混合肌）の平均絶対誤差

図 3: 口紅の結果



(a) ファンデーション（普通肌）の平均絶対誤差



(b) ファンデーション（混合肌）の平均絶対誤差

図 4: ファンデーションの結果

深層学習を用いた好意と悪意を含む表現の可視化による 文章推敲支援

Document Polishing Support by Visualizing Favorable and Malicious Expressions using Deep Learning

庵 翔太^{1*} 砂山 渡² 畑中 裕司² 小郷原 一智²
Shota IHORI¹ Wataru SUNAYAMA² Yuji HATANAKA² Kazunori OGOHARA

¹ 滋賀県立大学大学院工学研究科

¹ Graduate School of Engineering, The University of Shiga Prefecture

² 滋賀県立大学工学部

² School of Engineering, The University of Shiga Prefecture

Abstract: In recent years, the number of people who interact online has increased due to the increase of SNS users. Considering this situation, it can be said that it is important to revise sentences in consideration of the reader's emotions during dialogue. Therefore, the purpose of this research is to support the polish of sentences that give readers a favorable impression. For this purpose, we implemented functions to visualize expressions including good and bad intentions using deep learning and functions to encourage sentence improvement. As a result of the experiment, it was found that this system can support polish of sentences that give readers a favorable impression.

1 はじめに

近年、スマートフォンの急速な普及に伴うソーシャルメディアの利用者の増加によって、増々多くの人々が、オンラインでのコミュニケーションを行うようになってきている。しかしながら、こういったオンラインでの対話では、書き手の表情や細かいニュアンスが伝わらないため誤解が生じやすく、相手の気分を害してしまい、トラブルの原因となることがある。また、昨今ソーシャルメディアにおける不適切投稿によって生じるトラブル、いわゆる「炎上」が問題となっている[1]。

こういった現状を考えると、オンラインでのコミュニケーションを行う際に、読み手の感情を考慮した文章作成を行うことは、良好な人間関係を構築する上で重要であると言える。例えば、メールを作成する時に、そのメールを読む人の気分を害するような表現を避けることは我々が日常的に行っていることである。そこで、読み手が好感を抱く文章の作成を支援する事が出来れば、コミュニケーションの支援として有用となる。例えば、Twitterなどのソーシャルメディア上で、読み手の気分を害する投稿をしようとする書き手のユーザに注意喚起するサービスや、メールの自動添削システム

が考えられる。

そこで本研究では、メールやメッセージアプリ、ソーシャルメディアなどのオンラインでの他人とのコミュニケーションにおいて、他人に送りたい文章を入力として、読み手に好感を持ってもらえる文章推敲を支援するシステムを構築する。筆者の学部時代の研究[2]では、書き手の好意や悪意を表す単語を可視化する機能、文章修正を促すメッセージを表示する機能、文章をスコア付けして表示する機能を実装し、文章修正を支援する。本研究では、これらの機能に加えて、深層学習を用いて、文章の書き手の好意や悪意を含む文を抽出し可視化する。これによって書き手に文章推敲を促し、読み手から好感を持たれる文章の作成を支援することを目的とする。

以下本論文では、2章で関連研究について述べる。3章で提案する文章推敲支援システムについて述べる。4章で提案システムの効果を検証した評価実験について述べ、5章で本論文を締めくくる。

2 関連研究

文章作成を支援する研究は盛んに行われている。

段落の一貫性を数値的に評価する研究がある[3]。この研究では、段落一貫度を定義し、論文執筆などの際

*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
庵翔太

〒522-8533 滋賀県彦根市八坂町 2500
E-mail: of23sihori@ec.usp.ac.jp

の文章校正支援を期待している。また、文の接続関係を利用する論理方向を可視化し、段落中の話の展開を確認しながら執筆可能な支援環境を用意する研究がある [4]。さらに、英語文章作成時に、過去の添削履歴を提示することにより、学習者に変更を促す英語文章作成支援システムを構築する研究がある [5]。

また、文章から読み手の感情を予測する研究も行われている。

システムがユーザの感情を推測し、その時々感情状態に応じて感情表現を用いながら、ユーザの説得を行う対話システムを提案する研究 [6] が行われている。この研究では、感情表現が対話に与える影響に着目して、人を説得するための効率的な方法として感情表現を用いている。

また、オンライン上の対話において聞き手の感情を予測し喚起させる研究 [7] がある。この研究では、マイクログログから大規模な感情タグ付き対話コーパスを構築し、感情喚起モデルを学習することで読み手に感情を喚起させる応答を生成している。

本研究では、他人とのコミュニケーションを目的とした文章を、読み手に好感を持ってもらえる文章に推敲する支援を行うために、深層学習を用いて、文章中の書き手の好意や悪意を表す表現を抽出し可視化するインタフェースを提案する。

3 好意と悪意を含む表現の可視化による文章推敲支援システム

3.1 システムの概要

本研究における文章推敲支援システムの構成を図 1 に示す。本研究では、まず文章を入力として、入力に含まれる好意や悪意を表す単語と、好意や悪意の表現を含む文を抽出する。そして、抽出した単語や文をそれぞれ文章推敲支援インタフェースで可視化する。そのうえで、文の好意悪意度合い表示、文章推敲を促すメッセージ表示、文章のスコア表示、文章の編集機能を提供し、ユーザに文章の推敲を促す。

3.2 好意と悪意を表す単語の抽出 [2]

文章中の好意を表す単語、悪意を表す単語を抽出するために、好意、悪意を表す単語の単語辞書を用いる。単語辞書を構築するために、Twitter 上のツイートとリプライのデータ集合から「喜び」または「悲しみ」を表すリプライの元ツイートに含まれる高頻度語を抽出する。詳細は筆者の学部時代の研究 [2] を参照されたい。

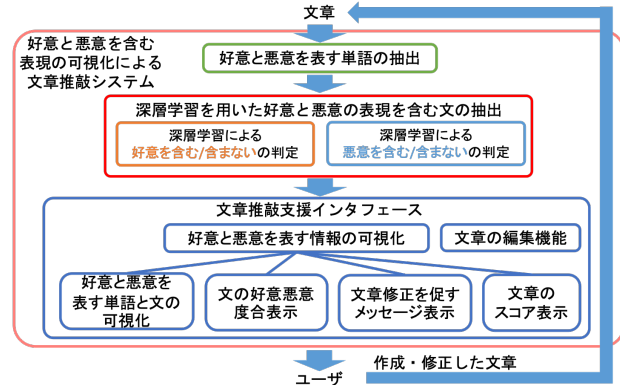


図 1: 文章推敲支援システムの構成

3.3 深層学習を用いた好意と悪意の表現を含む文の抽出

3.3.1 文の抽出に用いる深層学習モデル

文章中の好意の表現を含む文、悪意の表現を含む文を抽出するために、深層学習を用いて文中の単語の組み合わせや出現順などを考慮した文の判定を行う。具体的には、文章中の各文が、好意の表現を含む文か、含まない文かを判定する深層学習のモデル（以下これを「好意モデル」と呼ぶ）と悪意の表現を含む文か、含まない文かを判定する深層学習のモデル（以下これを「悪意モデル」と呼ぶ）の 2 つのモデルを学習し、文の判定を行う。

2 つのモデルの学習には、深層学習の手法の一つである LSTM(Long short-term memory) を用いる。LSTM とは、時系列データの分類や回帰をするモデルである RNN(Recurrent Neural Network) を記憶を長期にわたって保つように拡張された、長短期記憶と呼ばれる時系列データに対する再帰的モデルの一種である。これを用いることで、単語の組み合わせや出現順を考慮して、各文が好意や悪意の表現を含む文か、含まない文なのか評価する。

3.3.2 学習データの収集

深層学習の学習データには、Twitter におけるユーザ間の対話情報に筆者が人手でラベル付けしたものを 200 次元の分散表現にしたものを使用する。まず、Twitter Streaming API, Twitter REST API[8] を利用することにより、Twitter 上でのある発話（これを「ツイート」という）について、その発話に返答（これを「リプライ」という）が行われているツイートを収集する。収集の際には、Twitter 特有の表現を含むツイートは削除する。具体的には、「【定期】」のように使われる「【】」を含む

ツイートは削除する。「【定期】」のような文字列は、定期的に同じ内容のツイートを発信する際に用いられることから、何らかの宣伝や告知のために用いられる事が多い。こういったツイートには書き手の好意や悪意は含まれないことから今回収集するツイートからは削除する。また、ツイート中の URL や記号、絵文字は削除する。これは、URL は文章の上で意味をなさない文字列であるため、書き手の好意と悪意を表す単語にはなり得ないからである。また絵文字や記号は単語ではないため、今回収集したツイートからは削除する。さらに、Twitter にはボット (特定のキーワードに反応し自動返信するプログラム) と呼ばれるユーザが存在する。ボットの発言を収集すると発言と応答の内容に偏りが生じるため、ツイートのユーザ名に「bot」が含まれるツイートまたはリプライは削除する。

3.3.3 学習方法

好意モデルと悪意モデルを学習するための学習データには、3.3.2 項で取得した Twitter 上の対話データのリプライに、筆者が人手でラベルをつけたものを使用する。好意モデルの正例/負例のラベル付けの基準として、「このメッセージをもらった時に嬉しい気持ちになるか/ならないか」を判断する。前者のラベルをつけたデータは、好意の表現を含む文、つまり好意モデルの正例とする。後者のラベルをつけたデータは、好意の表現を含まない文、つまり好意モデルの負例とする。悪意モデルのラベル付けも同様に「このメッセージをもらった時に不快な気持ちになるか/ならないか」を判断する。前者のラベルをつけたデータは、悪意の表現を含む文として悪意モデルの正例とする。後者のラベルをつけたデータは、悪意の表現を含まない文として悪意モデルの負例とする。好意モデルの正例/負例の数はそれぞれ 681 データ、悪意モデルの正例/負例の数はそれぞれ 519 データを用いる。

こうして収集した正例と負例を単語に分ち書きし、各単語を 200 次元の分散表現にする。単語の分ち書きには、オープンソース形態素解析エンジンである MeCab を利用する。また、単語を分散表現にするために、単語分散表現モデルである hottoSNS-w2v[9] を利用する。hottoSNS-w2v は SNS や Web 上の文書を学習コーパスとして word2vec により学習させた単語分散表現モデルである。これにより、学習データの各単語を 200 次元の分散表現とする。最終的な各学習データの形状は [全データ中の最大単語数以上の数, 単語の次元数] となるように [150, 200] のベクトルとする。各学習データの単語数が 150 に満たない場合は、満たない分を 0 で埋める。これを学習データとして、好意モデル、悪意モデルの学習を行う。学習には、好意モデ

表 1: 好意モデル, 悪意モデルの学習パラメータ

エポック	100
バッチサイズ	256
中間層	1 層
中間層ノード数	50
活性化関数	relu
最適化アルゴリズム	Adam
学習率	0.001
ElasticNet 正則化	0.001
ドロップアウト	0.5

表 2: 好意モデル学習結果

	Precision	Recall	F1
正例	0.85	0.75	0.80
負例	0.79	0.88	0.83
平均	0.82	0.82	0.82

表 3: 悪意モデル学習結果

	Precision	Recall	F1
正例	0.74	0.81	0.78
負例	0.85	0.80	0.82
平均	0.81	0.80	0.80

ル、悪意モデル共に表 1 のパラメータを使用し、学習データの内 2 割をテストデータとして用いる。

好意モデル、悪意モデルの判定精度を表 2、表 3 に示す。本システムでは、「好意を含まない文」を「好意を含む文」と判定することを避けるために、好意モデルは正例の Precision が高いモデルを採用した。また同様に、「悪意を含む文」を「悪意を含まない文」と判定することを避けるために、悪意モデルは正例の Recall が高いモデルを採用した。上記の手法で学習した深層学習モデルを好意モデル、悪意モデルとし、入力文章の各文が好意、悪意の表現を含む文か、含まない文かをそれぞれ判定する。

3.4 文章推敲支援インタフェース

抽出した好意、悪意を表す単語や文を可視化し、ユーザに文章推敲を促すためのインタフェースを実装する。また、本システムはテキストデータのための総合環境 TETDM(Total Environment for Text Data Mining)[10][11] 上で構築する。図 2 はシステムの表示例であり、システムに文章を入力した時の出力を表示している。また、便宜上各機能を表示している部分を色付きの枠で囲い番号を割り振っている。

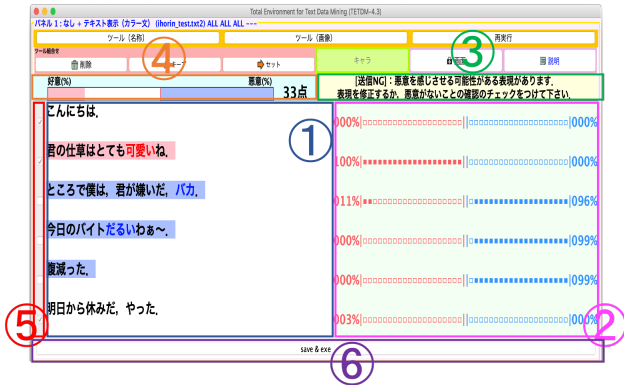


図 2: 文章推敲支援インターフェースの構成

3.4.1 好意と悪意を表す単語と文の可視化

表示例を図 2 の①に示す。3.2 節の手法により抽出された書き手の好意を表す単語は文字色を赤色に、悪意を表す単語は文字色で青色で表示している。また、3.3 節の手法により抽出された書き手の好意を含む文は薄い赤色で、悪意を含む文は薄い青色でハイライトしている。文のハイライトについては、3.3.1 項で前述した好意モデルによって、入力文章の各文が好意を含むと判定されると、該当の文を薄い赤色でハイライトし、悪意モデルによって悪意を含むと判定されると、薄い青色でハイライトする。好意にも悪意にも判定された場合は、悪意の判定を優先して薄い青色でハイライトする。この単語の文字色の変更と文のハイライトによって、書き手の好意や悪意を表す単語や文をユーザに視覚的に提示する事で、文章推敲を促す。

3.4.2 文の好意悪意度合い表示

表示例を図 2 の②に示す。文章の各文がどの程度読み手に好意を与えるか、また悪意を与えるか、その度合いを表示する。3.3 節で構築した構築済みの深層学習モデルである好意モデル、悪意モデルによって各文が好意を含む文として判定される確率、悪意を含む文として分類される確率を算出し、それぞれ薄い赤色と薄い青色の棒グラフで可視化する。好意と悪意を表す棒グラフの原点はそれぞれ左右の端であり、100 %の点は棒グラフの中央である。例えば、図 2 において②を見ると、3 文目、4 文目、5 文目の悪意度合いがそれぞれ 96 %、99 %、99 %となっている。これは、3 文目、4 文目、5 文目が読み手に悪印象を持たれる可能性が高い文となることを表している。このように各文が読み手にどんな印象を与えるかを視覚的に表示することによって、文章推敲を促す。

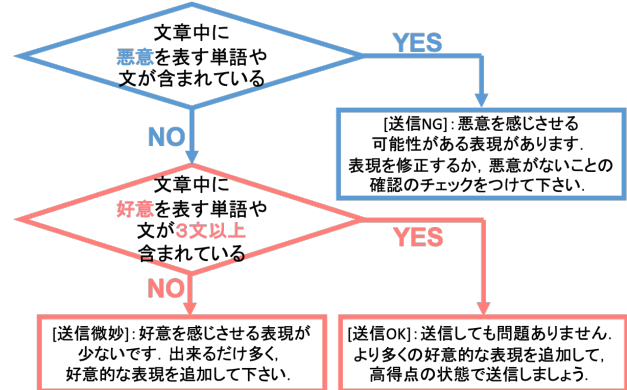


図 3: メッセージ表示までの流れ

3.4.3 文章推敲を促すメッセージ表示

表示例を図 2 の③に示す。この機能では文章の内容に応じて図 3 の流れに従って、ユーザに文章推敲の指針をメッセージとして表示する。これらのメッセージによって文章推敲を促す。

3.4.4 文章のスコア表示

表示例を図 2 の④に示す。文章の内容を点数化、また可視化表示することで、現在の文章を定量的に評価する。3.4.1 節で抽出した好意と悪意を表す単語の含まれる文の数、また、好意と悪意の表現を含む文の数を受けて、全文中における割合をパーセンテージで算出し、それぞれ薄い赤色と薄い青色の棒グラフで可視化する。好意と悪意を表す棒グラフの原点はそれぞれ左右の端であり、100 %の点は棒グラフの中央である。また、好意を表す単語が含まれる、または好意の表現を含む文の割合が 50 %の時に 100 点になるようにスコアを算出し表示している。文章中の色付きの単語やハイライトしている文の割合を視覚的に表現して、ユーザの文章推敲を促す。

3.4.5 文章の編集機能

この機能は以下の 3 つの機能に分かれている。

- 文章を編集する機能。図 2 の①の領域で行うことができ、文章を 1 文ずつ編集することが出来る。文末に全角の「。」または「。」を追加することで文の数を増やすことが出来る。
- 悪意を表す単語や文の推敲を促すチェックボックス機能。この機能は図 2 の⑤の領域で行うことが出来る。悪意を表す単語が含まれる文、または悪意の表現を含む文の先頭のチェックボックスに

チェックをすることで、チェックされた文を悪意を表す単語、また文として扱わず、図2の④の棒グラフを再描写する。悪意を表す単語として表示された単語、または文について、推敲しなくしてもそのままの表現で問題ない場合にチェックボックスにチェックをすることで、問題ないことを確認する。文章推敲を促すメッセージ表示機能との連携によって、ユーザに悪意を表す単語や文の確認を促すことが出来る。

- システムの再描写機能。この機能は図2の⑥の領域で行うことが出来る。編集した文章をシステムに再入力するボタンである。再入力する際に、チェックボックスへのチェックがリセットされるため、編集内容に応じた再確認を促すことが出来る。

これらの機能によってユーザに文章推敲を促し、読み手に好感持たれる文章の推敲を支援する。

4 文章推敲支援機能の効果の検証

4.1 実験目的

本実験では、文章推敲支援インタフェースにおける深層学習を用いた好意と悪意の表現を含む文の可視化を利用することによって、文章中の読み手に好感を持ってもらえる文章の推敲を支援出来るかを評価する。具体的には、悪意を含む表現を文章中から削除し、好意を含む表現を文章中により多く追加することで、読み手の印象を良くするという文章推敲の支援目標を達成できるかを評価する実験を行った。

この実験目的を達成するために、以下の2種類の実験を行なった。1つは、4.2節で述べる文章推敲支援システムを用いた文章作成実験、もう1つは、4.3節で述べる推敲された文章の内容評価実験となる。それぞれに別の被験者を用意し、実験を行なった。

4.2 文章推敲支援システムを用いた文章作成実験

4.2.1 文章推敲支援システムを用いた文章作成実験の実験方法

表4に提示する4つの状況において、指定の目的を達成するメールを、独力で作成してもらった。ここで、状況1、3は書き手の悪意が表れやすい状況、状況2、4は書き手の好意が表れやすい状況として用意した。それぞれの状況において、システムを使用する事で、書き手の悪意を表す表現が削除され、好意を表す表現が

表 4: メール作成課題の状況

	送信相手	送信相手について	送信したいメール
1	同じサークルの同性の先輩	自己中心的な性格で普段からあなたを含め周りの人のことを全く考えない発言や行動をする先輩。	あなたが企画したサークル旅行について先輩が楽しめる企画が少ないから計画を大幅に変更するようにと自分勝手な事を言われた。あなたの周りのメンバーは、計画に概ね同意している。そこで先輩に自分勝手な主張はやめてもらい普段から皆困っていると文句を言って欲しいとメンバーから言われている。
2	同じ学科の異性の先輩	成績優秀でいつもやさしく、先生や周りの人からの評価も高くあなた自身も憧れている先輩。	期末試験に向けて、勉強でわからないところを質問したら個人的に時間を取ってくれて、丁寧にわかりやすく教えてもらったので感謝のお礼を伝えたい。また先輩とずっと仲良くなりたい。
3	同じ研究室の同性の後輩	自分を含め、後輩から見て先輩にあたる人達に、普段からタメ口で文句を言うなど、研究室の多くの人からその態度や言葉遣いが問題視されている。	研究室の研究報告会で、質問者である自分がした指摘に対して不満そうな顔をした上で、ぶっきらぼうに「それはいい」と言った。先生からはその指摘は正しいという言葉も貰っている。報告会での態度を謝るように、また普段からも最低限の礼儀をわきまえるよう指導したい。
4	同じサークルの気になる異性	普段から仲良く話をする相手。とても気が合って数人で一緒に遊びに行く事もある。服装にこだわるおしゃれさん。	次の土曜日に、普段のおしゃれさから自分の服装にもアドバイスをもらいたいという話から入り自分と二人で遊びに行くように誘いたい。

多く追加されることを期待している。次に、作成したメールを「悪意を含む表現を使わない、また好意を含む表現を増やしたメールに修正してください」という指示のもと、独力で推敲してもらった。その後、独力で推敲したメールを提案システムに入力してもらい、システムを用いて同様の指示のもと、再度メールの推敲を行ってもらった。これを11名の理系学生を被験者として行った。ここで、独力で推敲したメールを「システム使用前のメール」とし、システムを用いて推敲した後のメールを「システム使用後のメール」と呼ぶ。

4.2.2 文章推敲支援システムを用いた文章作成実験の結果と考察

まず、文章の修正がシステムのどの表示による修正か確認するために、システム使用前後で修正が行われた文について、その修正方法を以下のように区別する。好意と悪意を表す単語を追加、削除することでの修正を「単語単位での修正」と呼ぶ。また、好意と悪意を表す単語の追加、削除以外での文単位での修正を「文単位での修正」と呼ぶ。文単位での修正は、3.4.1項で

前述した深層学習による好意や悪意の表現を含む文のハイライト表示を用いて修正された可能性が高いと考えられる。単語単位での修正、文単位での修正どちらも行われた場合は「単語と文による修正」と呼ぶ。

各被験者がシステムを用いて修正した文について、その修正方法の内訳を確認するために、各状況毎のシステム使用前後で修正された文の内訳を図4に示す。図4からわかるように、状況1、3、4ではシステムを用いることで全文中半数以上が修正がされており、状況2においても27%程度の文が修正されている。また、修正の内訳のうち、文単位の修正、単語と文での修正が大多数を占めている。これにより、システムの深層学習による好意や悪意の表現を含む文のハイライト表示を用いて文章のおおよそ半数程度の文修正するように文章推敲を促すことができたと考えられる。

また、修正が行われた文のうち、修正によって深層学習の結果によるハイライト表示がどのように変化したか確認するため、各状況毎のシステム使用前後での修正によるハイライト表示の変化の内訳を図5に示す。ここで、「修正でハイライトが改善しなかった文の数」とは、修正された文のうち、修正によって新たに好意ハイライトが表示された文、悪意ハイライトが消えた文以外の文を指す。図5より、各状況において、修正された文のうち、およそ半数以上の文で新たに好意のハイライトが表示されるか悪意のハイライトが消えている。特に、悪意の表れやすい状況である状況1、状況3では修正によって悪意のハイライトが多く消えている。また、好意の表れやすい状況である状況2で好意のハイライトが多く追加されている。これより、深層学習による好意や悪意の表現を含む文のハイライト表示を用いて好意のハイライトが表示される文を増やし、悪意のハイライトが表示される文を減らすように文章推敲を促すことができたと考えられる。

4.3 推敲された文章の内容評価実験

4.3.1 推敲された文章の内容評価実験の実験方法

4.2節の実験によって推敲されたメールが、自分宛てに送られてきたという仮定で、メールの送り主にどの程度好感を持ったか、理由とともに7点満点で評価してもらう。評価点と評価基準を表5に示す。また、システム使用前とシステム使用後のメールの各文を、どちらがシステム使用後か明示せずに、どちらに好感を持ったかを評価してもらう。これを4.2節の実験で作成された11名(被験者数)×2種類(システムなしでの推敲とシステム有りでの推敲)×4種類(状況)=88種類のメールに対して行ってもらった。この実験を8名の理系大学生を評価者として行った。

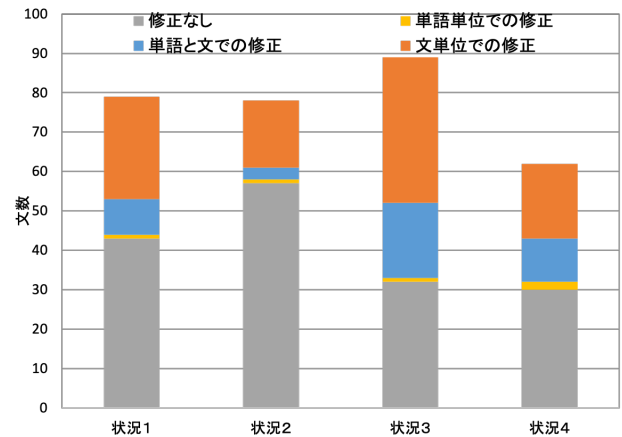


図4: 各状況毎のシステム使用前後で修正された文の内訳

表5: メールの評価点と評価基準

点数	評価
7	メールの送り主に、強く好感が持てる。
6	メールの送り主に、割りと交換が持てる。
5	メールの送り主に、少し好感が持てる。
4	メールの送り主に、特になんとも思わない。
3	メールの送り主に、少しだけ好感が持てない。
2	メールの送り主に、あまり好感が持てない。
1	メールの送り主に、全く好感が持てない。

4.3.2 推敲された文章の内容評価実験の結果と考察

システム使用前後で読み手により好感を持ってもらえる文章に修正されたか検証するために、各状況におけるシステム使用前と使用後のメールの評価点の被験者毎の平均を図6に示す。ここでの評価点の差とは、システム使用後のメールの評価点から使用前のメールの評価点を引いたものとする。本実験では、既に一度推敲したメールを、システムによってより読み手に好感を持たれるメールに修正していることに注意されたい。

図6より、ほとんどの被験者がシステム使用前に比べてシステム使用後の評価点が高いことがわかる。特に、状況1、状況3の被験者毎の評価点の差の平均はそれぞれ1.0、0.9となっている。これは、状況1、状況3がそれぞれ「先輩に文句を言う」状況と「後輩に注意する」状況であるため、深層学習によって、明示的に悪意を表す単語が使われないが、悪意を表す文をハイライトすることで文章修正が行われたためだと考えられる。これより、悪意を含む表現が使われやすい状況においては、システムを用いる事でおよそ1段階程度の評価をあげる程度の文章の推敲を支援できることがわかる。しかし、状況2、状況4に関しては評価点の差があまり大きくなかった。ここで、状況2、状

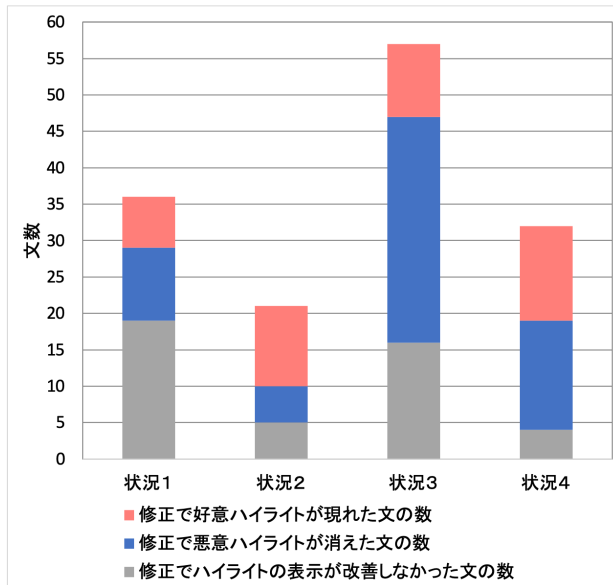


図 5: 各状況毎のシステム使用前後での修正によるハイライト表示の変化の内訳

況4はそれぞれ「先輩にお礼を言う」状況と「異性をデートに誘う」状況となっており、共に好意を含む表現が表れやすい状況となる。このため、システム使用前の評価点が悪意を含む表現が表れやすい状況に比べて1点ほど高く、独力でもある程度好意的な文になっていたと考えられる。全体として、評価はシステム使用前後で上昇した。特に、悪意を含む表現が発生しやすい状況では、深層学習を用いた好意と悪意の表現を含む文の可視化に意味があるとわかった。

システム使用前後で評価点が大きく上がったまたは下がったメールについて、システム使用前後のメールを図7、図8に示す。ここで、図7、図8ではシステム使用前後での差異を赤字で表示している。また、深層学習の結果によるハイライト表示で好意のハイライトがされた文を薄い赤色で、悪意のハイライトがされた文を薄い青色で塗りつぶしている。

図7は状況1における被験者Eのシステム使用前後のメールと好意悪意度合いの差を表している。被験者Eは、悪意ハイライトのされた「苦情が出ています」という文を修正して「相談をいくつか受けています」という表現にすることで、悪意ハイライトを消し、より読み手に悪意を与えない文にした。また、「ご理解いただければ幸いです」という表現を「ご理解いただければ幸いです」という表現に修正し、好意ハイライトを追加することで、より柔らかい印象を受ける文にした。この修正によって評価を4.0点から5.4点にあげている。

図8は状況1における被験者Aのシステム使用前後

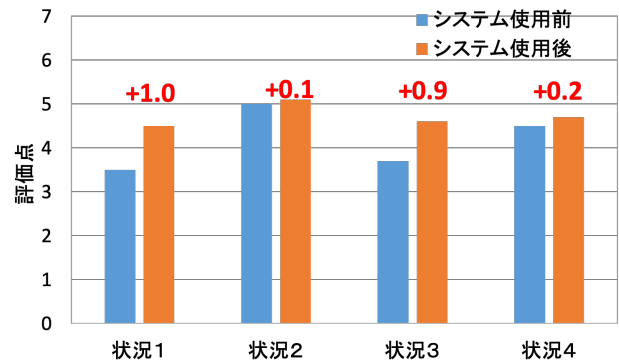


図 6: 各状況ごとのシステム使用前後での評価点の平均と差

のメールと好意悪意度合いの差を表している。被験者Aは、修正前後でほとんど修正をしておらず、好意を含む文も追加していない。語尾に「ね」をつけた修正を行っているが、これによって上から目線の文章になっており、評価点は4.6点から4.0点に下がっている。

全体として、評価点が大きく上昇したメールでは、悪意を含む表現を無くし好意を含む表現を追加する修正が多かった。これにより、システムを用いて悪意を含む表現を無くし好意を含む表現を追加する事で、読み手に好感を持たれる文章を作成できると考えられる。また、修正を促された時に、どのように修正したらよいかわからない場合もあると考えられる。

5 おわりに

読み手に好感持たれる文章推敲支援を行うために、深層学習を用いて文章の書き手の読み手に対する好意、悪意の表現を含む文を抽出して可視化し、書き手に文章推敲を促すシステムを実装した。

また、実装した文章推敲支援システムが、読み手に好感を持たれる文章の推敲に有用であることを検証するために評価実験を行った。ある状況において作成したメールを、一度独力で推敲したうえで文章推敲支援システムを利用して再度推敲してもらい、システム使用前後の推敲されたメールを別の被験者に評価してもらった。この実験より、文章推敲支援システムを用いて文章推敲する事で、より読み手に好感持たれる文章を作成できることがわかった。

今後の展望としては、修正が望まれる箇所に対して、修正案を提示できるシステムへの改良が期待される。

システム使用前(評価点4.0)	システム使用后(評価点5.4)
〇〇先輩、いつもお世話になっております、XXです。	〇〇先輩、いつもお世話になっております、XXです。
(中略)	(中略)
また、この場をお借りして申し上げますが、以前からサークル内で先輩の振る舞いに対し、 少し苦情が出ています。	また、この場をお借りして申し上げますが、以前からサークル内で先輩の振る舞いについて 相談をいくつか受けています。
(中略)	(中略)
重ね重ね失礼なことを述べましたが、 どうかご理解いただければと思います。	重ね重ね失礼なことを述べましたが、 どうかご理解いただければ幸いです。
以上、どうかよろしくお願いします。	以上、どうかよろしくお願いします。

図 7: 各状況 1 における被験者 E のシステム使用前後のメール (評価点 1.4 上昇)

システム使用前(評価点4.6)	システム使用后(評価点4.0)
〇〇様、お世話になっております、××です。	〇〇様、お世話になっております、××です。
先日、ご連絡いただいた企画変更につきまして、部内で検討した結果、 変更せずに企画を行う事となりました。	先日、ご連絡いただいた企画変更につきまして、部内で検討した結果、 変更せずに企画を行う事となりました。
大変申し訳ありませんが、 ご理解の程よろしくお願いいたします。	大変申し訳ありませんが、 ご理解の程よろしくお願いいたします。
次回、企画するときは一緒に考えましょう！	次回、企画するときは一緒に考えましょう ね ！
よろしくお願いします。	よろしくお願いします ね 。

図 8: 各状況 1 における被験者 A のシステム使用前後のメール (評価点 0.6 下降)

謝辞

本研究における深層学習を用いた好意と悪意の表現を含む文の抽出のために株式会社ホットリンクより、大規模日本語 SNS コーパスによる文分散表現モデル hottoSNS-w2v[9] を利用させて頂いたことについて感謝する。

参考文献

- [1] 総務省: ネット依存など新たな課題とインターネットリテラシーの重要性, 総務省平成 26 年版情報通信白書, pp.283-303 (2014)
- [2] 庵 翔太, 砂山 渡, 畑中 裕司, 小郷原 一智: 良好な人間関係構築のための好意と悪意を表す単語の可視化による文章作成支援, 第 32 回人工知能学会全国大会論文集, 3F2-OS-12b-03 (2018)
- [3] 板倉 由知, 白井 治彦, 黒岩 丈介, 小高 知宏, 小倉 久和: 様々な文書を対象とした段落一貫性の解析, 情報処理学会研究報告自然言語処理 (NL), Vol.192, No.9, pp.1-6 (2009)
- [4] 大野 博之, 稲積 宏誠: 文の接続関係を利用した論理方向の可視化による技術文章作成支援, 電子情報通信学会技術研究報告, Vol.107, No.48, pp.27-32 (2015)

- [5] 徳岡 拓弥, 竹内 章, 國近 秀信: 変更履歴を用いた英語文章作成支援システムの実現, 第 30 回人工知能学会全国大会論文集, 1C3-1 (2016)
- [6] 石川 葉子, 水上 雅博, 吉野 幸一郎, Sakti Sakriani, 鈴木 優, 中村 哲: 感情表現を用いた説得対話システム, 人工知能学会論文誌, Vol.33, No.1, pp.1-9 (2018)
- [7] 長谷川 貴之, 鍛冶 伸裕, 吉永 直樹, 豊田 正史: オンライン上の対話における聞き手の感情予測と喚起, 人工知能学会論文誌, Vol.29, No.1, pp.90-99 (2014)
- [8] Twitter.com, (URL)<https://dev.twitter.com/docs/api/>
- [9] 日本語大規模 SNS+Web コーパスによる単語分散表現モデルの公開: hottoSNS-w2v の配布, (URL)https://www.hottolink.co.jp/blog/20190311_101674/
- [10] Total Environment for Text Data Mining (テキストデータマイニングのための統合環境), (URL)<https://tetdm.jp>
- [11] 砂山 渡, 高間 康史, 西原 陽子, 徳永 秀和, 串間 宗夫, 阿部 秀尚, 梶並 知記: テキストデータマイニングのための統合環境 TETDM の開発, 人工知能学会論文誌, Vol.28, No.1, pp.1-12 (2013)

ギャップにより面白さを生むストーリーの生成に向けた単語の自動補完

Automatic Word Completion for Creating Interesting Story with Gaps

河村 芽依^{1*} 砂山 渡¹ 畑中 裕司¹ 小郷原 一智¹

Mei KAWAMURA¹ Wataru SUNAYAMA¹ Yuji HATANAKA¹ Kazunori OGOHARA

¹ 滋賀県立大学工学部

¹ School of Engineering, The University of Shiga Prefecture

Abstract: In everyday life, there are many opportunities to communicate with people. Sometimes I want to make a story that is interesting and interesting to others. The purpose of this research is to create a system that presents a story to be created with gaps of fun to support story generation. The story is created by using the relevance of the word and acquiring the related word of the input word and combining it with the prepared template sentences. The texts prepared by the system were evaluated from two viewpoints, fun and betrayal of expectations, and a system evaluation experiment was performed by asking what kind of correction could be made to make it even more interesting. Experiments have confirmed that the generated story can generate a certain amount of fun. In addition, after generating a story with the system, we propose more specific words as correction candidates, and we want to aim for a more satisfactory story by selecting final words by humans.

1 はじめに

日常生活では、人とコミュニケーションを取る機会は多く、親しい関係の人との会話ややりとりの中では、相手に話への興味や関心を持って、面白いと思ってもらえる話を作りたいと考える。相手に自分の思うことを伝えようとする時、どのようなことをどのような流れの内容か、自分の発想力やひらめき、知識を基に考えることを求められる。しかし、相手に伝えたいことはあるが、どんな言葉で表せばよいかうまく思い浮かばず、伝えられないこともある。そもそも、相手に面白いと思わせることを、難しいと思う人もいる。そのような場面で、ストーリー生成の支援をシステムによって行いたいと考える。

そこで本研究では、予想とのギャップによって面白さを生むストーリーをテンプレート内の単語の自動補完によって作成するシステムの構築を目的とする。単語を、用意したテンプレートに自動で補完することによって、ストーリーを形成し、そのストーリーを提案することで、支援を行う。本研究では、期待はずれの場面で起こる不調和のタイプの面白さについて扱う [1]。

形成するストーリーは、面白さを生み出す内容にすることによって、相手の興味や関心を得ようとしている。また、この提案を行うシステムを作成し、実際に日常的なやりとりの中で利用したり、会話の想定や練習の機会として、利用できるようにしたい。

以下本論文では、2章で関連研究について述べる。3で単語の自動補完によるストーリー生成システムについて述べる。4章で本システムの評価実験について述べ、5章で本論文を締めくくる。

2 関連研究

ボケとツッコミの対話形式で笑いを生み出す文章自動生成システムに関する研究がある [2][3]。予測される事柄をわざと誤ることでおかしさを生み出しており、誤りの単語は Google N-gram を用いて、入力文の前後を考慮している。また、文章を入力して、その一部の単語を面白みのある単語に変換することで、漫才の台本となる文章の生成を行っている。

面白いフレーズの自動生成に関する研究がある [4]。この研究では、身近にあることわざを利用して単語を置き換えることにより、ユーザの予想を裏切って笑いを生

*連絡先：滋賀県立大学工学部 電子システム工学科 河村 芽依
〒522-8533 滋賀県彦根市八坂町 2500
E-mail: os23mkawamura@ec.usp.ac.jp

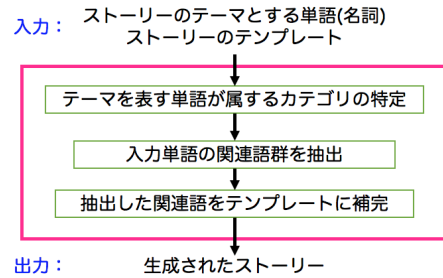


図 1: 単語の自動補完によるシステムの構成

み出している。ことわざを利用することによって、次はこうなるという予想を容易にさせることができ、音の長さや音韻の類似などを考慮して、単語を変換している。

テーマとキーワードからアイデアの発想を支援するシステムに関する研究がある [5]。この研究では、テーマやキーワードをそのまま発想支援の入力とすると、その要素に固執した発想しか得られない可能性を指摘している。これを解決するために、入力とするキーワードの共起関係を用いて関連する語を取り出し、関連性はあるが、共起度は低い要素を抽出し、発想のヒントとして提示している。このように、人のアイデアでは、発想し難い単語を取り出している。

本研究では、入力としてユーザが指定する単語のみとし、作ることができるストーリーの多様性を考慮する。また、単語の音や構造を考慮するのではなく、単語の持つ意味や状態などの情報を考慮して、意外性のある単語を取り出す手法を試みる。

3 単語の自動補完によるストーリー生成システム

3.1 システムの構成

本研究で提案する単語の自動補完によるストーリー生成システムの構成を図 3 に示す。

まず入力として、ストーリーのテーマとする名詞の単語と、ストーリーのテンプレートを受け取る。入力された単語がどのカテゴリに属しているか特定を行い、カテゴリを決定する。これらの入力された単語と特定されたカテゴリを基に、関連語を抽出し、あらかじめ用意したテンプレートに補完することによってストーリーを生成する。

3.1.1 ストーリーのテーマとする単語

入力として、ストーリーのテーマを表す 1 単語を受け取る。本研究では、話題の中での利用のしやすさを考

表 1: ストーリー生成用テンプレート

A 「[特徴 1(形容詞)][カテゴリ]って何やと思う？」
B 「一番に思いつくのは、[入力単語] やなあ。」
A 「でもそれは、[特徴 2(動詞)] ものではないねん。」
B 「なら [単語 A] やろ。」
A 「いや、[特徴 3(動詞)] もでもないねん。」
B 「じゃあそれは一体なんやねん。」
A 「実は、[単語 B] なんや。[特徴 4(形容詞)] ものやねん。 [入力単語] や [単語 A] より、[単語 B] の方が [特徴 4(形容詞)] やん。」
A 「[特徴 1(形容詞)] ものって何やと思う？」
B 「それで思いつく [入力単語のカテゴリ] は [入力単語] なんちゃう。」
A 「いや、[入力単語とは異なるカテゴリ] のことやから [単語 C] やねん。 [単語 C] こそが、[特徴 1(形容詞)][入力単語とは異なるカテゴリ] や。」
A 「友達が好きな [カテゴリ] があるらしんやけど、 その名前を忘れたみたいやねん。 いろいろ聞いても、全然名前がわからへんねんて。」
B 「じゃあ、その友達の好きな [カテゴリ] を考えてあげるから、 特徴を教えてよ。」
A 「確か [特徴 1(形容詞)] もので [特徴 2(動詞)] って言ってた。」
B 「[入力単語] じゃない、それ？ その特徴は完全に [入力単語] や。」
A 「自分もそうやと思ってんけど、[特徴 7(形容詞)] ものらしいねん。」
B 「なら、[入力単語] ではないな。 [入力単語] は [特徴 7(形容詞)] ものじゃないもんね。」
A 「それで [特徴 5(形容詞)] もので [特徴 6(動詞)] っのも言ってた。」
B 「それはやっぱり [入力単語] やろ？ その特徴はやっぱり [入力単語] や。」
A 「でもな、[特徴 8(動詞)] ものらしいねんて。」
B 「なら、[入力単語] ではないな。 [特徴 7(形容詞)] ので [特徴 8(動詞)] なら、もう [単語 D] でええやろ。」

慮して、日常生活で現れやすい単語のカテゴリを 8 つ用意し、そのカテゴリのいずれかに当てはまる単語を入力する。本研究では用意したカテゴリは、洋菓子、動物、スポーツ、植物、食べ物、遊び、日用品、施設の 8 つとなっている。

3.1.2 ストーリーのテンプレート

ストーリーのテーマとする入力単語と、その関連語を用いてギャップが生じるストーリーとしてテンプレートを 3 つ用意した。表 1 に用意した 3 つのテンプレートを示す。

これらのストーリーは、入力された 1 単語をテーマとして、連想ゲームでの「〇〇といえば」というやりとりをする内容となっている。例えば、「甘いもの」といえば何だろう、といった時、マカロンやクッキーなどが想像できる。しかし、実際は別のものだった、という形を利用して、マカロンやクッキーといった予想を裏切っている。

各テンプレートは、予想を裏切るパターンが異なり、使用する関連語および使い方によりパターンを形成している。テンプレート 1 は、単語を表す特徴に着目したやりとりを行い、テンプレート 2 は、単語のカテゴリに着目したやりとりを行い、テンプレート 3 は、入力された単語に着目したやりとりを行っている。

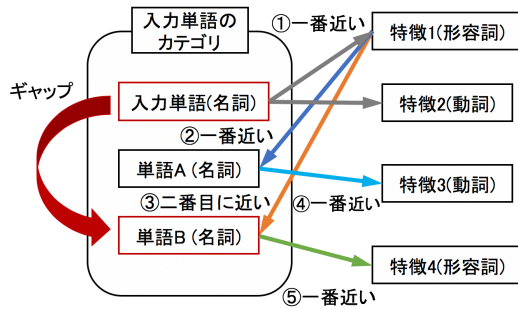


図 2: テンプレート 1 に使用する関連語

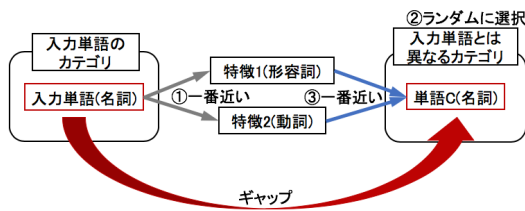


図 3: テンプレート 2 に使用する関連語

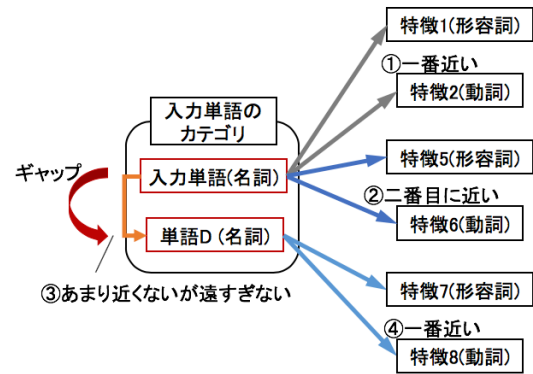


図 4: テンプレート 3 に使用する関連語

ストーリー生成結果

A 「【凄い】【スポーツ】って何やと思う？」
 B 「一番に思いつくのは、【野球】やなあ。」
 A 「でもそれは、【投げる】ものではないねん。」
 B 「なら【カバディ】やる。」
 A 「いや、【鍛える】ものでもないねん。」
 B 「じゃあそれは一体なんやねん。」
 A 「実は、【気球】なんや。【熱い】ものやねん。
 【野球】や【カバディ】より、【気球】の方が【熱い】やん。」

図 5: 入力単語「野球」をテンプレート 1 に補完した時の生成例

3.2 単語の表現

本研究では、単語の分散表現を利用して単語間の関係を比較する。単語の分散表現を得る手段として fast-Text¹を利用している。単語のベクトルを 300 次元とし、インターネット上での百科事典と言える Wikipedia²の全データの文章を利用して、学習を行う。

単語ベクトルを取得できることによって単語間の cos 類似度を求めることができ、単語同士の意味的な関係を捉える。

3.3 関連語の導出

それぞれのテンプレートに使用する関連語の関係図を図 2, 図 3, 図 4 に示し、テンプレート 1 の関連語の抽出の順序を以下に示す。

1. 入力単語との cos 類似度が最も大きい形容詞 (特徴 1) および動詞 (特徴 2)
2. 入力単語のカテゴリ内の単語で、特徴 1 の形容詞との cos 類似度が一番大きい単語 (単語 A)
3. 特徴 1 の形容詞との cos 類似度が 2 番目に大きい単語 (単語 B)
4. 単語 A との cos 類似度が最も大きい動詞 (特徴 3)

¹<https://fasttext.cc>

²<https://ja.wikipedia.org/wiki/メインページ>

5. 単語 B との cos 類似度が最も大きい形容詞 (特徴 4)

この順序で関連語を全て抽出することができたら、テンプレートの決められた箇所にそれぞれの単語を当てはめて、補完を行う。また、赤色で囲まれた単語 (入力単語, 単語 B) を裏切ることによってギャップを生み出している。

3.4 システムの出力例

図 5 に、入力単語を「野球」としてテンプレート 1 に補完した時のストーリー生成結果を示す。野球という単語から、カテゴリを「スポーツ」と特定している。ギャップを生み出しているのは、「野球」ではなく「気球」だったという単語の裏切りである。

3.5 システムのインターフェース

この単語の自動補完によるストーリー生成システムのインターフェースは Python で扱っている。本システムを作成するにあたり、Python からインターフェースを

構築するための標準ライブラリである Tkinter を使用している。単語を入力および、テンプレートを選択すると、用意したカテゴリに入力単語が含まれている場合、カテゴリを特定する。この場合、含まれるカテゴリが複数ある場合は、ユーザーに選択してもらう。入力単語を基に、形容詞および動詞の $\cos$ 類似度が高い 10 単語を表示させる。そして、入力単語から各関連語を抽出し、選択されたテンプレートに抽出した関連語を当てはめることによって生成されたストーリーも表示させる。

4 ストーリーの生成に向けた単語の自動補完システムの評価実験

本研究で提案したシステムの評価をするため、評価実験を行った。その方法と実験結果を示す。

4.1 実験方法

評価実験の方法として、まず単語を自動補完しストーリーを生成するシステムを用いて、3つのテンプレートを使ってそれぞれ5つずつ、全部で15個のテキストを用意した。そして、被験者として大学生、大学院生11名にそれぞれのテキストを読んでもらい、5段階の評価を行ってもらった。評価してもらう項目は、AとBの会話のやりとりを面白いと感じたか、AとBのどちらかが予想を裏切られているかの2つである。これらの項目から、本システムでギャップを生むことができているか、それによりおもしろさを生むことができているのかを検証する。また、用意されたテキストをより良くするための修正を1単語単位で行ってもらった。これにより、さらに面白いストーリーに作り変える形で、実際にストーリー生成に向けた支援ができているのか、検証する。

4.2 面白さと裏切りの評価についての結果

表2に、評価実験に用いた15のテキストの面白さおよび裏切りの評価の順位を示す。この表より、面白さの評価について、全テキストにおいては、テンプレート1,3のテキストが上位を占めているのに対して、テンプレート2のテキストは下位を占めていることがわかる。裏切りの評価についても同様に、テンプレート1,3のテキストが上位を占めているのに対して、テンプレート2のテキストは下位を占めていることがわかる。特にテンプレート2は文章量も他の二つに比べて少ないため、情報量や予想を裏切る度合いも小さく感じられたことも考えられる。面白さの評価に比べて、裏切りの評価の方が全体的に高くなっている。よって、裏切りの評価が高いにも関わらず、面白さの評価が低いものもある。こ

表 2: 15 の文章の面白さおよび裏切りの評価の順位

面白さの評価			裏切りの評価		
順位	テキスト	評価点	順位	テキスト	評価点
1	3-4	3.82	1	1-1	4.55
2	1-3	3.55	2	1-2	4.36
2	1-5	3.55	3	1-3	4.27
4	3-1	3.45	4	2-1	4.27
5	3-3	3.36	5	2-4	3.82
6	2-1	3.27	6	3-1	3.82
7	3-5	3.09	7	3-4	3.82
8	1-1	2.91	8	2-2	3.73
8	3-2	2.91	9	3-3	3.73
10	1-2	2.82	10	3-5	3.64
10	1-4	2.82	11	1-5	3.55
10	2-4	2.82	12	1-4	3.18
13	2-2	2.64	13	3-2	3.18
13	2-5	2.64	14	2-3	2.45
15	2-3	2.55	15	2-5	2.36

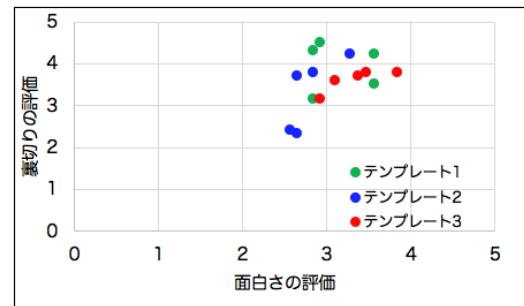


図 6: 面白さと裏切りの評価の関連性

れらより、ただ裏切るだけではなく、ギャップを生む名詞の単語間の関係性は一定の距離を考慮するための新たな抽出条件が必要と考えられる。

図6に、面白さと裏切りの評価の分布を示す。本システムは、ギャップを生み出して面白さを感じさせるため、面白さと裏切りの評価の間には、比例の関係である必要がある。また、本実験での二つの評価に対する相関係数は0.36であった。このことから、面白さと裏切りの評価の間には弱い正の相関があるといえる。本研究で用いたテンプレートでは、予想が裏切られたテキストは面白さを生み出しているといえる。

実験で使ったテキストの中で、最も裏切りが面白さにつながった例を、図7に、面白さの評価と裏切りの評価を実験者ごとに棒グラフで示す。このテキストは表3に示すテキスト3-3となっている。実験者1, 8, 9を除いた実験者は、裏切りの面白さの評価が同等であることがわかる。「おもしろい」「遊ぶ」「転がす」「楽しい」の特徴から「ピリヤード」しかない、形容詞と動詞の関連語により、ギャップに関する名詞の単語を強く予想できていると考えられ、面白さにつながったと言える。また、「雪合戦」は「遊ぶ」「楽しい」といったピ

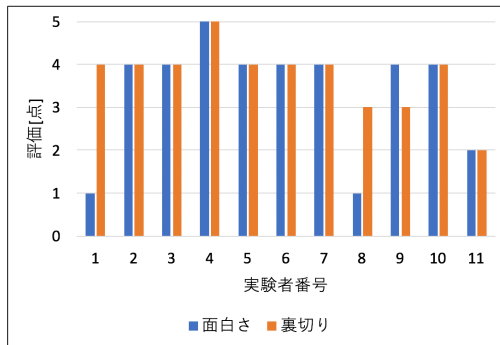


図 7: 裏切りが面白さにつながった例

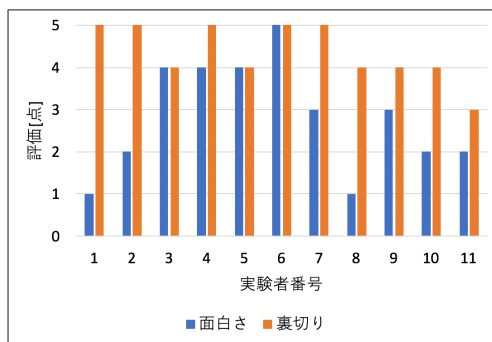


図 8: 裏切りが面白さにつながらなかった例

リヤードの特徴にも当てはまっているため、ビリヤードと雪合戦は遠い関係ではないといえる。

実験で使ったテキストの中で、最も裏切りが面白さにつながらなかった例を、図 8 に、面白さの評価と裏切りの評価を実験者ごとに棒グラフで示す。このテキストは表 3 に示すテキスト 1-2 となっている。裏切りの評価は全体的に高いが、面白さの評価は低いものが多いことがわかる。丸い動物で連想をしていき、特徴を否定されていくというストーリーだが、最後にジャガーときて裏切りはあるものの、はじめの「丸い動物」に対して違和感を感じて面白さの評価が低くなってしまったと考える。

よって、これらのことより、ギャップに関する二つの単語は一定の距離の関連度を考慮して抽出することが重要と考えられる。本システムにおいて、関連語を抽出する際、 $\cos$ 類似度が大きいといったルールだけではなく、近すぎず遠すぎないといった、抽出条件を増やす必要があると考えられる。

表 3: 評価実験に用いたテキスト 1-2, テキスト 3-3

テキスト	
1-2	A「丸い 動物 って何やと思う？」 B「一番に思いつくのは、アザランやなあ。」 A「でもそれは、泳ぐ ものではないねん。」 B「なら、ブタ やろ。」 A「いや、飼う もでもないねん。」 B「じゃあそれは一体なんやねん。」 A「実は、ジャガー なんや。速い ものやねん。」 アザラン や ブタ より、ジャガー の方が 速い やん。」
3-3	A「友達が好きな スポーツ があるらしんやけど、その名前を忘れたみたいやねん。いろいろ聞いても、全然名前がわからへんねんて。」 B「じゃあ、その友達の好きな スポーツ を考えてあげるから、特徴を教えてよ。」 A「確か おもしろい もので 遊ぶ って言ってた。」 B「ビリヤード じゃない、それ？その特徴は完全にビリヤード や。」 A「自分もそうやと思ってんけど、痛い ものらしいねん。」 B「なら、ビリヤード ではないな。」 ビリヤード は 痛い ものじゃないもんね。」 A「それで 転がす もので 楽しい ってのも言ってた。」 B「それはやっぱり ビリヤード やろ？その特徴はやっぱり ビリヤード や。」 A「でもな、滑る ものらしいねんて。」 B「なら、ビリヤード ではないな。」 痛い ので 滑る なら、もう 雪合戦 でええやろ。」

4.3 文章の修正の結果

用意したテキストを面白くするための修正をしてもう場合と、裏切るための修正をしてもう場合の二つを行った。修正されたパターンとして、同じ品詞の同カテゴリの単語、同じ品詞のより具体的な単語、の 2 つに分けることができる。表 4 に、面白くするための修正案、および裏切るための修正案を 2 つのパターンに分けた平均件数を示す。

面白くするための修正案と裏切るための修正案の件数が最も多いのは、テンプレート 2 であることがわかる。テンプレート 2 の評価において、面白さ、裏切り共に平均も低かったため、修正案も多くなったと考えられる。全体的に面白さの観点の方が、裏切りの観点からの修正案の件数が多かった。これは、面白さの評価が裏切りの評価より低かったため、面白くするための修正案を促されたと考えられる。

修正する観点について、関連語を同じ品詞の別の単語に言い換える修正が多いことがわかる。また、変換された名詞に関しては、より具体的な単語に言い換える修正が多く、ストーリーが具体的に変わり、確かに面白くなるといえる。具体的な単語に置き換えることによって、よりイメージしやすく、裏切りが強く感じられる。これらより、本システムにおいて、関連語を抽出し

表 4: 修正案の平均件数

修正する観点 テンプレート	面白さ			裏切り		
	1	2	3	1	2	3
同カテゴリの単語 [件]	3.8	4.2	5.6	3.4	4.4	2.2
具体的な単語 [件]	3.4	4.4	2.2	2.0	3.0	2.4

た後, さらにギャップに関する関連語の名詞の具体的な単語を, 取り出すことを可能にしたいと考える.

5 おわりに

日常的な会話やメッセージのやりとりにおいて, 面白さを生むストーリーの生成に向けた単語の自動補完により支援を行うシステムを作成した.

入力単語 1 単語から, 関連語を取り出し用意したテンプレートに当てはめてストーリーを作成するシステムを作成した. 作成したテキストが面白いと感じられるか, また裏切られていると感じられるかを評価するための実験を行った. その結果, 単語の自動補完により, ギャップによる面白さを生むストーリーを提示できることが確認できた. また, 単語を裏切る際には, 二つの単語間の意味は近すぎず遠すぎない単語を抽出する条件を増やす必要があると考えられる. さらに予想を裏切るために, 具体的な単語を修正候補として提案する機能を追加していきたいと考える.

これらのことから, 単語の関連語を抽出する際の条件の追加, ストーリー生成後にさらに具体的な単語を修正候補として提案する機能を追加, テンプレートおよびカテゴリの追加による, システムの改善をすることでより多様性のあるシステム作成を目標としたい.

参考文献

- [1] 志水 彰: 笑い-その異常と正常, 勁草書房, (2000)
- [2] 吉田 祐介, 萩原 将文: 漫才形式の対話文自動生成システム, 日本感性工学会論文誌, Vol. 11, No.2, pp.265-272(2012)
- [3] 真下 遼, 梅谷 智弘, 北村 達也, 灘本 明代: 文の感情を考慮した漫才ロボット台本自動生成 手法の提案, DEIM, Forum, F4-4(2015)
- [4] 山根 宏彰, 萩原 将文: 笑いを生むことわざかしの自動生成システム, 日本知能情報ファジィ学会誌, Vol.24, No.2, pp.671-679(2012)
- [5] 深津 春文, 伊藤 毅志: テーマとキーワードからアイデアの発想を支援するシステム, 情報 処理学会 研究報告 (Web), Vol.2016, No.11, HCI, pp.167-175(2016)

自動車運転者の行動分析のための注視情報抽出法

A New Method of Fixation Identification for Driving Behavior Analysis

針ヶ谷勇星^{1*} 白山晋¹
Yusei Harigaya¹ Susumu Shirayama¹

¹ 東京大学大学院工学系研究科

¹ School of Engineering, The University of Tokyo

Abstract: Determining the starting time, duration, and target of drivers' fixation is useful for driving behavior analysis. However, this information is currently extracted from raw gaze data by hand or by algorithms that are not suitable for mobile eye tracking. In this paper, we firstly propose a novel method for extracting drivers' fixation information that is robust to drivers' head movement utilizing spatial and temporal features of drivers' foveal vision. Then we introduce the gaze data processing system that semi-automatically extracts drivers' fixation information based on our proposed method. It is shown that the proposed method outperforms previous methods in terms of both validity and efficiency. The result indicates that our system is effective in driving behavior analysis.

1 はじめに

日本では高齢運転者による交通事故が増加しており、社会問題となっている [1]。高齢運転者の事故率が高い原因として心身機能の低下が挙げられるが、その程度には大きな個人差があるため、個々の高齢者自身にどのような運転上の問題があるかを把握し、一人ひとりの特性に応じた指導や支援を行うことが必要となる。Palazzi ら [2] は、このような個人に即した支援にあたっては、運転者が運転中に受容する情報を分析し運転者個人の認知状態を推定することで、より効果的な支援が可能であるとしている。

このような背景から、運転者が受容した情報をもとに彼らの行動分析を行うことで、運転者の行動分析に基づく運転者への助言・指導や運転支援システムの構築に活用できるような知見を得ることが求められている。運転行動は一般的にその 90 % が受容した視覚的情報に基づいてなされているため、運転者に対して視線計測を行うことで、彼らが受容した視覚的情報を分析することができると考えられている。藤本ら [3] は、熟練運転者が注目した領域をドライブレコーダーの映像に重ねて表示することで熟練運転者が受容したと思われる情報を可視化し、これを初心運転者に見せることで、初心運転者による映像中の危険因子発見率が有意に向上したと述べており、運転者の注意機構が視線移動に反映されていることが示唆されている。Martin

ら [4] は、運転者がどのような物体および領域を注視したかを分析することで、運転者の状況理解度の把握や次に行う運転行動の予測を行うことが可能であると述べている。このため、運転中の視線計測データを分析する際は運転者の注視情報に着目して分析を行うことが有用であり、様々な形で研究が行われている。

Martin ら [4] は、注視情報として以下の 3 つを挙げている。

- 注視開始時刻
- 注視終了時刻
- 注視対象の属性

すなわち、注視情報の抽出を行うためには、まず各時刻における視線データを注視状態であるか否かの 2 つに分類し、その後注視対象を識別することが必要である。注視状態の分類手法は広く研究が行われており、代表的な手法に視線移動速度に閾値を設けることによって分類を行う I-VT [5] がある。しかしながら、従来手法は頭部を固定した状態でスクリーンと対峙するような特殊な実験室環境下で計測されたデータに対して適用することを前提としており、実環境、とりわけ運転中のような大きな頭部運動を伴う環境において注視状態を安定して検出する手法は未だに確立されていない。

したがって、運転者の注視状態に着目した既存研究は、その多くが研究者が手作業で各視線データに対して注視状態の検出および注視対象の識別を行うものである。手作業での注視情報抽出を行う場合、運転者の注視状態を分析することは原理上可能ではあるが、視線研究の専門家でなければ作業が非常に困難である上

*連絡先：針ヶ谷 勇星，東京大学大学院工学系研究科システム創成学専攻 白山研究室，文京区本郷 7-3-1
E-mail: yharigaya429@gmail.com

に、膨大な視線データ一つひとつを手作業で処理することは非常に高コストであり大規模な分析は非現実的である。したがって、運転者の注視状態の分析を低コストかつ非専門家によって行うことを可能にするためには、実環境下における高精度な注視状態の検出、および注視対象の識別を自動で行う必要がある。

以上を踏まえ本研究では、自動車運転者の注視情報抽出において課題となる

- (i) 自動車運転時のように頭部運動の大きい環境下での注視状態の機械的な検出が困難である
- (ii) このような環境下における視線データから、注視している対象物を機械的に識別することが困難である

の2点を解決する注視情報抽出法を提案する。また、提案した注視情報抽出法に基づく注視イベント抽出システムを構築する。これらによって、運転者の行動分析の精度向上と効率化に寄与する。

2 関連研究

2.1 注視状態検出手法の研究

先述した、頭部運動に対して頑健な注視状態検出手法が確立されていない課題の解決を目的とするいくつかの研究を紹介する。

Hossain と Miléus [6] は、頭部装着型視線計測装置に搭載されたジャイロセンサ、加速度センサからの出力の積分値から頭部運動を推定し視線位置の補正を行うことで I-VT [5] の適用を試みている。彼らの検証実験では I-VT の適切な角速度閾値を絞り切れておらず、この原因としてセンサの出力値のノイズによって頭部運動の推定が正しく行えていない可能性が言及されている。

Steil ら [7] は、視野カメラ映像中の隣接するフレームの視線位置周辺の画像を切り出し、これを CNN に入力することで2つの画像の類似度を定量化し、この類似度の値に閾値を設けることで注視状態検出を行う手法を提案している。視線位置の座標値でなく画像特徴を用いることで理論的に頭部運動に頑健な手法となっているが、類似度計算を行った後閾値処理を行うという完全に独立した2段階からなるため、フロー全体での最適化がなされていない。また、彼らが用いたデータセットは非公開となっている。

2.2 運転者の視線分析に関する研究

Sun ら [8] は、運転中の高齢運転者に対して視線計測を行い、視線位置の空間的広がりから視線移動パターンを分類することで、事故に繋がりやすい視線パター

ンを特定する手法を提案している。彼らは、分析の結果認知負荷が高いとされる特定の状況において特有の視線行動がみられると述べているが、分析には視線位置の座標値のみを用いており、運転者の注視した内容を考慮した分析を行っていないため、議論の妥当性を示すことが困難である。

Tawari ら [9] は車内に設置したカメラで運転者の顔を撮影し、映像に含まれる顔特徴点の情報を入力として、予め分割した車内の領域のいずれに視線が向いているかの多クラス分類を行い、運転者の注視対象を試みている。Tawari らの手法は、運転者の前方に設置したカメラのみを用いる非侵襲的なシステムである点で優れているが、広い領域内のどこが注視されているかが認識できず、また先行車、歩行者など車外環境に存在する対象への注視を検出することは想定されていないため、運転中の注視行動を分析するに足る情報を得ることができているとはいえない。

3 提案手法

関連研究の課題を踏まえ、視線計測データから自動で注視情報を抽出する手法を提案する。運転者の視線計測データとして視野カメラ映像、車載カメラ映像、視線座標データの3つが入力されると、出力として注視開始時刻、注視終了時刻、注視対象からなる注視情報が抽出される。図1に提案手法のフローを示す。

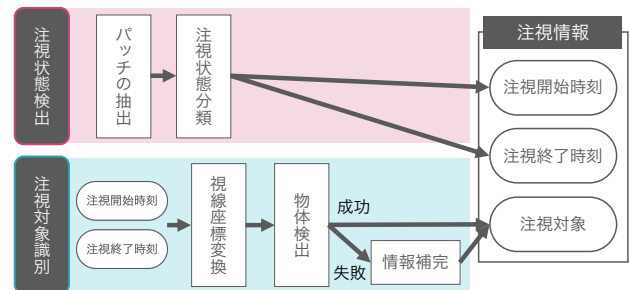


図1: 提案手法のフロー

提案手法は、以下の2つの段階に大別される。以下では、それぞれの詳細を述べる。

3.1 注視状態検出

はじめに、視線計測データから運転者が注視を開始した時刻および終了した時刻の情報を抽出する。

3.1.1 視線座標欠損データの削除

視線計測データには計測装置のデータ計測漏れおよび被験者の瞬目による欠損値が含まれるため、本研究ではこのような異常データを削除して分析を行う。具

体的には、ある視線座標データ $\mathbf{Q} = \{q_1, q_2, \dots, q_N\}$ について、欠損データが $q_k, q_{k+1}, \dots, q_{m-1}, q_m$ であった場合、これら $m - k + 1$ 個のデータを削除し、その前後に存在する $\mathbf{Q}_{\text{before}} = \{q_1, q_2, \dots, q_{k-1}\}$ および $\mathbf{Q}_{\text{after}} = \{q_{m+1}, q_{m+2}, \dots, q_N\}$ をそれぞれ独立した時系列として保持する。

3.1.2 視野カメラ映像からのパッチ抽出

次に、視線計測データの座標情報をもとに、視野カメラ映像の対応するフレームから視線位置周辺のパッチを抽出する。パッチ抽出の概念図を図2に示す。注視状態にあるときには、視線位置周辺の画像特徴は変化が少ないと考えられる。これを利用し、ある時間内の視線位置周辺の画像特徴の変化を入力として学習器により注視状態検出を行うことが本手法の着想である。

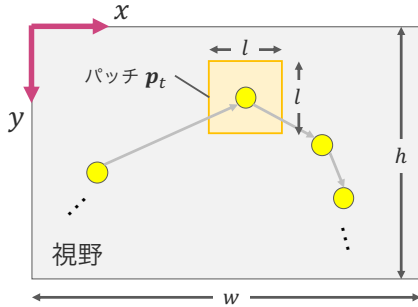


図2: 視野カメラ映像からのパッチ抽出の概念図

抽出するパッチは、視線座標 (x, y) を中心とする一辺 l の正方形領域とする。ここで、 l の値はヒトの中心視野領域に相当するサイズに周辺視野の影響を考慮したマージンを設けたものとする。本研究では中心視野領域の視野角 2° に対して 4° の範囲の画像を抽出した。

パッチ抽出は時系列内のデータのそれぞれについて行い、各時系列内の視線座標データ $\mathbf{Q} = \{q_1, q_2, \dots, q_n\}$ に対してパッチデータ $\mathbf{P} = \{p_1, p_2, \dots, p_n\}$ が作成する。各パッチを注視状態検出部への入力サイズにリサイズした後、 n 個のパッチの各ピクセルの輝度値の平均と分散を求めることにより、各パッチの輝度値の正規化を行う。

3.1.3 各時刻における注視状態分類

抽出したパッチデータを入力として注視状態分類、すなわち注視状態であるか否かの分類を行う。図3に、本研究で提案する注視状態検出部の構造を示す。本研究では Donahue ら [10] の LRCN の構造を参考にした。具体的には、画像特徴を抽出する CNN に、時系列処理を行う LSTM を接続し、最後に全結合層によって注視状態・非注視状態の二値分類を行う学習器を作成する。ここで、LSTM 層としては Bidirectional LSTM [11]（以下 BLSTM）を2層に重ねたものを用いる。これによ

り、予測サンプルの未来の情報を入力に加えた予測が可能となる。

はじめに、パッチデータからウィンドウ幅 s のデータ $\mathbf{P}_{\text{input}} = \{p_t, p_{t+1}, \dots, p_{t+s-1}\}$ を取得する。次に、ウィンドウ内の各パッチを CNN を用いて畳み込み、これにより得られた特徴量 $\mathbf{F} = \{f_t, f_{t+1}, \dots, f_{t+s-1}\}$ を BLSTM に入力することで、ウィンドウ内のある時刻における注視状態の二値分類（注視: 1, 非注視: 0）を得る。

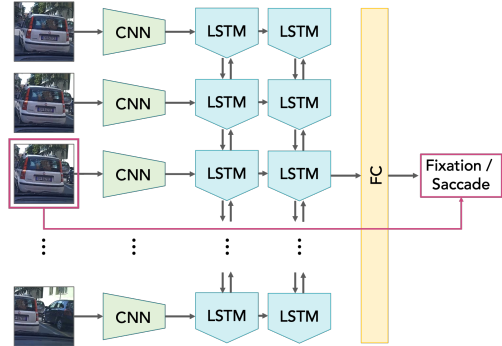


図3: 注視状態検出部の構造

3.1.1 項から 3.1.3 項までの手順によって、全視線計測データは注視状態・非注視状態のいずれかに分類される。このうち、連続して注視状態であると分類されたデータ群それぞれを注視イベントと称する。各注視イベントに属する最初および最後のデータに対応する時刻がそれぞれ注視開始時刻、注視終了時刻となる。

3.2 注視対象識別

次に、上述の手順で検出された各注視イベントにおける運転者の注視対象を識別する。視線位置周辺において物体検出を行うことで注視対象の候補を求め、この中から注視対象を決定する。

3.2.1 車載カメラにおける視線座標の推定

はじめに、注視イベントを構成する各データについて、車載カメラにおける視線座標を推定する。運転者が視野の周縁部を注視している可能性を考慮すると、注視対象の検出は視野カメラよりも画角が広い車載カメラ映像上で行うことが望ましいと考えられる。したがって、視野カメラ座標系における視線座標 (x_s, y_s) から車載カメラ座標系における視線座標 (x_c, y_c) への変換を行う必要がある。

まず、視野カメラ映像、車載カメラ映像それぞれの対応するフレームから AKAZE [12] により画像特徴点を検出する。次に、k 近傍法により双方の画像から検出された特徴点同士のマッチングを行う。一般的に、このマッチングには誤対応が含まれるため、RANSAC [13] を

用いて誤対応の除去を行う。最後に、誤対応が除去された対応点を用いて2画像間の透視変換行列を推定し、これを用いて視線座標を車載カメラ座標系に射影する。

3.2.2 各注視イベントにおける注視対象の識別

車載カメラ座標系に射影された視線座標を用いて、その時刻の車載カメラ映像中に存在する物体の検出を行う。提案手法では視野内の画像特徴から注視状態検出を行うため、各注視イベント中の視野はそれぞれ類似した画像特徴を持っていると考えられる。したがって、運転者は各注視イベントにおいて同一の対象を注視していると推察される。そこで本研究では、各注視イベントに対して注視対象を一つ対応付ける。

まず、車載カメラ映像の各フレームを入力として物体検出を行う。出力された物体リストの中から物体領域の範囲内に視線位置座標を含む物体を注視対象候補として保持する。この操作を当該注視イベント中に含まれる全フレームに対して行い、注視対象候補として最も多く挙げられた物体クラスを注視対象とする。ただし、大半のフレームで物体が検出されなかった場合などには、対象の識別結果が妥当でない可能性がある。このような場合に対処するため、データ数 n に対して閾値 n_0 を超える数のデータに同一の物体が対応付けられた場合のみその物体を注視対象とすることで、注視対象として妥当なもののみを記録する。本研究では $n_0 = 0.5n$ とした。なお、物体クラスのラベルとしては、ドライブレコーダー映像を用いた物体検出データセットである BDD100k [14] のラベルに、背景として運転者が注視し得る建物・道路・空・植物の4種および車内空間を示すラベルを加えて用いる。

本研究では、物体検出手法として M2Det [15] の著者が公開している学習済みモデルを用いる。

3.2.3 注視対象情報の補完

物体検出に失敗した場合や、注視対象を表すラベルがデータセットに含まれない場合等は上述の手順では注視対象を識別することができない。このような場合には、例外処理として運転者の注視対象を人間が判断し情報を補完することとする。

3.3 注視イベント検出システム

本研究では、提案手法に基づく運転者の注視イベント抽出システムを、上述の補完作業を行う作業者が操作するアプリケーションとして実装した。実装は、Unityを用いてスタンドアローンな Windows アプリケーションとして行った。図4に本研究で実際に実装したアプリケーションの作業画面を示す。

3.1.1 項から 3.2.2 項に示した処理が機械的に行われる。ある注視イベントに対して自動で注視対象の識別ができなかった場合、作業者にはアプリケーションの



図 4: 注視イベント抽出システムの作業画面

画面上で当該イベント時間内の視野カメラ映像、車載カメラ映像および視線位置周辺の映像を拡大した映像の3つが提示される。各映像中には、視線位置の座標が黄丸で示されている。

補完作業は以下の手順により行われる。

1. 注視イベント中の映像の確認。作業者は、まず図4に示した画面左下の領域にある再生ボタンを押下し、1つの注視イベントの時間内の映像を確認する。この映像は、再度再生ボタンを押下することで繰り返しの確認が可能である。
2. 注視対象の決定。映像の確認が終了した後、作業者はその映像中で運転者が注視していると思われる対象を示すラベルを1つ選び、対応するボタンを押下する。

作業者は、以上の手順を 3.2.2 項の処理において物体検出が成功しなかった全ての注視イベントについて繰り返す。この作業により、3.1.3 項の処理において検出された全ての注視イベントに対して注視開始時刻、注視終了時刻、注視対象の3つの情報が視線計測データから抽出される。

3.4 視線計測データの事前処理手法

本研究では、注視状態検出部の精度向上を図るため、入力となる視線計測データに適用する事前処理手法として以下の2つを合わせて提案する。

3.4.1 視線座標データの平滑化

一般に、視線計測装置の計測誤差は 0.5° 程度とされ、視線位置の座標にはデータ点ごとに細かなずれが発生する。このずれによりパッチの画像特徴が変化することで、注視の最中の視線位置周辺の画像特徴が類似することを利用した提案手法の精度を低下させている可能性が考えられる。

この計測誤差の影響を低減するため、視線座標データに対してガウシアンフィルタによる平滑化を行う。本

研究では、 x 座標、 y 座標それぞれのデータに前後 1 近傍の 1 次元ガウシアンフィルタを畳み込む平滑化を行った。

3.4.2 訓練データの選択的利用

人間の視線行動においては、平均 200 ms～300 ms の注視状態のあと 30 ms～80 ms の非注視状態が発生する、ということが繰り返される。注視対象の複雑度によっては注視状態がさらに長く続くことも考慮すると、視線計測データは一般に注視状態のサンプル数が非注視状態のそれより顕著に多い不均衡データであるといえる。

このデータの不均衡を低減するため、多数派クラスである注視状態のデータの削減を行う。具体的には、視線計測データからあるウィンドウサイズで切り出される、注視状態検出部への入力の時系列を以下の 4 種類に分類する。

- (a) ウィンドウに含まれるデータ全てが注視状態
- (b) ウィンドウに含まれるデータ全てが非注視状態
- (c) 両方の状態が含まれ、予測サンプルが注視状態
- (d) 両方の状態が含まれ、予測サンプルが非注視状態

このうち、後述する 2 つのデータセットのいずれも (a) にあたるウィンドウを全体の 6 割以上含んでいた。(a)(b) のように全サンプルが同一ラベルのものは分類容易であると推察されるので、学習困難な例を残すために (a) にあたるウィンドウのうち 75 % をランダムに削除したものを訓練データとして用いる。ただし、この操作により訓練データとテストデータでクラス確率に際が生じ、テストデータに対する予測において非注視状態の事前確率が実際より大きくなると考えられる。したがって、以下の式を用いて実際の学習器による予測確率 p_s を本来の確率である p に補正する。

$$p = \frac{p_s}{p_s + \frac{(1-p_s)}{\beta}} \quad (1)$$

ここで、 β は注視状態のデータのうち、訓練データとして選択される確率である。

4 注視状態検出の評価実験

3.1.1 項から 3.1.3 項に示した注視状態検出手法の評価実験を、以下の記述に従って行った。

4.1 使用したデータセット

4.1.1 GazeCom

GazeCom [16] は予め撮影した日常の映像を被験者に見せた際の視線計測データである。被験者はディスプレ

イに正対しそこに映された映像を見るため、常にディスプレイ方向を向いており頭部運動は発生しない。データセットは 18 種の映像を平均 47 人が見る計 477 試行の視線計測データから構成されており、各試行の長さはそれぞれ約 20 秒である。各サンプルは (Fixation, Saccade, Pursuit, Noise) のいずれかにラベリングされているが、本研究で扱う注視状態の定義より、Fixation と Pursuit を合わせたものを注視状態のラベル、Saccade を非注視状態のラベルとして処理した。また、視線計測データは 250 Hz で記録されているが、一般的な頭部装着型視線計測装置の時間分解能に合わせ 50 Hz にダウンサンプリングして用いた。

4.1.2 DR(eye)VE

DR(eye)VE [17] は運転中の運転者に対して行った視線計測結果のデータセットである。被験者は頭部装着型視線計測装置を着用しており、運転中の左右・後方確認等によって大きな頭部運動が発生している状況を含む。計 74 試行の視線計測データから構成されており、各試行の長さはそれぞれ約 300 秒である。各サンプルは (Fixation, Saccade, Noise) でラベリングされているため、Fixation を注視状態のラベル、Saccade を非注視状態のラベルとして処理した。データは 60 Hz で記録されている。

4.2 実験設定

本実験では、注視状態検出部の CNN として畳み込み層 4 層からなる簡素な構造を使用した。各畳み込み層はバッチ正規化、プーリングの操作を含む。活性化関数には ReLU を用い、プーリングのウィンドウサイズは 2×2 とした。入力となるパッチのサイズは 64×64 ピクセルとし、ウィンドウサイズは 100 ms に相当する長さとした。バッチサイズ 128 のミニバッチ学習を行い、epoch 数は学習状況を考慮し 20 とした。最適化には Adam を用い、学習率は 1.0×10^{-3} より始め、学習が 50 % 経過した段階で 1.0×10^{-4} とした。

提案する注視状態検出部 (CNN-BLSTM) は、ウィンドウ内の最後の時刻のサンプルを予測するもの (CNN-BLSTM_ls)、中央の時刻のサンプルを予測するもの (CNN-BLSTM_md)、視線座標データの平滑化を行うもの (CNN-BLSTM_sm)、訓練データの削減を行うもの (CNN-BLSTM_ds) の 4 つを比較した。前処理を行う 2 つに関しては、ウィンドウの中央の時刻を予測する構造を使用した。比較手法として、注視状態検出で一般的に用いられる手法である I-VT と、頭部運動に頑健な手法として提案されている Steil ら [7] の手法を用いた。

各手法の性能評価はそれぞれのデータセットを 6 分割することで $k = 6$ の交差検証を行い、テストデータに対する分類結果について評価指標を算出することに

よって行った．評価指標としては注視状態および非注視状態それぞれに対する F 値, Cohen の κ 係数, 各手法により検出された注視状態の継続時間分布を用いた．

4.3 実験結果

4.3.1 数値実験 1 : GazeCom

はじめに, 頭部運動の存在しないデータセットである GazeCom について注視状態検出を行った結果を考察する．表 1 に, GazeCom データセットで評価を行った結果を示す．紙面の都合上, 評価指標として F 値のみを示す．

表 1: GazeCom データセットにおける評価指標

	F 値 (注視)	F 値 (非注視)
CNN-BLSTM_ls	0.975 $\pm$ 0.002	0.785 $\pm$ 0.011
CNN-BLSTM_md	0.977 $\pm$ 0.002	0.811 $\pm$ 0.010
CNN-BLSTM_sm	0.966 $\pm$ 0.006	0.772 $\pm$ 0.023
CNN-BLSTM_ds	0.977 $\pm$ 0.003	0.808 $\pm$ 0.009
Steil et al. [7]	0.966 $\pm$ 0.004	0.709 $\pm$ 0.038
I-VT	0.977 $\pm$ 0.001	0.808 $\pm$ 0.006

表 1 より, 提案手法と I-VT はほぼ同等の性能を示し, Steil らの手法はこれらに劣る結果となった．GazeCom は被験者の頭部運動が発生しないデータセットであるため, 頭部運動に対応できない I-VT でも高い精度で注視状態検出が可能であったと考えられる．提案手法は被験者の頭部運動に頑健な注視状態検出を目的とする手法であるが, 頭部運動の発生しない状況においても, そのような状況に特化した I-VT と同等の精度で検出が可能であることが示唆された．これに対して, Steil らの手法も提案手法と同じく頭部運動に頑健な検出を目的とするものであるが, 頭部運動の発生しない状況において I-VT よりも低い精度であることから, 様々な状況で用いることのできる汎用的な手法ではないと考えられる．

また, 提案手法ではウィンドウの中央の時刻の予測を行った場合に, 最後の時刻を予測した場合よりも高い精度を示し, 特に非注視状態の F 値に大きな改善がみられた．中央の時刻の予測を行った場合, 最後の予測の場合と比較して入力に含まれる過去の情報量が半分程度となるが, この影響を踏まえても注視状態検出では予測時点より未来の情報を入力に含めることがより重要であることが示唆された．これは, 特に非注視状態において隣接するサンプル間で画像特徴が大きく異なることから, この変化の前の情報だけでなく変化後の情報を考慮することが精度改善に大きく寄与するからであると考えられる．

次に, 提案した訓練データの前処理手法の効果を考察する．まず, 提案手法において視線座標データの平

滑化を行ったものにおいては, わずかな精度の低下がみられた．この原因としては, 座標の平滑化によって注視状態と非注視状態の境界部分で切り出されるパッチの画像特徴の変化が低減されてしまい, この部分の分類性能が劣化した可能性が考えられる．一方で, 訓練データを間引いて学習を行った場合は, 訓練データの選択を行わない場合とほぼ同等の性能となった．訓練データ量が約半分になっていることを鑑みると, これは提案した訓練データ選択手法によって分類困難なデータを残して学習ができたことを示唆している．

4.3.2 数値実験 2 : DR(eye)VE

次に, 頭部運動の大きい状況を含むデータセットである DR(eye)VE について注視状態検出を行った結果を考察する．表 2 に, DR(eye)VE データセットで評価を行った結果を示す．こちらも同様に評価指標として F 値のみを示す．

表 2: DR(eye)VE データセットにおける評価指標

	F 値 (注視)	F 値 (非注視)
CNN-BLSTM_ls	0.916 $\pm$ 0.005	0.419 $\pm$ 0.035
CNN-BLSTM_md	0.926 $\pm$ 0.007	0.566 $\pm$ 0.028
CNN-BLSTM_sm	0.922 $\pm$ 0.007	0.521 $\pm$ 0.035
CNN-BLSTM_ds	0.917 $\pm$ 0.011	0.602 $\pm$ 0.042
Steil et al. [7]	0.871 $\pm$ 0.016	0.439 $\pm$ 0.037
I-VT	0.899 $\pm$ 0.017	0.551 $\pm$ 0.042

表 1, 表 2 を比較すると, DR(eye)VE では全体として各手法の分類能力が低い．これは, DR(eye)VE が GazeCom よりも分類困難なデータセットであったことが原因であると考えられる．

GazeCom の分類において提案手法と同等以上の性能を示した I-VT は, DR(eye)VE の分類においては提案手法より顕著な精度の低下が見られた．これは, GazeCom には含まれなかった被験者の頭部運動が大きい状況のデータが DR(eye)VE に少なからず含まれることが原因であると考えられ, I-VT に代表される従来手法では頭部運動の大きい状況において適切に注視状態検出を行うことが困難であることが示唆された．また, このような状況においては提案手法が従来手法よりも高精度の検出を行うことができたために, 4.3.1 項の結果と合わせて考えると, 提案手法によって, 頭部運動の有無にかかわらず頑健な注視状態検出が可能であると示唆された．

提案手法において, 視線座標データの平滑化を行った場合, 訓練データを間引いて学習を行った場合ともに通常より精度が僅かに低下する結果となった．しかしながら, 訓練データを間引いた場合には少数派クラスである非注視状態の F 値が改善された．したがって, 提案した訓練データの選択手法が, 運転者の視線計測

データの注視状態および非注視状態の不均衡の是正に有効であることが示唆された。他の場合と比較して各指標の標準偏差は増大しているが、これは訓練データの選択により約半分の訓練データ量で学習を行ったことが主因であると考えられる。そのため、データ削減前と同程度の規模の訓練データを用いて学習を行うことでより安定した学習が可能になると推察される。

5 注視イベント抽出の評価実験

5.1 実験概要

本実験では、提案した注視イベント抽出システムの有効性を示すため、従来行われてきた手作業による注視イベント抽出（従来手法）と作業の効率性や妥当性を比較する。抽出作業を行うにあたって、被験者がそれぞれの手法で作業を行う際の UI の差異による影響を可能な限り低減するため、手作業による注視イベント抽出についても提案手法と同様に作業用アプリケーションを実装した。図 5 に手作業用アプリケーションの作業画面を示す。



図 5: 手作業による注視イベント抽出の作業画面

作業者は、作業対象のデータ範囲の映像をコマ送りで確認し、各フレームで示されている視線位置の変化を確認することで対象への注視が開始したフレーム、終了したフレーム、その間の運転者の注視対象を判断し、これらの情報をアプリケーションに入力する。

5.2 実験設定

実験の被験者は、正常な視力を有する男子大学生・大学院生計 6 名とした。6 名全員が普通自動車運転免許を所持し、年に数回以上の頻度で運転を行っていた。被験者は、提案システムによる注視イベント抽出作業および手作業により 3 つの注視情報全てを抽出する作業をそれぞれ一度ずつ行った。実験結果として、各試行における所要時間および抽出された注視情報が記録された。実験データとして、DR(eye)VE から注視状態・

非注視状態の切り替わりのタイミングを多く含む 300 フレーム（5 秒）の範囲のデータを抽出したものを 3 種類用意しこれを用いる。

5.3 実験結果

はじめに、提案手法に基づくシステムの効率性を評価する。実験の結果、各被験者が提案手法および従来手法による作業に要した時間を表 3 に示す。

表 3: 注視イベント抽出作業の所要時間平均 [s]

提案手法	従来手法
221.2 ± 95.2	683.5 ± 242.0

各被験者の提案手法による作業時間と従来の手作業による作業時間の対応をみるため対応あり t 検定を行ったところ、2 つの手法による作業時間の分布に有意差がみられた ($p = 0.006$)。したがって、提案手法による作業で手作業よりも有意に短い時間で作業を完了させることができることが示唆された。また実験後、各被験者が作業前後の疲労度の差を 10 段階で主観評価した結果、提案手法において有意に疲労度が小さくなった ($p = 0.02$)。したがって、提案手法に基づくシステムでは作業者の負担が軽減されることが示唆された。以上の考察より、提案手法による作業は所要時間及び作業者への負担の両面において従来の手作業よりも優れており、より効率的な手法であると考えられる。

次に、システムによる抽出結果の妥当性を評価する。2 つの手法による注視イベント抽出結果から、視線計測データに対して注視状態・非注視状態の分類結果が得られる。これを DR(eye)VE の教師ラベルと比較することで、抽出結果の評価指標（F 値）を算出することができる。表 4 に、注視情報抽出結果の評価指標を示す。

表 4: 注視イベント抽出結果の評価指標

	F 値（注視状態）	F 値（非注視状態）
提案手法	0.925 ± 0.002	0.689 ± 0.048
従来手法	0.798 ± 0.069	0.376 ± 0.189

表 4 に示す通り、提案システムによる分類が全ての指標で従来の手作業による分類を上回り、提案手法による分類結果が従来手法より妥当なものであることが示唆された。従来のような手作業での注視イベント抽出は通常専門家によって行われるため、実際は両手法による結果の妥当性の差は本実験よりも小さくなることが予想される。しかしながら、提案手法を用いることによって非専門家でも従来より妥当に注視イベント抽出を行うことが可能となるため、運転者の注視情報をより幅広い機会で活用できると考えられる。

6 おわりに

本研究では、従来手作業で行われてきた運転者の視線計測データからの注視情報の抽出を自動化することで視線計測データを運転者の行動分析や運転者の支援システムで活用可能にすることを目的として、自動車運転者の注視情報抽出法を提案した。また、提案手法に基づき運転者の視線計測データから半自動で注視情報抽出を行うシステムを提案・実装し、その有効性を検証した。

評価実験の結果、提案した注視状態検出手法は視線計測データにおける頭部運動の有無にかかわらず従来手法以上の精度で注視状態・非注視状態の分類が可能であることを示した。また、本研究で提案した注視情報抽出法による作業は、従来のように完全に人手で情報の抽出作業を行う場合と比較して非常に短時間で完了するものであり、かつ作業への負担も有意に小さいことを確認した。抽出された情報に関しても、従来手法による情報抽出結果と比較してより実際の視線行動に近い結果が得られ、提案手法によって従来の作業より妥当な情報抽出が可能であることが示唆された。

以上の結果から、提案した注視情報抽出法および注視イベント抽出システムを用いることによって、自動車運転者の注視情報を従来の手作業と比較して短時間、低負担、高精度で抽出することが可能であり、本研究は運転者の注視情報を広く活用可能な情報にすることに貢献すると考えられる。

今後の展望としては主に次の2点が挙げられる。まず、より高性能な画像特徴抽出器を用いて注視状態検出部の学習を行うことで、より高精度での分類が可能になると考えられる。次に、運転者の視線がある領域に停留している際に、運転者がそれを注視しているのか、視野を俯瞰しているのかを区別して識別することによって、運転者の認知過程をより妥当に推定することが可能であると考えられる。

参考文献

- [1] 平成 29 年交通安全白書 (概要) - 内閣府. <https://www8.cao.go.jp/koutu/taisaku/h29kouhaku/gaiyo/features/feature01.html>. (最終閲覧日: 2019/11/12). 2017.
- [2] A. Palazzi, F. Solera, S. Calderara, S. Alletto, and R. Cucchiara. Learning Where to Attend Like a Human Driver. *2017 IEEE Intelligent Vehicles Symposium (IV)*. 2017, pp. 920–925.
- [3] 藤本武司, 砂山渡, 山口智浩, 谷内田正彦. 視線行動の可視化による着眼スキル伝達支援. *人工知能学会論文誌* 19 3 (2004), pp. 174–183.
- [4] S. Martin, S. Vora, K. Yuen, and M. M. Trivedi. Dynamics of Driver's Gaze: Explorations in Behavior Modeling and Maneuver Prediction. *IEEE Transactions on Intelligent Vehicles* 3 2 (2018), pp. 141–150.
- [5] D. Salvucci and J. Goldberg. Identifying Fixations and Saccades in Eye-tracking Protocols. *Proceedings of the Eye Tracking Research and Applications Symposium*. Jan. 2000, pp. 71–78.
- [6] A. Hossain and E. Miléus. “Eye Movement Event Detection for Wearable Eye Trackers”. Master Thesis. Linköping University, 2016.
- [7] J. Steil, M. X. Huang, and A. Bulling. Fixation Detection for Head-mounted Eye Tracking Based on Visual Similarity of Gaze Targets. *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. 23. 2018, pp. 1–9.
- [8] Q. Sun, J. (Xia, T. Falkner, and H. Lee. Investigating the Spatial Pattern of Older Drivers' Eye Fixation Behaviour and Associations with Their Visual Capacity. *Journal of Eye Movement Research* 9 6 (2016), pp. 1–16.
- [9] A. Tawari, K. H. Chen, and M. M. Trivedi. Where is the Driver Looking: Analysis of Head, Eye and Iris for Robust Gaze Zone Estimation. *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. 2014, pp. 988–994.
- [10] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 2625–2634.
- [11] A. Graves and J. Schmidhuber. Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures. *Neural Networks* 18 5 (2005), pp. 602–610.
- [12] P. Fernández Alcantarilla, J. Nuevo, and A. Bartoli. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. *Proceedings of the British Machine Vision Conference*. 2013, pp. 1–12.
- [13] M. A. Fischler and R. C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM* 24 6 (1981), pp. 381–395.
- [14] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell. BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling. *arXiv preprint 1805.04687* (2018).
- [15] Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, and H. Ling. M2Det: A Single-Shot Object Detector Based on Multi-Level Feature Pyramid Network. *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (2019), pp. 9259–9266.
- [16] M. Dorr, T. Martinetz, K. Gegenfurtner, and E. Barth. Variability of Eye Movements When Viewing Dynamic Natural Scenes. *Journal of Vision* 10 10 (2010), pp. 1–17.
- [17] S. Alletto, A. Palazzi, F. Solera, S. Calderara, and R. Cucchiara. DR(eye)VE: A Dataset for Attention-Based Tasks with Applications to Autonomous and Assisted Driving. *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2016, pp. 54–60.

アダプティブ e-ラーニングのための図形問題の類似度算出法

A Similarity Calculation Method of Geometric Problems for Adaptive e-Learning

多田 舜一^{1*} 白山 晋¹
Shunichi Tada¹ Susumu Shirayama¹

¹ 東京大学大学院工学系研究科システム創成学専攻

¹ Department of Systems Innovation, Faculty of Engineering, The University of Tokyo

Abstract: In recent years, attention has been paid to adaptive e-learning, a learning method that uses educational big data to flexibly change the learning content according to the learner's proficiency. In this research, we proposed an overview of an adaptive e-learning system that focuses on the solution procedure that means what kind of knowledges and operations the learner used to solve the problem. As for the method of construction, we implemented the figure problem of arithmetic as an example. In the proposed system, more detailed support for learners can be expected compared to the conventional method.

1 序論

1.1 背景

現在、日本政府が提唱している「Society 5.0」の一環として、EdTech による教育改革が推進されている [1]. EdTech とは、「Education (教育)」と「Technology (技術)」の融合を意味する造語であり、「教育における AI、ビッグデータ等の様々な新しいテクノロジーを活用したあらゆる取組」と定義されている. EdTech による教育改革の推進の中で、すぐにでも着手すべき課題としてアダプティブラーニングが挙げられている. 文献 [2] によれば、アダプティブラーニングとは、「学習者の要求や知識、好みなどの要素に基づき、適切なコンテンツを適切な方法・タイミングで提供する教育手法」であり、学習者が電子端末を用いて行う学習 (e-ラーニング) におけるサービスの一種である. このため、アダプティブラーニングはアダプティブ e-ラーニングと呼ばれることも多い.

既存のアダプティブ e-ラーニングサービスでは、コンテンツの提供方法に関して様々な課題がある. 例えば、Truong は、2004 年から 2014 年までのアダプティブ e-ラーニングに関する研究のレビューを行った中で、学習スタイルがアダプティブ e-ラーニングにおいて重要な役割を担っていると提唱している [3]. Truong によれば、多くのアダプティブ e-ラーニングに関する研究で Felder と Silverman によるモデル [4] を利用して

いるが、このモデルでは、学習スタイルを 4 軸 (Active - Reflective, Sensing - Intuitive, Visual - Verbal, Sequential - Global) それぞれの傾向を算出することで決定している. しかし、このモデルは学習分野に依存しない個人の学習行動の傾向について分類しているに過ぎず、このモデルに基づいたアダプティブ e-ラーニングサービスでは、各学習コンテンツをどう解いたかや、どのような間違い方をしたかなど、学習行動中における学習分野特有のパラメータについて考慮できていない. 加えて、Barr らは、既存の LMS (Learning Management System) に基づいた学習環境に加え、LRS (Learning Record Store) を活用することで、学習者の分析や学習者へのサポートなどを、学習行動中にリアルタイムに行えると主張しているが [5]、アダプティブ e-ラーニングサービスに LRS を活用する研究はまだ発展途上にある. また、学習分野特有のパラメータを考慮したアダプティブ e-ラーニングに関する研究として、那須野らは (株) リクルートマーケティングパートナーズとの共同研究を行っている [6]. この共同研究の中で、大規模オンライン講座アプリ「受験サプリ」(現「スタディサプリ」[7]) の数学教材における各講義の依存関係について、まず学習指導要領を参考に系統 (単元) 間の依存関係をネットワーク化して表現し、各講義の所属系統を講義名や使用教材から判断することで各講義間の依存関係を決定している. さらに、学習者の動画視聴や問題解答などの学習ログを深層学習を用いて解析し、学習者が次に取り組む問題の正誤を高い精度で予測している [8]. しかし、この手法では、那須野らが指定した系統を最小単位としているため、講義内のどこでつ

*連絡先: 東京大学大学院工学系研究科システム創成学専攻
〒113-8656 東京都文京区本郷 7-3-1
E-mail: tada-shunichi513@g.ecc.u-tokyo.ac.jp

まずいたかを考慮できておらず、また、膨大な量の各講義を全て手入力で分類しており、導入への障害となる。

以上を踏まえ、既存のアダプティブ e-ラーニングに関する研究及びアダプティブ e-ラーニングサービスにおいては、以下のような課題が挙げられる。

- 学習行動中の学習者に対するリアルタイムな分析・サポートに関する研究は発展途上にある
- 既存のアダプティブ e-ラーニングサービスでは、学習分野特有のパラメータを考慮した例は少ない
- 学習分野特有のパラメータを考慮した研究例でも、各システムを最小単位としているため、系統内により詳細なつまづきに対応できない
- 膨大な量の教材に対し、内容に応じた分類を全て手入力で行う必要があり、導入への障害となる

1.2 目的

本研究では、1.1 項の課題を踏まえ、以下に挙げる特徴を持つアダプティブ e-ラーニングシステムの開発を目指す。

- 学習者の学習行動中にリアルタイムに分析・サポートを行う
- 学習者の学習スタイルだけでなく、学習分野特有のパラメータとして、解答手順を考慮する
- 各教材を解答手順に基づいて自動で分類する

なお、本研究で提案するアダプティブ e-ラーニングシステムが対象とする学習者と学習内容、及び目的を表 1 に示す。

表 1: 本システムの対象

対象の学習者	中学校受験を控えた小学生
対象の学習形態	電子端末を使用した個人学習
対象の学習内容	算数の図形問題
学習の目的	中学受験で必要とされる図形問題の解答パターンの習熟

2 AL システムの設計

本節では、AL (Adaptive e-Learning) システムに要求される機能及びその検証のためのモデルを説明する。

2.1 学習環境のモデル化

一般に AL とは、学習者がコンピュータ端末やタブレット端末を用いて一人で学習する仕組みであるため、教授者が存在しない、いわゆる個人学習と呼ばれる学習形態である。文献 [9] で提案された、組織のモデル化のためのメタネットワークモデルを、複数の人間が共存する教育形態に拡張し、個人学習におけるメタネットワークモデルを定義する。個人学習を構成する要素を表 2 に、また、その要素間の相互関係を表 3 に示す。

表 2: メタネットワークを構成する要素

種別	内容
人間	対象となる学習者 (個人)
スキル 知識	解答に必要な知識・作業
リソース	解答行動中の手順や解答結果などの学習者の学習履歴
タスク	用意されている問題

2.2 AL システムに要求される機能

本研究で開発する AL システムの導入によって、前節で述べたモデル内において従来の個人学習からどのような変化を期待するか考察し、そのために要求される機能を検討する。なお、従来の個人学習として、小学校で課された宿題を自宅で行うことを想定している。2.1 項で述べたモデルを構成する要素や要素間の相互関係について、AL システムの導入によって期待する変化を表 4 にまとめる。また、そのために要求される機能を表 5 に示す。これらの機能を満たすために必要な手法とその実装方法は、3 節で説明する。

3 AL システムの実装

本章では、3.2 節で述べた AL システムに要求される 3 つの機能を実現するために必要な手法を述べる。

3.1 機能①:問題情報・学習履歴の保持

問題情報を計算機上で扱える形式で表現した研究として、国立情報学研究所の新井らによる「東ロボくん」[10] や、岡本による「GEOMEX」[11] などがある。これらの研究に倣い、論理型プログラミング言語 Prolog [12] の事実と規則を用いてデータベースを作成し、問題情報及び学習履歴を表現する手法を提案する。データベースは以下の 5 種のデータで構成される。

表 3: メタネットワークを構成する要素間の関係

種別	スキル・知識	リソース	タスク
人間	知識・作業の習得状況	学習履歴の参照可否	学習者への出題内容
スキル・知識	知識・作業の依存関係	どの知識・作業を用いた学習履歴であるか	問題を解答するのに必要な知識・作業
リソース		学習履歴の時系列	どの問題に関する学習履歴であるか
タスク			問題を解く順序

表 4: AL システムによる個人学習の変化

変化	項目	従来の学習	アダプティブ e-ラーニング
変化①	リソース	ノートや問題集などの紙媒体	タブレットなどの電子端末
	タスク		
変化②	人間 - スキル・知識	学習者自身の理解度や進捗を自ら解答箇所を振り返ることで参照	過去の学習履歴から学習者の理解度や進捗を自動で可視化
	人間 - リソース		
変化③	人間 - タスク	取り組むべき問題と順序が指定されており、解答行動中にはサポートを得られない	学習者の理解度に応じて取り組むべき問題と順序が動的に変化し、解答行動中に必要に応じたサポートを行う
	タスク - タスク		

表 5: AL システムに要求される機能

変化	機能
変化①	計算機上で扱える表現形式で問題情報及び学習履歴を保持する【機能①】
変化②	解答に必要な作業の習得度、知識の理解度を数値化する【機能②】
変化③	作業の習得度、知識の理解度、問題を解答するのに必要な作業・知識から問題の優先度を変動させる【機能③】

- 各問題における図形構造のデータ
図形を構成する部品として、点・線分・面の3つがある。また、部品間には様々な図形的な関係性がある。これらの部品・関係を計算機上で扱えるよう、Prologの述語を定義し、事実として記述した。それぞれの部品の記述方法を表6に、部品間の関係を表7にまとめる。
この形式に従い、各問題の図形構造をPrologの事実として表現したデータを作成する。
- 各問題における設問条件のデータ
各問題において設定されている長さや角度などの物理量を表8の命名規則に従いPrologの事実として表現し、また問題で問われている対象についてもPrologの事実として表現し、データを作成する。
- 作業及び知識のデータ
図形問題を解くために必要な作業及び知識で構成されるデータである。補助線の引き方のパターンや面積を求める公式などが相当する。
- 知識推論による解答情報のデータ
1. 2. 3. で作成したデータに基づいてPrologで知識推論を行い、問われている対象を求めるまでに使用した作業及び知識の組み合わせと、最終的に得られた解答を出力し、データを作成する。
- 学習履歴のデータ
4. で説明した作業及び知識の組み合わせと得られた解答、その正否と合わせて出力し、データを作成する。
これら5種のデータによって図1に示すようなネットワーク型のデータベースを構築する。このネットワーク構造を利用し、機能②、③を実現する。

表 6: 図形構造を構成する部品

部品	記述方法	内容
点	point(P, X, Y).	P : 点の名称 X : 点の x 座標 Y : 点の y 座標
線分	segment(L, P1, P2).	L : 線分の名称 P1, P2 : 線分の端点の名称 (辞書順に P1 ; P2)
多角形	polygon(Poly, Ps).	Poly : 多角形の名称 Ps : 多角形の頂点リスト
正多角形	regular_polygon (Poly, Ps).	Poly : 正多角形の名称 Ps : 正多角形の頂点リスト

表 7: 部品間の関係

部品 - 部品	記述方法	意味
点 - 点	reflect(P1, P2, P0).	点 P1 は点 P2 の点 P0 に対する対称点である
点 - 線分	on_line(P, L).	点 P は線分 L 上にある
点 - (正) 多角形	vertex(P, Poly).	点 P は多角形 Poly の頂点である
線分 - 線分	straight(L1, L2).	線分 L1 と線分 L2 は同一直線上にある
	parallel(L1, L2).	線分 L1 と線分 L2 は平行である
線分 - (正) 多角形	edge(L, Poly).	線分 L は多角形 Poly の辺である

表 8: 与えられている物理量

物理量	記述方法	意味
長さ	distance(P1, P2, Length).	点 P1 と点 P2 の間の距離は Length[cm] である
角度	angle(P1, O, P2, Angle).	線分 OP1 から線分 OP2 までの角は反時計回りに Angle[度] である
面積	area(Poly, Area).	多角形 Poly の面積は Area[cm ²] である
分数	value(Value, N, D).	Value の値は $\frac{N}{D}$ である

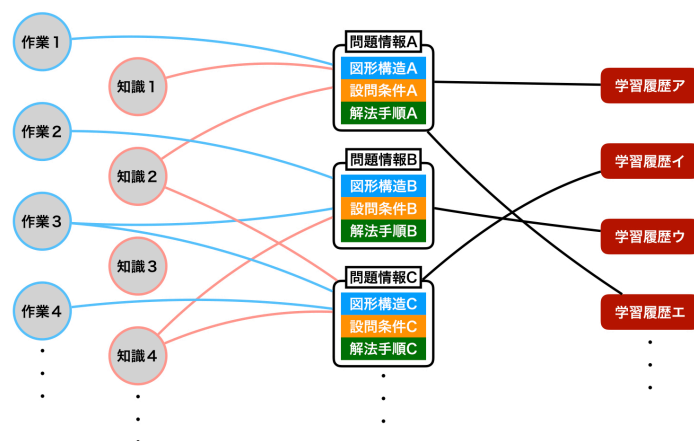


図 1: ネットワーク型データベースの概要図

4 類似度算出法

本節では、図形問題間の類似度を算出する方法を説明する。学習者に適切な出題・サポートを行うため、用意する問題群における各問題間の比較や、学習者が行った解答手順と正しい解法手順の比較は必要である。図形問題の解法手順は、前章で述べた個々の公理・公式・作業からなる有限個の重複順列として表現される。このような配列間の類似度を評価する代表的な手法として Levenshtein 距離と Needleman-Wunsch 法を説明する。Levenshtein 距離とはある文字列から他の文字列に変換する際に要する編集操作の回数である。ここでの編集操作は、文字の挿入・削除・置換であり、比較する文字列が似ている場合は文字列の変換に要する編集操作の回数は少なく、逆に比較する文字列が著しく異なる場合は変換に要する操作の回数は多くなるはずである。すなわち、Levenshtein 距離の大小が文字列間の類似性を表す。Levenshtein 距離と同様に文字列の類似度を算出する手法として、Needleman-Wunsch 法（NW 法）がある。NW 法はアミノ酸の配列や DNA 分子配列間の共通部分の抽出法として開発され、バイオインフォマティクス分野では一般的な手法となっている。

5 エキスパートシステムによる解法の出力

エキスパートシステムによる計算実験についてまとめる。本研究では、自作した及び問題集の平面図形問題を 31 問用意し、前述した Prolog に基づくエキスパートシステムで解答の出力を試みた。全問題のうち、正しい解答は少ししか得られなかった。原因として、様々な理由が考えられる。そのうちの一つは、Prolog の推論を行う順序にかかわるものである。Prolog の推論では、ソースコードの上から参照を順番に行うため、参照されず、正しい結果に結びつかないことがあった。図 2 にエキスパートシステムで解けた問題を、図 3 にエキスパートシステムで解けなかった問題を示す。

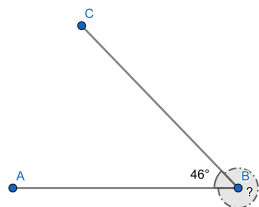


図 2: エキスパートシステムで解けた問題

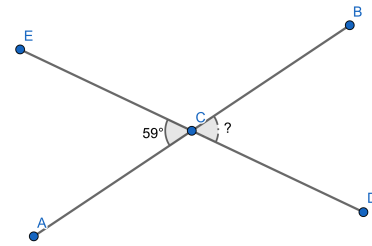
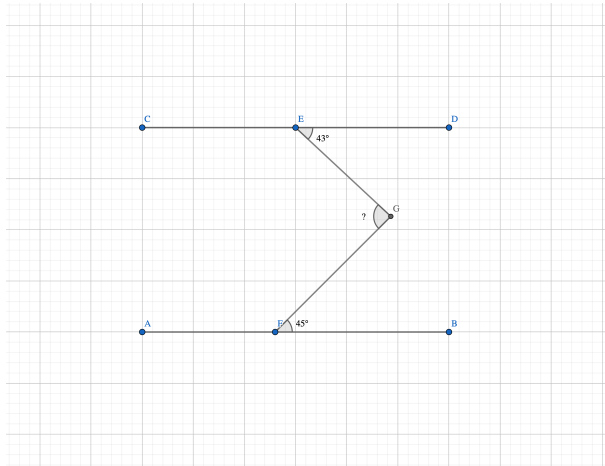


図 3: エキスパートシステムで解けなかった問題

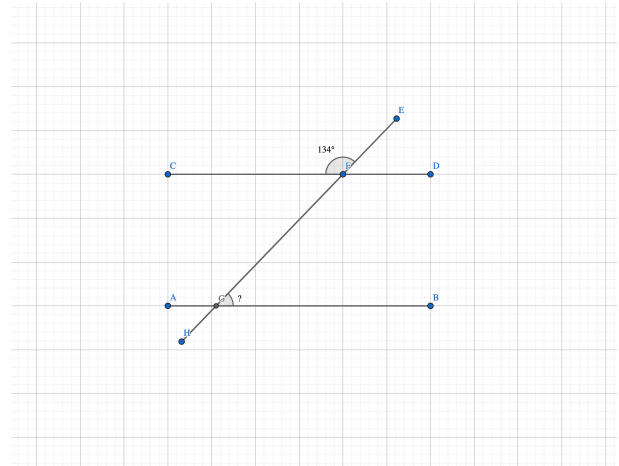
6 解法間の類似度算出

図 4 に示す問題群における解法に対し、Levenshtein 距離と NW 法による類似度算出を試みた。選択した問題について、Problem 1 と Problem 2 は錯角などの平行線に関わる作業を必要とし、Problem 3 と Problem 3 は正多角形や二等辺三角形を利用した作業を必要とするため、Problem 1 と Problem 2 の間、及び Problem 3 と Problem 4 の間で高い類似度を示すことを期待した。

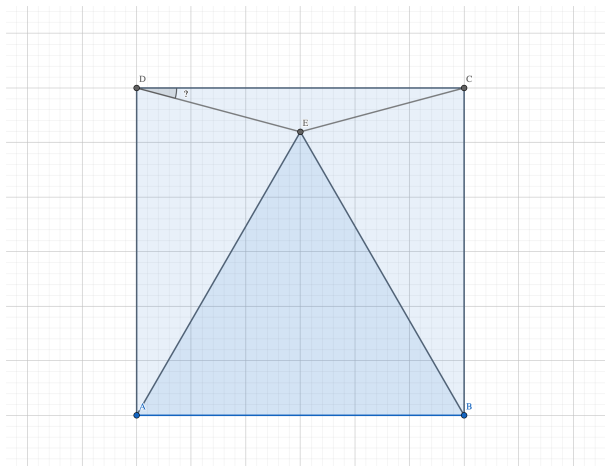
表 9 に Levenshtein 距離の計算結果を、表 10 に NW 法の計算結果を示す。Levenshtein 距離では、数値が小さいほど類似度が高く、NW 法では、数値が大きいほど類似度が高くなる。期待した通り、どちらの手法においても Problem 1 と Problem 2 の間と Problem 3 と Problem 4 の間で高い類似度が算出された。このことから、どちらの手法も問題間の類似度を解法に基づいて算出できることが示された。



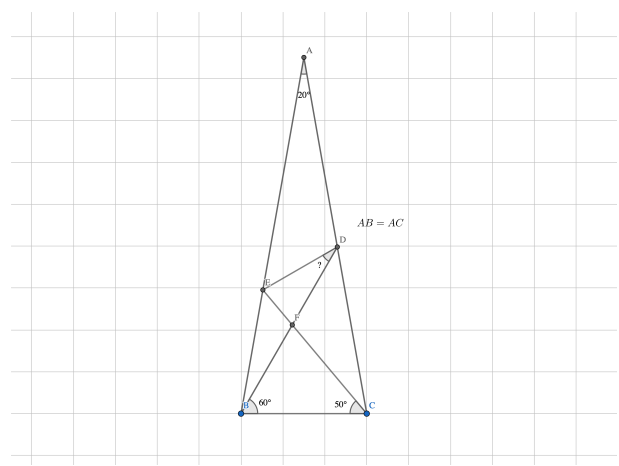
(a) Problem 1



(b) Problem 2



(c) Problem 3



(d) Problem 4

図 4: 類似度算出を行なった問題群

表 9: Levenshtein 距離の計算結果

	Problem 1	Problem 2	Problem 3	Problem 4
Problem 1		0.75	1.0	0.95
Problem 2			0.83	0.95
Problem 3				0.75
Problem 4				

表 10: NW 法の計算結果

	Problem 1	Problem 2	Problem 3	Problem 4
Problem 1		0.25	0	0.05
Problem 2			0.17	0.05
Problem 3				0.25
Problem 4				

7 結論

本研究のまとめと今後の課題について述べる。
本研究の成果は以下である。

- アダプティブ e-ラーニングシステムについて提案を行い、解答手順を考慮したシステムの全体像を示した。
- AL システムの導入に期待する環境の変化を考察することで、そのために要求される3つの機能を抽出した。
- データベースの概要を示し、そのうち図形構造の部分や知識・作業部分の実装を完了させた。
- 図形問題の類似度算出法について提案を行い、一部実験により確認した。

今後の課題として、

- エキスパートシステムの改良
- さらなる類似度算出法の実証実験の試行
- システム全体の完成へ向けた取り組み

が挙げられる。

参考文献

- [1] 文部科学省 新たな時代に対応するための EdTech を活用した教育改革推進プロジェクトチーム.: Society5.0 における EdTech を活用した教育ビジョンの策定に向けた方向性, http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/002/siryo/_/_icsFiles/afieldfile/2018/06/20/1406021_18.pdf (2019/11/17 最終アクセス)
- [2] Shi, L., Cristea, A. I., Foss, J. G., Al Qudah, D., Qaffas, A.: A social personalized adaptive e-learning environment: a case study in Topolor, *IADIS Int. J. WWWInternet*, Vol. 11, No. 2 (2013)
- [3] Truong, H. M.: Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities, *Computers in human behavior*, Vol. 55, pp. 1185–1193 (2016)
- [4] Felder, R. M., Silverman, L. K.: Learning and teaching styles in engineering education, *Engineering education*, Vol. 78, No. 7, pp. 674–681 (1988)
- [5] Barr, A., Robson, R.: Missing Pieces: Infrastructure Requirements for Adaptive Instructional Systems, *Proceedings of the Workshop on Adaptive Instructional Systems at HCI International*, pp. 169–178 (2019)
- [6] 那須野薫, 萩原静蔵, 井上綾香, 伊藤岳人, 浜田貴之, 川上登福, 松尾豊.: 大規模オンライン講座における自己適応学習者に着目した学習項目の理解度予測, 人工知能学会全国大会論文集 第 29 回全国大会, pp. 1N25–1N25 (2015)
- [7] スタディサプリ, http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/002/siryo/_/_icsFiles/afieldfile/2018/06/20/1406021_18.pdf (2019/11/17 最終アクセス)
- [8] 東京大学松尾研究室× BAI×リクルートマーケティングパートナーズ 共同研究. 学習者の「解けない問題」の予測に成功「つまづき予防」の実現へ: University of Tokyo Matsuo laboratory × BAI × recruit Marketing Partners joint research: Successful prediction of unsolvable problems of learners. Realization of preventing stumbling blocks., https://ring.education/research/recruit_tokyo-univ_matsuo-lab_bai_joint-research_20160915/ (2019 年 12 月 16 日 最終アクセス)
- [9] Carley, K. M., Pfeffer, J.: Dynamic network analysis (DNA) and ORA, *Advances in Design for Cross-Cultural Activities Part I*, pp. 265–274 (2012)
- [10] 新井紀子, 松崎拓也.: ロボットは東大に入れるか, 人工知能学会誌, Vol. 27, No. 5, pp. 463–469 (2012)
- [11] 岡本敏雄.: 知的 CAI 最近の動向: 初等幾何 CAI システム: GEOMEX, 情報処理, Vol. 29, No. 11, pp. 1301–1308 (1988)
- [12] Kowalski, R.: *Logic for problem solving*, Department of Computational Logic, Edinburgh University (1974)

時系列データを扱った研究の動向調査

Systematic Review of Studies on Time Serise Data

柴田 祐樹^{1*} 高見 玲¹ 菊池 圭祐¹ 武川 紘也¹ 白井 佑¹ 張 錦程¹
笹方 育也¹ 李 穎豪¹ 高間 康史¹

¹ 首都大学東京 システムデザイン研究科

Abstract: 時系列データを扱った研究は未来を推測する枠組みを与え、季節の予測等生活に不可欠であることから古くより行われている。現在では、対象とする手法や問題は分野横断的となっているため、全体像の把握が難しくなっていることから、時系列データを扱った研究の動向調査を行い、その結果を報告する。近年医学系の分野で用いられる体系的調査を参考に、本調査に適した形式に改良した調査方法についても報告する。

1 研究背景

本稿では人工知能分野における時系列データを扱った研究の動向調査の結果報告と、本調査に適した体系的な調査方法 (Systematic Review) [2] の提案を行う。

時系列の解析に対する要求は、季節の予測など生活に直結することから古くから存在する。これら要求は現代でも同様であり、加えて近年では計算機を用いた数値解析手法の発達、時系列に対する汎用的な数理モデルの導入により [6]、貨幣価値の変動、機械の制御、解析可能な範囲は広がり続けている。この現状に対し、時系列データを扱った研究を体系的にまとめた文献が見当たらないため、全体像を俯瞰可能とする資料作成が、本調査の目的の一つである。

別の目的として、一般性の高い体系的な調査手法を構築し、様々な調査に対して適用可能な手法を構築しておくことは、今後においても価値があることである。体系的調査は医療分野ではすでに広く行われており、この分野での手法は確立されていると言えるが、医療分野における文献調査では人命に関わることから漏れが認められない一方、人工知能分野においては明確な基準があるわけでもなく、このように調査趣旨が異なるため、人工知能分野に適した、効率的な体系的調査手法を構築することも本調査の目的とする。

調査対象とする文献発行媒体として、本調査では、2016 年から 2019 年までの間の日本人工知能学会全国大会予稿集を調査対象とした。本媒体を選んだ理由は、国内、特に人工知能分野での動向を幅広く知りたいため、また、文献あたりの容量がそれほど大きくなく、試行錯誤が行われる調査手法確立のための調査に適していると判断したためである。

*首都大学東京システムデザイン研究科
191-0065 東京都日野市旭ヶ丘 6-6
E-mail: hshibata@tmu.ac.jp

2 関連研究

2.1 体系的調査

本調査で採用する体系的調査はもともと医療の分野において、網羅的な調査に基づき全体として優勢なものを明らかにする目的で行われるものである。典型的には以下の、2 段階に分けて調査が行われる [2]。

- 1 次選別: 要約や表題を参考にした網羅的に文献を収集
- 2 次選別: 上記収集された文献の全文を確認し、関係する文献を選別

1 次選別では、取りこぼしが無いように、少しでも疑わしい文献ならば収集される。2 次選別では関係のない文献が除外される。どちらの作業も複数人で確認が行われる。手法や材料の出現頻度などを集計し、統計的に議論も行われる。

この体系的調査は、計算機科学など他分野でも利用されている [1,3,4]。基本的には [2] と同様に網羅的な、取りこぼしの無い調査方法で行われるが、選別の基準に文献の特徴を使うなど、調査対象や目的に応じて幾らかの拡張が行われる [4]。文献の収集方法は重要であるが、使う検索エンジンや、選別の効率性に言及している文献は見当たらない。

得られた知見をまとめるために、分野の分類を作成し、体系的に論じる手法がしばしば用いられる [5]。ACM のページでは¹、計算機科学の分野の分類を提供されている。広く知られるように、分野の分類は階層構造をもたせることで、様々な粒度で議論が可能となる。これに関し、文献 [5]、ACM どちらも、分類を階層的に定義している。

¹<https://dl.acm.org/ccs> [2020 年 3 月閲覧]

2.2 時系列

時系列といえば、ある時刻 t の値 $f(t)$ が $\tau < t$ であるような $f(\tau)$ に依存して決まる系列のことである。この系列を微分方程式として定式化し、未来の挙動を予測することは広く行われている。19 世紀以降、ブラウン運動をはじめとする確率的な系列の研究が進み、確率過程として経済学、物理学をはじめとする様々な分野に適用可能な理論的枠組みが完成している [6]。

3 調査方法

既存の調査手法に効率性へ言及したものが見あたらないが、限られた時間において調査を繰り返し行う為に効率性の高さは必要である。そこで、本稿では、次に示す調査全体を 5 段階に分けた調査方法を提案する。

- 事前調査: 媒体と文献の傾向を調査
- 一次選別: 文献を広範囲に網羅的に収集
- 二次選別: 大まかな内容確認による選別と分類
- 三次選別: 全文精読による選別
- 報告書作成

一次選別では、調査範囲の偏りを防ぐために、統計的かつ機械的手法を用いて文献を網羅的に収集することを目的とする。範囲を広くするために一文献あたりの作業量を減らす必要があり、調査方法は文献の表題のみの確認、特徴語による機械的検索などに限定したため、不適切な文献が結果に含まれることが想定される。

二次選別では、上述の不適切な文献を除外を行う。ここでは内容の確認を伴うが、同時に報告書の作成に必要な体系化なども行ったほうが効率的であり、かつ理解の助けにもなると考えるため、本調査では文献の分野分類や、当該分野における位置づけの把握も同時に行う。この時、効率と柔軟性を向上させるために、導入部のみを把握することで文献を選別する。医療系における症例把握目的など、内容の見逃しが許されない調査では二次選別の段階で全文を確認することが望ましいと考えるが、本調査は研究動向調査であり、網羅性だけでなく時間的負担を減らし結果として全体の質にかけ余力を確保することが重要であるとの観点から、この方式とする。

三次選別では、全文を詳細に確認し、必要に応じて体系的にまとめ最終報告書作成のための準備を行う。不適切な文献は除外する。

上記大枠による作業は人でのみでも可能な範囲であるが、これに加え、本調査では近年利用可能である、検索エンジン、高い計算機性能を活用した調査とするた

めに、上述した事前調査、一次、二次、三次選別のそれぞれの間で、計算機による統計処理を行えるように全体の方式を設計する。以降の節でこれらの詳細について説明する。

3.1 事前調査

3.1.1 調査媒体の予備的調査

調査媒体には様々な特徴がある。例えば、表題、引用、節の構成は各会議で異なり、論文が提供される媒体には複製防止策（コピープロテクト）が施されていたり、入手困難な文献が存在したりするため、本格的な調査を開始する前に事前調査を行い、調査方法の詳細を決定する。このために、本調査でははじめに、調査協力者全員に各自興味を持つ文献を報告してもらい²、その作業を通して媒体の特徴を把握してもらう。また、文献の内容を報告書にまとめてもらうことで、調査対象の分野への足がかりとなる知見の獲得を目指す。

報告書には当該文献中で時系列に関する研究分野で関連が深いと思われる特徴語を近傍語として記載してもらう。この時得られた近傍語を用いて文献の検索を行うことも考えられたが、本節の作業により選ばれた文献は恣意的であるため、次の節で説明する無作為な予備調査により得られる近傍語を、このような検索語拡張に用いる。

3.1.2 一次選別のための無作為な予備調査

前節の作業により、調査媒体の仕様を確認したのち、検索エンジンを利用しより広範囲かつ無作為な文献の抽出を行う。さらに、調査範囲に対し網羅的な文献を検索結果として得られるように、次の手順により検索語を拡張する。二次選別以降の指針となるよう、同時に各研究の分類を木構造により行うことも試みる。

- 検索語「時系列」により文献を検索
- 得られた文献の導入部を読み、関連深いと考えられる単語を収集し、それらを近傍語と定義
- 簡単な分類を作成

3.2 一次選別

3.1.2 節の作業で得られた近傍語に対し、調査協力者は検索語に相応しいどうか判定し、検索語を選別する。この具体的な手順を以降説明する。

²YAML 形式の様式まとめて報告をしてもらい、後の解析を容易にした <https://yaml.org/spec/1.2/spec.html> [2019 年 12 月閲覧]

評価者（調査協力者）の集合を U 、近傍語の集合を W_{nbr} とする。 $u \in U, w \in W_{\text{nbr}}$ に対し、 $k(u, w)$ は u が w を採用すべきと主張した時 1 を取り、そうでないとき 0 を取るものとする。以上の定義のものと、 w の評価 $\tilde{s}(w)$ を式 (1) により計算する。

$$\tilde{s}(w) = \frac{\sum_{u \in U} k(u, w)}{\sum_{u \in U, w \in W_{\text{nbr}}} k(u, w)} \quad (1)$$

また、近傍語の中に存在する同義語を一つにまとめるために、代表語を定義する。この代表語の集合を W_r と表し、 $w \in W_r$ に対して同義語の集合を $W_{\text{id}}(w)$ とする。以上の定義から、代表語 w のスコア $s(w)$ を次式で計算する。

$$s(w) = \frac{\sum_{w' \in W_{\text{id}}(w)} \tilde{s}(w')}{\max_{w'} \tilde{s}(w')} \quad (2)$$

式 (2) では同義語の多い代表語が過大に評価される可能性があるため、近傍語への評価を行うとき、評価者には同義語に対してできるだけ重複した採用結果としない評価とするよう依頼し、重複による過大評価となる危険性を防ぐ³。 $s(w)$ の値が高い方から順に、幾らか選び、検索語とする。検索語の数は再現率を確保する必要性と検索の負担を考慮した上で決定する。

次に、検索語を用いて文献を選別する。本調査では検索語を J-Stage⁴が提供する検索エンジンへ一つづつ入力し、得られた文献を記録した後、検索語の評価値を合計し、文献に優先度を定義し、優先度の高い方から順に文献を選ぶ。選ぶ文献の数は先ほどと同様に再現率を確保する必要性と、二次選別の負担を考慮した上で決定する。

次に上述した文献の優先度の定義を具体的に述べる。文献の集合を D と表し、文献 $d \in D$ に含まれる単語の集合を $W(d)$ とする。また、 $w \in W$ に対する文献 d の検索結果を示す関数を $I(w, d)$ とし、式 (3) で定義する。

$$\begin{aligned} [\exists w' \in W_{\text{id}}(w), w' \in W(d)] &\Rightarrow I(w, d) = 1 \\ [\nexists w' \in W_{\text{id}}(w), w' \in W(d)] &\Rightarrow I(w, d) = 0 \end{aligned} \quad (3)$$

この定義のもと、文献 d の評価 $s_1(d)$ を式 (4) で定義する。

$$s_1(d) = \sum_{w \in W_r} I(w, d) s(w) \quad (4)$$

³同義語をまとめてから評価を行うべきであったが、今回の方式では同義語の選定と近傍語の評価を同時に行ったため、評価にあたりこのような煩雑な注意を与えている。本来は同義語のみを評価したほうが良い。

⁴<https://www.jstage.jst.go.jp/browse/-char/ja> [2019 年 12 月閲覧]

3.3 二次選別

二次選別では人による評価を重視した選別方法とする。ここで、一次選別を通過した文献は単語による検索によってのみ選ばれたものであり、調査内容に対し不適切なものも多く含まれているため、これらを除外する必要がある。このために、まずは主に導入部のみ把握することで、大まかな選別を行う。また、導入部を把握することで得られる知見から、分類を作成する。分類の作成にあたり、3.1.2 節の結果や関連研究に上げた文献の分類を参考にする。

報告書では、動向を研究の分野、つまり手法などの分類毎に体系的に説明することが望ましいため、分類毎に評価者の担当をわけ、精読、つまり三次選別に望む方式が良いと考える。このときの担当分けの目安とするために、選別中、文献に対する自らの専門分野との関連性を 3 段階、「関連なし」、「少し関連有り」、「専門に該当」で記載してもらう。

3.3.1 選別方法

一次選別の結果得られた文献の列 $d = d_1, d_2, \dots, d_n$ の各要素の全体を $D_{1\text{st}}$ とする⁵。ただし n は文献の数で、 $|D_{1\text{st}}| = n$ ある。つぎに、評価者 u ($\in U$) の $d \in D_{1\text{st}}$ に対する評価を $I(u, d)$ とし、 u が d を参考にすべきと判断した場合、 $I(u, d) = 1$ 、参考にする必要がないと判断した場合 $I(u, d) = -1$ とする。また、これらいずれにすべきか判別できないもしくは評価を行っていない場合 $I(u, d) = 0$ であるとする。以上の定義において、式 (5) 中 s_2 の値が 1 以上の文献を採用し、その集合を $D_{2\text{nd}}$ とする。

$$s_2(d) = \sum_{u \in U} I(u, d) \quad (5)$$

ただし、 $\sum_{u \in U} |I(u, d)| \leq 3$ を満たすものとする⁶。

3.3.2 分類の作成

報告書を分類を目安にまとめることを想定しているが、このために、選ぶべき分類を定量的に議論する必要があるため、ここでは分類の定式化を行う。

分野の分類は階層構造を持つものと仮定する。分類全体の集合を V とし、 $v \in V$ に対し v の下位に属する分類の集合を $f(v)$ ($\subset V$) と表す。また、 $v \in V, d \in D_{1\text{st}}$ に対し定義された分類を $f'(d, v)$ と表す。ここで、3.3 節の報告書に記載された分類は断片的なものであるため、

⁵ $\{d_1, d_2, \dots, d_n\} = D_{1\text{st}} \subset D$ である。

⁶つまり、1 か -1 の評価を行った評価者が 3 人となった時点で当該の文献にはそれ以上評価者を割り当てないものとする。また、 $|s_2(d)| = 2$ となった時点で、同様にこれ以上評価を行っても結果が変わらないため評価者の割り当てを停止するものとする。

これらを統合し分類体系を作るために、 $f(v)$ を式 (6) で定義する。

$$f(v) = \bigcup_{d \in D_{1st}} f'(d, v), v \in V, \quad (6)$$

また、文献 d の評価 $s_2(d)$ を用い、分類 v の評価 $s_d(v)$ を式 (7) で定義する。

$$s_d(v) = \ln \left(\sum_{d \in D_{1st}} s_2(d) \min(I(u, d), 0) \right), v \in V, \quad (7)$$

次に、得られた分類の構造と、分類それぞれの評価値から、着目すべき分類を抽出する方法を説明する。

式 (6) で分類を階層的に定義しているため、分類の数は階層が深くなるにつれ指数関数的に増加すると推測する。このため、式 (10) に示す指数分布から分類 v が生成されるとモデル化すれば、分布をよく表現する統計量が得られると考える。

$$p(v) = \lambda \exp(-\lambda s_d(v)) \quad (8)$$

このとき、分類の集合 V に対し、この分布に対する対数尤度関数は以下で定義される。

$$l = |V| \ln \lambda - \lambda \sum_{v \in V} s_d(v) \quad (9)$$

上記対数尤度の λ による偏微分が 0 となるときの λ を求めることで、最尤推定値 $\tilde{\lambda}$ は式 (10) により求めることができる。

$$\tilde{\lambda} = \frac{|V|}{\sum_{v \in V} s_d(v)} \quad (10)$$

この推定量 $\tilde{\lambda}$ を使い、着目すべき分類を推定することを以降考えるが、本調査では次の仮定を用いることとする。

1. 得られた分類は正しいものではなく、真の分類の一部を切り取ったものである。
2. 分類の階層関係は見かけの他に、調査の着目点に従い暗黙的に決定されている。

1 の仮定より、理想的な分布を仮定することが可能である。理想的な分布とは、どの階層でも対称性の高い議論が可能であることを要請すると考えるため、分類の階層構造がある一つの定数 c ($\in N$) に対し、 $\forall v \in V, |f(v)| = c$ を満たすと仮定する。つまり、最上位分類から c ずつ均等に枝分かれしていくものとする。このとき、すべての階層間の要素数の比は c となるため、 i 番目の階層に属する要素の数 n_i は次のように表される。

$$n_i = c^{i-1} \quad (11)$$

上記のような分類を行うことは一般に困難であるが、高度に統計的振る舞いをする系、つまり近年見られる高度に枝分かれした学術体系では潜在的に成り立っていると仮定することは妥当だと考える。

そして、2 の仮定より、得られた分類体系の最上位が「時系列」⁷でなくとも、時系列データの動向を調査するのに適切な文献から生成されたものであるため、「時系列」を最上位に持ってきた階層構造を構築可能な分類集合が得られていると考える。

以上の解釈に基づき、「時系列」以下の分類の数を c とすることを目標としたとき、式 (8) の分布の「時系列」を示す地点から、一様な階層構造を仮定した場合に、小分類が c となるまでの区間内に入る分類を、調査報告で着目する分類の候補とする。具体的には、 $v_{ts} (\in V)$ を「時系列」に対応する分類とし、区間 $[s_d(v_{ts}) - \Delta s_d, s_d(v_{ts}))$ 中に入る分類を求める。このとき、 Δs_2 は次のように求める⁸。

$$\begin{aligned} c^{-1} &= \frac{\lambda \exp(-\lambda (s_d(v_{ts}) - \Delta s_d))}{\lambda \exp(-\lambda s_d(v_{ts}))} \\ &= \exp(-\lambda \Delta s_d) \end{aligned}$$

すなわち、式 (12) となる。

$$\Delta s_d = \frac{1}{\lambda} \ln c \quad (12)$$

v_{ts} よりも評価⁹の低い分類を選ぶ理由は、 v_{ts} よりも評価の高い分類は、多数ある子分類の親となっている、着目すべき分類としては粒度が大きすぎると考えるためである。

得られた分類の候補を議論により、その数を c となるように統合するして調整する。

3.4 三次選別

二次選別時に記載してもらった関連度を元に、前節で得られた分類への最適割当を組み合わせ最適化法により決定する。その後、二次選別を通過した文献を各分類に割当て、分類担当者は精読を行い、必要に応じて除外や、他の分類に移すなどの対処を行う。ここで、報告書としてまとめるために、各研究における共通点などをまとめておく。

⁷「時系列データ」の同義語として扱っているが、分類体系としてデータが付くことは妙なので、単に時系列としている。

⁸本来比ではなく、 $c^{-1} = \int_{s_d(v_{ts}) - \Delta s_d}^{s_d(v_{ts})} \lambda \exp(-\lambda x) dx$ を求めるべきであるが、調査中の間違いにより、結果は式 (12) により出されている。しかし、今回の場合はどちらを用いても結果に影響はない。

⁹ $s_d(v)$ の値、つまり v の出現頻度と文献の評価の積の対数である。

表 1: 近傍語および重み

重み	近傍語
1.00	時系列データ (time series data)
0.78	市場動向予測 (market trends predicting)
0.67	行動分析 (behavior analysis)
0.67	トレンド予測 (trend prediction)
0.56	株価予測 (stock prices prediction)
0.56	セグメンテーション
0.56	時系列文節化
0.56	センサデータ (sensor data)
0.44	注文データ (pos data)
0.44	シーケンシャルパターンマイニング
0.44	時間的粒度
0.44	モーショントラッキング
0.44	隠れセミマルコフモデル

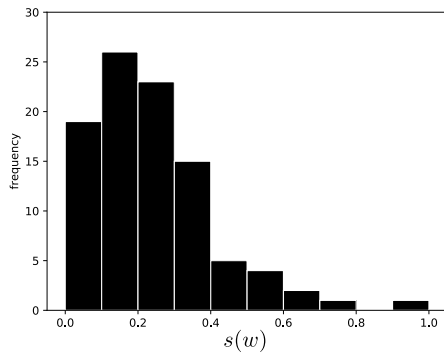


図 1: 検索語 w に対する重み $s(w)$ の頻度分布

4 調査結果

4.1 事前調査の結果

事前調査では、時系列データという語を含む文献は数 100 件以上と多数存在することを確認したため、効率的な選別方法の考案が必要であることを確認した。また、「時系列データ」の近傍語を選出した。選出された近傍語を元に得られた代表語と $s(w)$ (式 (2)) の値の一覧を表 1 に示す。また、頻度分布を図 1 に示す。

網羅性と作業量の兼ね合いを考慮し、検索語は 16 程度が良いと判断した。また、重みの高い方から順に並べたとき、 $s(w_1) > s(w_2) > \dots s(w_{12}) = s(w_{13}) > s(w_{14}) \dots$ であったため、13 番目までを検索語として採用した。また、13 番以降の単語のうち、特に含めるべきと判断した「強化学習」、「再帰的ニューラルネットワーク」、「クラスタリング」を検索語に含め、これら計 16 語を最終的な検索語とした。

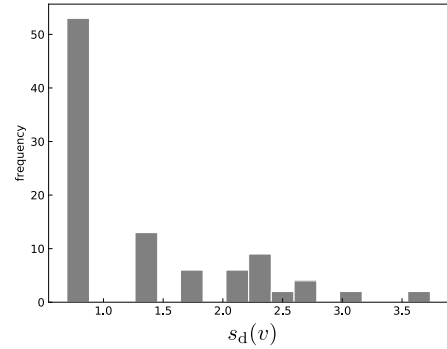


図 2: 分類 v に対する評価 $s_d(v)$ の頻度分布

4.2 1 次選別の結果

網羅性と作業量の兼ね合いから、上位 32 件程を二次選別の対象とすることに決定した。しかしながら、 $s_1(d_1) \geq s_1(d_2) \geq \dots$ となる $d_1, d_2, \dots (\in D)$ の列に対し、 $s_1(d_{32}) = s_1(d_{33}) = s_1(d_{34}) \neq s_1(d_{35})$ であったため、 d_{34} までを二次選別対象とした。検索語が一つでも該当する文献は 451 あったため、採択率は約 0.075%である。

4.3 二次選別の結果

二次選別の結果残った文献は 21 件で採択率は 61.8%であった。除外理由として、「時系列データに関係がない」、「内容の読み取りが困難である」などがあつた。この二次選別であるが、他の評価者の報告結果が見れる状態で行った。本来は評価の独立性を保つために、他人の結果を最後まで秘匿すべきであつたが、円滑に行うために様式の記載方法、分類の情報共有をしたほうが良いとの判断から、秘匿せずに行った。

二次選別の結果得られた分類が、指数分布に従うという仮定の妥当性を検証するために、 $s_d(v)$ の頻度分布を計算したところ、図 2 に示す結果となった。これより仮定はある程度妥当であると考ええる。なお、分類の総数は 97 である。

分類の選択に使う式 (12) 中 c の値について、本調査では調査員が 8 人であったため、各分類精査に複数人を割り当て可能な最大値である 4 とし、分類の候補区間を決定した。この区間 $[s_d(v_{ts}) - \Delta s_d, s_d(v_{ts})]$ に属する分類の一覧を表 2 に示す。また、この結果を議論によりまとめた結果を以下に示す。

1. 統計学と確率模型
2. 機械学習
3. マルチメディア処理
4. 人間主体型コンピューティング

表 2: 分類と評価値

$s_d(v)$	分類名
2.89	時系列
2.64	機械学習
2.48	確率過程
2.48	統計学
2.30	人文社会科学
2.08	センサとアクチュエータ
2.08	ロボット工学
1.79	教師なし学習
1.79	bayes モデル
1.39	マルチモーダル
1.39	自然言語処理
1.39	人間主体型コンピューティング
1.39	動作追跡
1.39	認知心理学

1, 2については上位からそのまま採択した。3のマルチメディア処理についてだが、センサデータ、マルチモーダルは個別に扱われることが少なく、近年では多数のセンサからのデータを複合することが一般的である、また、センサに限らず、様々な媒体が複合的に用いられると考えることから、この分類とした。最後に4であるが、人文社会学、認知心理学やロボット工学等が挙げられているが、これらは人を主体に置いた研究であると見受けられ、人間主体型コンピューティングにまとめることが可能と判断した。

4.4 三次選別の結果

調査報告作成のための担当分けに関し、3.4節で述べたとおり組み合わせ最適化を試みたが、関連度に8人中6人の評価者が「関連なし」とだけ記載していたため、2人を除き無作為な組み合わせとなった。除外すべきと判断された文献は特になかったが、精読を行っていく中で、統計学と確率モデルは数理模型としてまとめられると考え、第一分類を数理模型に変更した。また、各分類を特徴づける研究の内容について記録を残した。

4.5 研究動向について

4.5.1 数理模型

ここでは数理模型の観点から研究動向を把握する。使われる手法は主に連続値を対象とする（連続模型）か、離散値を対象とする（離散模型）かに分けられる。また、実世界から得られる時系列データは、もともと連続値であるため、本来は連続値を模擬可能な模型を

使うべきとも考えられるが、センサから得られるデータはサンプリング周期により離散化されていることから、離散模型の適用も可能である。

連続模型を用いている文献は9件 [7,9,10,11,12,14,17,20]、離散模型のものを用いている文献は13件であった [8,13,15,16,18,19,21,22,23,24,25,26,27]。市場や生活様式のような、対象とする範囲が年単位に及ぶ文献は、すべてにおいて離散模型を用いていた。連続模型は身体の動作など数分以内の規模においてのみ利用されていた。RNNは連続模型に用いられることはなく、この理由として、RNNは自然言語分野で特に発達している影響から、連続値に対する適用事例が少なく、応用がしづらいのではないかと考える。

時系列を分節化する研究が多く見られた。時系列を分節に分けることは、音声認識では文字と発音を対応付けるために一般的であるが、人の生活活動や [10]、動作に意味付けを行う方法としても盛んに用いられていた [7,9,10,11]。分節化を行う手法に関しては、自然言語における形態素解析と同様に隠れ Markov 模型を使い潜在因子の系列を推定する方法と [7]、時系列の特徴量をクラスタリング (k-means 等) により分割することで推定する手法が見られた [16]。HMM と k-means 以外によく用いられていたのは DTW (Dynamic Time Wrapping) である。DTW は時系列の伸縮による相似性、例えば歩く速度が異なってもその動作は概ね同じであること、を考慮し、時系列データ間の類似度を測ることが可能な手法である。DTW を適用する参照用時系列を予め教師有り学習などで作成しておき、適合する区間を特徴ある時系列区間として定義することが可能である [17]。

4.5.2 機械学習

多くの文献で機械学習手法が用いられていた。特に、隠れ Markov 模型が多く、Neural Network (NN) 系では、畳み込み NN、長短記憶を含む再帰的 NN が用いられていた。

用いられている模型はおもに確率的生成模型であった。確率的生成模型としては Recurrent Neural Network (RNN)、隠れ Markov 模型が用いられていた [7]。トピックモデルも見られたが、特徴抽出にトピックモデルを利用し、それらの時間的変化は RNN で捉えるという枠組みで利用されていた [18]。

生成模型で、実現値として時系列データが連続値であることを仮定している模型では回帰を用いており、これは当然といえるが、時系列データの未来を予測する模型において、回帰を用い外挿するのではなく、変化を検知する2値識別模型とする文献がいくらかみられた [15,19,26]。この理由として、文献 [26] は噴火予測で、ほかは市場動向予測であるが、変化が起きるかど

うかわかることが最も重要であり、それ以降の対応はあまり考慮する必要のない問題を対象としているということが共通すると考える。

4.5.3 マルチメディア処理

センサデータの中から知見を見出そうとする研究がいくつか見られた [10, 20, 21, 22, 23, 27]。これは、近年の IoT (Internet of Things) の普及により、センサから多量に時系列データを取得する下地が整ってきているためだと考える。

文献 [21, 22, 24] では、得られたデータを解析する前段階として、センサから得られた時系列データを管理するシステムを提案している。これらは直接時系列データの解析手法を提案していないが、将来的にシステムへ蓄積されたデータを解析することを目指す研究に位置づけられている。

文献 [20] はセンサデータに含まれる欠損に対処する手法を提案している。特に、工学系のセンサでは、欠損に瞬間的なものから長期的なものなど様々な種類があることを言及しており、それらを統一的に扱うことが可能なモデルを提案している。

文献 [27] は [20] とは別の観点で、センサデータ特有の問題である、雑音の影響を除去する手法に取り組んでいる。畳み込み NN と再帰的 NN を組み合わせることで、雑音に埋もれる信号を検出することに成功している。

文献 [8] は、複数センサを同時に使う場合に、それぞれのセンサ間で発生する時間遅れを考慮することの重要性を主張している。例えば同じ事象に起因する変化であっても、観測地点ごとにずれが生じる。このことを考慮しなければ、分節の境界が曖昧になるため、分節化が難しくなる。文献 [8] ではこれに対し時間遅れを考慮したモデルを提案し、考慮しない場合に比べ、分節化の精度が高くなる結果を示している。

4.5.4 人間主体型コンピューティング

ヒューマン・コンピュータ・インタラクションやヒューマン・ロボット・インタラクションのような、人間が介在する研究領域においても時系列データに関する研究事例が複数確認される [11, 21, 24]。本節では、人間主体型コンピューティング (Human Centric Computing) と、関連分野における時系列データを対象とした研究事例を述べる。時系列データに対する当分類に該当する研究は教育学や動物行動学、ロボット工学 [11, 21] のような、学際的な研究として発表される傾向がある。

教育学における時系列データを対象とした研究事例として、鍵山らは学習内容の理解の補助を目指した可

視化システムの改善手法を提案している [24]。この手法では、学習者の理解内容を node-link ベースの可視化手法であるキットビルド概念マップにより可視化し、学習内容の時間的変化を複数のマップの重畳により表現する。学習者の知識の時間的変遷を外在化することで、教育者による指導方法の効果の理解を促進できると主張している。

上述の文献のように、人間主体型コンピューティングにおける時系列データを対象とした研究は、時系列データの量的な側面よりもマルチモーダルデータとしての質的な側面に焦点を当てている傾向がある。他説とは異なり、データの時間的粒度が大きい [24]、個人単位のデータを対象とした分析を行う [14]、データの巨視的な傾向把握を目的としている点が当該分類の研究において特徴的である。

5 まとめ

本稿では、人工知能分野における時系列データを扱った研究動向の調査結果を報告し、また、効率を重視した体系的調査手法を提案した。提案手法により、定量的に着目すべき観点の候補を推測し、4つの観点、「数理モデル」、「機械学習」、「マルチメディア処理」、「人間主体型コンピューティング」からみた動向をまとめた。

今後、同じ調査を別の媒体を行うことで、より洗練された知見が得られると考える。また、時系列データを扱った研究以外の動向調査を通じて、提案した調査手法の検証を行い、より一般性の高いものとする事ができると考える。

参考文献

- [1] Sacha, D., Kraus, M., Bernard, J., Behrisch, M., Schrech, T., Asano, Y., and Keim, D.: SOM-Flow: Guided Exploratory Cluster Analysis with Self-Organizing Maps and Analytic Provenance, IEEE Transactions on Visualization and Computer Graphics, Vol. 24, No. 1, 2018.
- [2] Waffenschmidt, S., Knelangen, M., Sieben, W., Buhn, S., and Pieper, D.: Single Screening Versus Conventional Double screening for Study Selection in Systematic Reviews: A Methodological Systematic Review, BMC Medical Research Methodology, Vol. 19, No. 132, 2019.
- [3] Vaidhehi, V., Suchithra, R.: A Systematic Review of Recommender Systems in Education, International Journal of Engineering & Technology, Vol. 7, No. 3.4, 2018.

- [4] Dutt, A., Ismail, M., Herawan, Tutut.: A Systematic Review on Educational Data Mining, IEEE Access, Vol. 5, 2017.
- [5] 山口昌哉, 佐和隆光, 坂本賢三, 富永茂樹: 学問の現在, 1989.
- [6] 伊藤清: 確率過程, 岩波書店, 第2版, 2007.
- [7] 長野匡隼, 中村友昭, 長井隆行, 持橋大地, 小林一郎, 高野渉: HVGH: 高次元時系列データの深層圧縮と教師なし分節化, 2019年度人工知能学会全国大会(第33回), 1L3-J-11-01, 2019.
- [8] 真矢滋, 山口晃広, 稲木達哉, 植野研: 多変量時系列データの柔軟な分割方法の提案, 2019年度人工知能学会全国大会(第33回), 1I4-K-2-05, 2019.
- [9] 長野匡隼, 中村友昭, 長井隆行, 持橋大地, 小林一郎, 金子正秀: ノンパラメトリックベイズ法に基づく時系列データの分節化, 2018年度人工知能学会全国大会(第32回), 2G4-04, 2018.
- [10] 岩田健輔, 池田成満, 青木達哉, 西原成, 中村友昭, 長井隆行: 動作の分節化に基づく家事タスクにおける行動の構造化, 2016年度人工知能学会全国大会(第30回), 1O2-4, 2016.
- [11] 布川遼太郎, 宮澤和貴, 中村友昭, 長井隆行, 金子正秀: 時系列マルチモーダル情報の分節・分類に基づく物体と動作の統合概念学習, 2018年度人工知能学会全国大会(第32回), 1G3-05, 2018.
- [12] 中村友昭, M., Attamimi, 長井隆行, 持橋大地, 小林一郎, 麻生英樹, 金子正秀: ガウス過程の隠れセミマルコフモデルに基づく身体動作の分節化, 2016年度人工知能学会全国大会(第30回), 1O3-5, 2016.
- [13] 野中尚輝, 中山浩太郎, 松尾豊: RNNを利用したコンテンツ産業の消費トレンド予測, 2016年度人工知能学会全国大会(第30回), 1D2-2, 2016.
- [14] 三村喬生, 中村友昭, 松本惇平, 西条寿夫, 須原哲也, 持橋大地, 南本敬史: 霊長類における身体動作時系列の分節推移構造推定, 2019年度人工知能学会全国大会(第33回), 1C4-J-3-01, 2019.
- [15] 松井藤五郎, 和泉潔: 新聞記事の時系列テキスト分析による株式市場の動向予測, 2016年度人工知能学会全国大会(第30回), 3L3-OS-16a-6, 2016.
- [16] 中塚祐喜, 白山晋: 外生変数を伴う長期予測における主要因子推定法, 2019年度人工知能学会全国大会(第33回), 1I4-J-2-03, 2019.
- [17] 中村凌子, 鈴木浩子, 渡部勇, 高木友博: Shapeletsによる時系列特徴抽出とオンラインギャンブルでの問題行動の予測, 2019年度人工知能学会全国大会(第33回), 4A2-J-3-04, 2019.
- [18] 内田脩斗, 吉川大弘, 古橋武: FKC コーパスを用いたトピック遷移分析手法に関する検討, 2019年度人工知能学会全国大会(第33回), 4A2-J-3-03, 2019.
- [19] 和田裕貴, 長尾智晴: Deep Neural Networkを用いた株式売買戦略の構築, 2016年度人工知能学会全国大会(第30回), 4G4-1, 2016.
- [20] 武石直也, 矢入健久: 工学システムの間欠的センサデータのための動的混合モデル, 2016年度人工知能学会全国大会(第30回), 4M1-2, 2016.
- [21] 稲邑哲也: 大規模なマルチモーダル対話経験からの概念獲得を支援するクラウド型VRのアーキテクチャ, 2016年度人工知能学会全国大会(第30回), 1O3-1, 2016.
- [22] 川進也, 今田兼太, 神谷直輝, 石川翔吾, 竹林洋一: 認知症ケア改善のためのマルチモーダル映像センシング解析基盤の検討, 2016年度人工知能学会全国大会(第30回), 2H3-NFC-03a-4in2, 2016.
- [23] 浅野秀平, 切通恵介, 泉谷知範: 特徴アテンションを用いた時系列データのアトリビュション抽出, 2019年度人工知能学会全国大会(第33回), 2H3-J-2-05, 2019.
- [24] 鍵山貴一郎, 林雄介, 平嶋宗: 学習者集団および個人についての理解状態とその遷移の可視化の試み, 2016年度人工知能学会全国大会(第30回), 1C5-OS-13b-3, 2016.
- [25] 中川慧, 今村光良, 吉田健一: 時系列勾配ブースティング木による分類学習, 2016年度人工知能学会全国大会(第30回), 2J2-03, 2018.
- [26] 村田剛志, L., Heip, 井口 正人: Stacked Recurrent Neural Networkによる桜島噴火予測, 2018年度人工知能学会全国大会(第32回), 2A1-02, 2018.
- [27] 茂木貴弘, 中澤友哉, 田原鉄也: 1d CNN-LSTMによる調節弁内部の異常検知, 2018年度人工知能学会全国大会(第32回), 3Pin1-44, 2018.