

高等学校における問題解決型探究指導の知識モデリング ～PPDACサイクルオントロジーの構築と探究過程の可視化～

Knowledge Modeling and Visualization of the Inquiry based Learning Instructional Process in Japanese High Schools PPDAC Cycle Ontology

林宏樹^{1*} 笹嶋宗彦¹
Hiroki Hayashi¹ Munehiko Sasajima¹

¹ 兵庫県立大学
¹ University of Hyogo

Abstract: 高等学校では、問題を発見し解決する探究学習を通して問題解決能力の育成を目指している。しかし、探究学習の指導法が確立されておらず、指導経験の浅い教員が指導計画を立てることは困難な状況である。本研究では、統計的探究プロセスを基盤とした探究学習指導を支援するPPDACサイクルオントロジーを提案した。研究方法として、探究学習における知識の獲得、体系化を行い、構築した知識体系とPPDACサイクルの対応付けを行い、知識モデルを構築した。アンケート調査によって検証した結果、構築したオントロジーは探究学習の指導経験の浅い教師が指導計画を立案する支援を行うことができる可能性があることがわかった。

1 はじめに

テクノロジーの発展によって単純作業はコンピュータに代替され、人間に求められる力は高度化している。OECDが実施する国際学習到達度調査(PISA2012) [1]で相互作用的な操作を通じた問題解決力に関する測定が行われているように、近年、問題発見・解決能力の育成が重視されている。これらを踏まえ、日本の高等学校教育では、問題解決プロセスを用いた探究活動を行うことで問題解決力を育む教育活動が盛んとなってきた。

高等学校における学習指導要領では問題発見・解決力の学習基盤となる資質・能力の育成が求められている。ここで、学習指導要領とは、文部科学省が定める日本の小学校、中学校、高等学校の各教科の到達目標や、教育すべき内容を定めたものである。全ての教員は学習指導要領に定められた内容に基づいた指導を行うため、学習指導要領は国全体の指導内容を方向付ける基準となっている。

問題解決力の育成の1つの方策として、学習指導要領の改訂によって「総合的な探究の時間」が設立された。「総合的な探究の時間」では、解決の道筋がすぐに明らかにならない課題や、唯一の正解が存在しない課題に対して、最適解や納得解を見いだすことを重視し

ている点が挙げられている [2]。このような「問題解決を目的とした探究活動」(探究活動)が問題解決力の育成のために必要となり、高等学校教員は生徒が探究活動を行う支援となる探究学習の指導(探究指導)を行う必要性が高まってきた。

生徒が探究活動を円滑に進めていく1つの手段として、問題解決プロセスに従って指導する方法がある。そこで、高等学校学習指導要領や教員研修用教材に提示されている問題解決プロセスについて整理する。一つ目として、総合的な探究の時間では「課題の設定、情報の収集、整理・分析、まとめ・表現」の探究の過程の4つのプロセス [2]がある。第二に、数学科「数学I」では「問題-計画-データ-分析-結論」の統計的探究プロセス [3]がある。第三に、情報科「情報I」教員研修用教材では「問題の発見、情報の収集と分析問題の定義、計画の立案・解決方法の探索、計画の実行・結果の予測、振り返り、次の問題解決へ」の学習活動の流れ [4]がある。これらの3つのプロセスは、問題解決につながる研究経験があり、大学などで論文を書いたことのある人ならば、実質的に同じプロセスであることは理解できるであろう。しかしながら、稲永 [5]は修士レベルの学識経験がないと探究学習の指導は難しく、探究学習の指導を行うには学問的な経験とともに一定の経験が必要であることを指摘している。現状調査すると、現役の高等学校教員の学歴構成は、学士修了者の割合が84.1%、大学院修了者の割合が15.9%であ

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

る [6]. このことから、日本の高等学校教員がこの3つのプロセスを同じであると判断できるかどうかは疑問である。そのため、多くの高等学校教員が問題解決プロセスに従って探究指導を行うのは困難であることが想像できる。本稿では、問題解決プロセスに関連した経験が少ないまたは未経験である、もしくは、探究指導の経験が少ないまたは未経験である教員を「指導経験の浅い教員」と呼ぶことにする。関連して考えると、指導経験の浅い教員が問題解決につながる探究学習の指導計画を立てることも困難であると考えられる。例えば、「問題発見」の段階では、問題を発見するまでの状態遷移の過程が暗黙的に含まれており、どのような状態が存在し、その状態を遷移するためには何を達成すればよいかのわかりづらい。このような過程が暗黙的に含まれている探究指導において、学習指導要領が示す問題解決プロセスは、その定義が詳細ではないため、指導経験の浅い教員が指導計画を立てるための参考にはなりにくいと考えられる。

本研究では、問題解決プロセスを活用した探究学習の指導方法の標準化を目指す。まず、問題解決プロセスに関する知識を獲得し、獲得した知識を体系化する。次に、知識体系と問題解決プロセスのフェーズを対応させる。探究学習の指導方法は暗黙的な要素が多いため、暗黙知を明示化するのに適したオントロジー構築ツールを用いることにした。本研究の成果は、探究指導の手引きの作成につながり、探究指導ができる教員の育成に貢献することが期待される。

本稿の構成は以下の通りである。第2章では、オントロジー工学に関する関連研究について述べる。第3章では、探究学習のプロセスをモデル化する方法について述べる。第4章では、PPDAC サイクルに基づいたオントロジーを構築し、モデル化する。第5章では、構築したオントロジーについて、探究学習の専門家として判断した高等学校教員や研究経験が豊富である大学の教員からのアンケート調査とインタビュー調査による評価について述べる。第6章では、結論と今後の展望についてまとめる。

2 オントロジー工学の関連研究

2.1 オントロジー構築について

本研究は、探究学習の指導経験が浅い教員でも理解し実践することができるような探究学習指導方法の標準化を目指している。日本の高等学校における探究指導の現状として総合的な探究の時間があるが、専門的な指導ができる教員免許をもった教員はおらず、他教科の教員免許をもった教員が割り当てられている。また、文部科学大臣の検定を経た教科書がない状態であ

る。そのため、教員自身の研究経験の有無に加え、探究指導の模範となる教員もおらず、教科書もなく、指導方法を確立するための体制が乏しい状態である。このような状況を解決する方法として、探究学習の指導方法のモデルを構築することが1つ挙げられる。指導方法のモデルがあれば、モデルを指針として指導内容の全体像を理解し、どのフェーズにどれだけの時間が必要であるのか、どのような状態になれば次の状態に遷移することができるのかを想定した年間指導計画を立てることが可能であると考えられる。そこで、本研究では、共通語の提供、暗黙情報の明示化、知識の体系化に適したオントロジー工学を用いることにした。

オントロジー工学とは、溝口 [7] によって提唱された、現実に存在する本質的な概念構造をコンピュータ上で表現するための理論と技術である。オントロジーで定義された概念は、知識を共有するための共通概念として利用することができ、知識の共有や使い勝手を向上させることができる。オントロジー構築では、is-a 関係に基づく概念クラスの構築が重要である。概念クラスは単なる分類を行うのではなく、概念間の意味的な差異を明確にすることで概念を分類する [9]。探究指導に必要な知識を獲得し、その知識の差異をプロパティによって明確に分類することで、生徒の探究学習の状態を把握し、次への状態に遷移するための指導助言ができる。オントロジーを構築することは、これまで暗黙的であった探究活動の状態を可視化し、具体的な指導方法を明示化することにつながる。よって、状態とプロパティが教員同士の共通語となり、一貫した生徒の指導が可能となると考えられる。

2.2 教育とオントロジー

教育分野に関連するオントロジーには以下のようなものがある：オントロジーを用いた情報教育目標の体系的記述 [12]、オントロジー教育における複雑な協同学習の場の設計を支援するオントロジー [10]、教育設計モデルのためのオントロジー教育設計モデルのためのオントロジー [11]、オントロジーを活用した授業設計システム開発 [13] などがある。教育分野では、暗黙的となっている情報の明確化や体系化のためにオントロジーが利用されている。

2.3 オントロジーと知識獲得方法

オントロジー構築のための知識獲得方法について、一般的なプロセス知識を構造化する方法 [14] や、介護現場ごとに依存して変化する固有プロセスを獲得する知識発見の方法、CHARM(Convincing Human Action Rationalized Model) と呼ばれる暗黙的行為の明示的記

述モデル [15][16] などがある。また、クラシックギター
の専門家から暗黙的な知識を獲得し手続き知識を構築
したドメインオントロジーの構築に関する研究がある
[17]。これらの研究を参考に、2種類の専門家から知識
を獲得することにした。1つは、探究指導に十分な経
験を持つ高等学校の教員である。もう1つは、研究経
験と研究指導が豊富な大学の教員である。専門家の知
識に基づいて、探究指導に関わる知識体系をモデル化
する。

3 探究指導プロセスのモデル化

探究指導における指導方法をモデル化するために、基
盤とする問題解決プロセスと PPDAC サイクルの説明、
探究指導の知識獲得とオントロジー構築の手順と評価
方法について述べる。

3.1 問題解決プロセス

問題解決プロセスに関する研究は盛んに行われてお
り、多くの問題解決プロセスが提案されている。その
中で、我々は日本の学校教育において、小学校、中学
校 [18]、高等学校 [19] の授業実践で用いられている統
計的探究プロセス Problem (問題) - Plan (計画) -
Data (データ) - Analysis (分析) - Conclusion (結
論) (PPDAC サイクル) に着目した。先進的な統計教
育を実施しているニュージーランドの学校教育 [20] で
も活用されている。本研究の成果は全国の高等学校教
員を対象に普及することを考えている。このことから、
学習指導要領に記載されているプロセスの1つである
PPDAC サイクルを採用することにした。

3.2 PPDAC サイクルについて

PPDAC サイクルは、Problem (問題) - Plan (計
画) - Data (データ) - Analysis (分析) - Conclusion
(結論) の5段階のフェーズに従って統計学を用いて問
題解決を行うためのフレームワークである。フェーズ
の内容について、学習指導要領 [3] では Problem は問題
の把握、問題の設定、Plan はデータの想定、収集計画、
Data はデータの収集、表への整理、Analysis はグラフ
の作成、特徴や傾向の把握、Conclusion は結論付け、振
り返りと表記している。深澤ら [21] では、Problem は
問題をとらえる、Plan は計画を立てる、Data はデー
タを集める、Analysis は分析する、Conclusion は結果
を読み取ると表記している。このことから、各フェーズ
の表記には探究活動を行う上で広義にも狭義にも解釈
できる表記となっており、具体的に何を指導したらよ
いのか指針が示されていないことがわかる。したがっ

て、獲得した問題解決プロセスの知識を体系的に可視
化し、PPDAC サイクルに対応させることで、探究指
導において PPDAC サイクルを活用した指導方法が明
確になると考えた。

3.3 知識獲得

日本の高等学校における探究活動の問題解決プロセ
スに関する知識を獲得するために、スーパーサイエン
スハイスクール (SSH) で開発されている評価を行うた
めのルーブリックが有効な資料となる。SSH とは、理
数系教育を通して、生徒の科学的な探究能力等を培い、
国際的に活躍し得る科学技術人材等の育成を行う文部
科学省が指定した中学校、高等学校である。ルーブリ
ックとは、数レベル程度の尺度と記述語から構成され
る評価基準表である。SSH 校では 2013 年前後から科
学的探究を通して獲得した力を評価することが意識さ
れ、多くの SSH 校が独自でルーブリックを作成し評価
に活用している。例えば、アメリカの科学的探究の議
論を参考に科学的探究力の質を判断するルーブリック
が開発されている [22]。

このような状況において、8校の SSH によって標準
化されたルーブリックが開発されている (標準化ルー
ブリック) [23]。標準化ルーブリックは、生徒の問題
解決プロセスに即した取組の質を判断するために、生
徒を捉える観点 (評価規準) と質の深まり (評価基準)
が明文化されている。このことから、先進的な問題解
決に関わる探究指導を行っている SSH 校が開発した
ルーブリックに基づいて問題解決プロセスの知識を獲
得することで、日本の高校生の実態に対応したもの
と考えている。

一方、標準化ルーブリックでは、問題解決プロセス
に特化しているわけではない。探究活動の評価を行う
ためのものであり、探究活動におけるフェーズ間の
つながりが表現されていない。探究活動の状態基準
を1から5の状態では表しているが、探究活動の過
程において状態を遷移する方法が明確に記述されて
いるとは言い難い。このことから、標準化ルーブリ
ックに基づいて獲得した知識をプロパティを用いて
体系的に記述することで、フェーズ間や状態間の遷
移を明確にすることが指導方法につながると考えら
れる。本研究では、評価のための標準化ルーブリ
ックとは異なり、探究指導の指導方法につながる可
視化がオントロジーによって明示化される。

3.4 オントロジー構築の手順

本研究では、次の手順で行うことにする。

手順1 問題解決プロセスに関する知識を獲得する。

手順2 獲得した知識を体系化する。

手順3 知識体系とPPDACサイクルに対応させて構造化する。

手順1では、標準ルーブリックの記述語を参考に、問題解決プロセスの状態に関連する用語を獲得した。獲得した用語は、概念間の重複、用語間の関係、用語がクラスかクラスでないかを考慮することなく作成した。手順2では、手順1で獲得した知識を体系化した。下位クラスは、プロパティを用いて上位クラスを特殊化した。手順3では、手順2で体系化した知識をPPDACサイクルに対応させて構造化した。各フェーズの状態を初期状態、中間状態、発展状態、終了状態として設定し、フェーズに対応した状態を割り当てた。構造化では探究学習の状態とPPDACサイクルの各フェーズとの対応関係を考慮した。

3.5 オントロジーの評価方法

探究指導の経験豊富な専門家がアンケート調査によって構築したオントロジーを評価した。本研究では、専門家は高等学校での探究学習の指導経験が4年以上、もしくは大学における研究指導がある実務経験者と定義した。第一著者が個別に60分から90分程度の時間を要してオントロジーの説明を行い、その後専門家はアンケートの回答を行った。アンケート内容は、オントロジーを評価するために21の項目から構成した。表1に質問項目を示す。回答は5件法(1:いいえ, 2:あまり, 3:どちらともいえない, 4:まあまあ, 5:はい)を用いた。

4 PPDACサイクルオントロジーの構築

PPDACサイクルオントロジーはオントロジー構築・利用環境「法造」[24]上で実装した。

4.1 知識の獲得

標準化ルーブリックでは、4つの観点があり、課題の設定、調査計画の立案と実施、情報収集と情報の評価、結果からの考察が設定されており、観点ごとに基準、徴候、指導方略が定義されている。これらを参考に、探究学習の状態に関する知識を獲得し、49の状態概念を含む93の概念を定義した。具体的には以下のように知識を獲得した。基準に基づいて、「問いを立てることができている」という記述は「問い状態」、「見通しをもった計画となっている」という記述は「実行不

明状態」、「実行可能状態」と概念化した。徴候に基づいて、「結果と考察が分離できず、結果のみとなる」という記述は「言語化状態」、「未考察状態」、「考察状態」と概念化した。

4.2 知識の体系化

獲得した知識と標準化ルーブリックに基づき、上位クラスと下位クラスをis-a関係によって体系化した。各観点に基づいて、上位クラスを「定式化状態」、「仮説設定状態」、「実行状態」、「データ収集状態」、「分析状態」、「結論状態」として特殊化を行った。以下に、具体例として「定式化状態」に関する知識を体系化した状態を図1に示す。

第一に、標準化ルーブリックにおける観点「課題の設定」の基準2「問いを立てることができている」、基準3「研究の目標を踏まえて、問いや仮説を設定できている」の記述から、問い状態から仮説状態へ状態が遷移することが読み取れた。

第二に、徴候3「曖昧な語を含んでいるものの、研究を通じて明らかにしたいことを目標や仮説といった形で表現できている」の記述は、「明らかにしたいこと」を問いに対する「予想した解」と判断し、プロパティに「予想した解の有無」のbooleanを設定し、Falseであれば「問い状態」、Trueであれば「仮説状態」として体系化した。

第三に、徴候3「仮説を立てている」と徴候4「数多くの実験をした上でそれを踏まえた仮説を立てている」、徴候5「先行研究がある場合、それらと比較できるような課題が設定できている」の記述は、「仮説状態」ではあるが客観的な根拠をもった仮説とそうでない仮説によって区別することが示唆されており、状態は変わらず、仮説の質の高さを区別する必要があると考えた。よって、「仮説状態」を特殊化し、「仮説設定・客観性有」、「仮説設定・客観性無」を設定した。

このような手順によって、それぞれの状態について、「仮説設定状態」、「実行状態」、「データ収集状態」、「分析状態」、「結論状態」に定義した概念を体系化した。「仮説設定状態」、「実行状態」は状態の遷移だけ、「定式化状態」、「データ収集状態」、「分析状態」、「結論状態」は状態の遷移と質の高さを表現することができた。本研究における構築したオントロジーは、ウェブページ¹に掲載する。

4.3 PPDACサイクルへの構造化

PPDACサイクルに基づき、獲得した知識体系を構造化した。まず、問題解決プロセスである各フェーズの

¹<https://sites.google.com/view/education-ontology>

表 1: 専門家に実施した質問項目

項目 1	オントロジーの可読性
1-1	PPDAC サイクルのオントロジーの概要を理解できたか
1-2	クラスとプロパティについて理解できたか
1-3	Problem 段階における生徒理解状態の分類を読み取ることができますか
1-4	Plan 段階における生徒理解状態の分類を読み取ることができますか
1-5	Data 段階における生徒理解状態の分類を読み取ることができますか
1-6	Analysis 段階における生徒理解状態の分類を読み取ることができますか
1-7	Conclusion 段階における生徒理解状態の分類を読み取ることができますか
項目 2	オントロジーの適切性
2-1	上位の概念は適切でしたか
2-2	プロセスの状態の分類は適切でしたか
2-3	Problem に含まれる状態の定義は適切でしたか
2-4	Plan に含まれる状態の定義は適切でしたか
2-5	Data に含まれる状態の定義は適切でしたか
2-6	Analysis に含まれる状態の定義は適切でしたか
2-7	Conclusion に含まれる状態の定義は適切でしたか
項目 3	オントロジーの有用性
3-1	PPDAC サイクルオントロジーによって、手続き的知識の理解が深まったか
3-2	手続き的知識の構造化において、PPDAC サイクルオントロジーの用語が役に立ちましたか
3-3	手続き的知識のプロセスが、PPDAC サイクルオントロジーで適切に分類されていたか
3-4	PPDAC サイクルオントロジーは、PPDAC サイクルに対する意識や認識を高めるのに役に立つと思えましたか
3-5	本オントロジーを用いることで、あなたは指導計画を立てることができたか
3-6	本オントロジーを用いることで、あなたは探究指導を行うことができたか
3-7	本オントロジーを用いることで、あなたは評価基準を立てることができたか

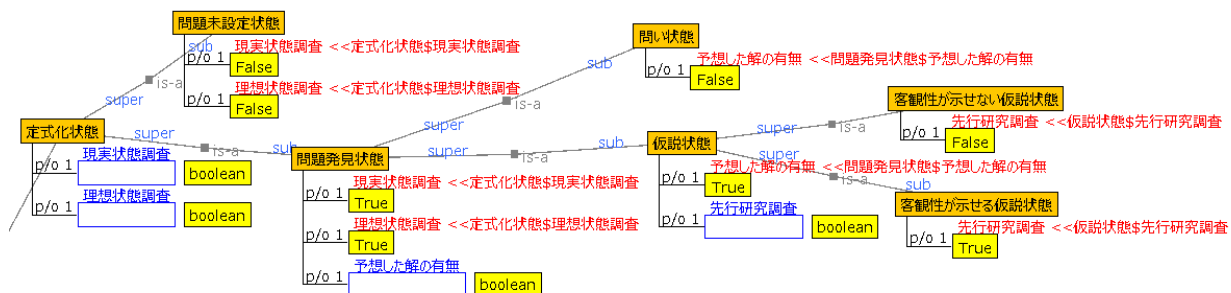


図 1: 定式化状態

上位クラス「プロセス」には初期状態，中間状態，発展状態，終了状態を割り当てた．初期状態と終了状態は必ず到達する状態であると定義した．次に，初期状態から終了状態へ遷移するためには，中間状態が存在する場合があり，それらの状態を配置した．最後に，必ず遷移するとは限らないが，状態の遷移ではなく，状態の質が高い状態を発展状態とし，該当する状態がある場合は配置した．図 2 に Problem フェーズ，図 3 に Plan フェーズと Data フェーズ，図 4 に Analysis フェーズと Conclusion フェーズを示す．

4.4 アンケート調査結果

専門家として判断した高校教員 (n=5) と大学教員 (n=4) を対象としたアンケート調査の結果が表 2 である．「はい」，「まあまあ」の回答を肯定的な回答とす

る．どの項目も概ね肯定的な回答が多く，各項目の平均値は 4.0 以上であった．高校教員と大学教員の全体回答結果から，肯定的な回答割合が 100 %であった項目は，オントロジーにおける概要の可読性，クラス・プロパティの可読性・適切性，Problem フェーズの可読性，Data フェーズの可読性・適切性であった．また，本オントロジーは PPDAC サイクルの認識向上に有効であることが示唆された．しかしながら，Plan，Analysis，Conclusion の各フェーズの可読性・適切性は肯定的な意見が少なく，検討の必要性が示唆された．

高校教員の回答結果では，5 名全員が「はい」と回答した項目は，1-2，1-7，3-1，3-2，3-4 であった．可読性については，クラスとプロパティの理解，Conclusion 段階の生徒理解状態の分類，有用性については，PPDAC サイクルの手続き的知識の理解，意識や認識の向上，オントロジーの用語の有用性の点において効果がある

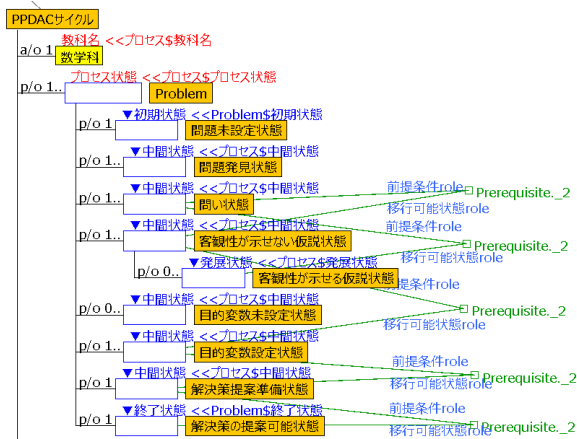


図 2: Problem フェーズ

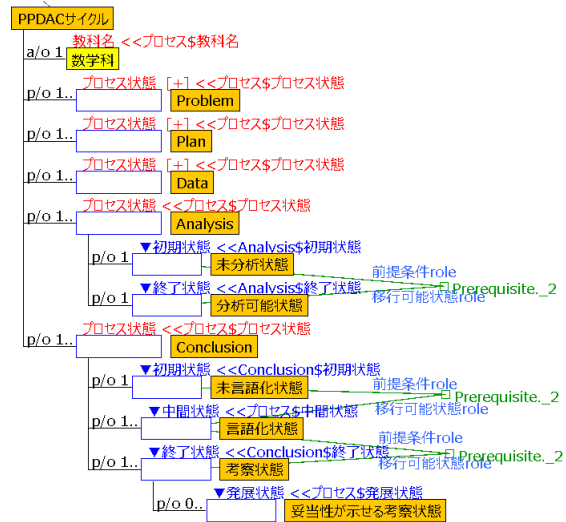


図 4: Analysis · Conclusion フェーズ

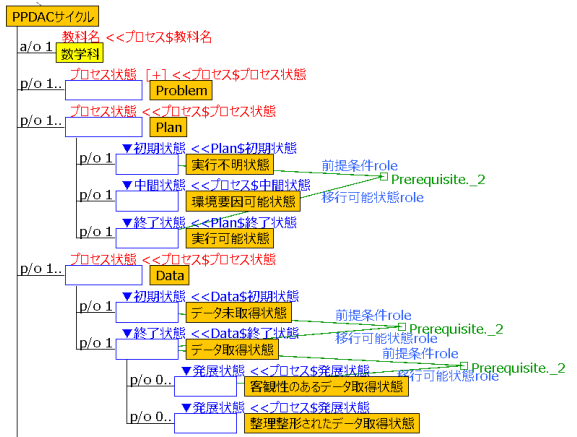


図 3: Plan · Data フェーズ

5.1 オントロジーの評価

高校教員と大学教員を対象としたアンケート結果の全体の回答によって評価した。オントロジーの可読性、適切性について、(1)オントロジーの概要の可読性、(2)クラス・プロパティの可読性・適切性、(3)Problem フェーズの可読性、(4)Data フェーズの可読性・適切性、以上の4項目で良好な評価を得た。

アンケートの記述回答では、探究指導の全体像が明確になったことと、専門家自身が探究指導をする上で暗黙的な用語を使用していることを認識できたという意見があった。具体的には、Problem フェーズを学習指導要領から判断すると「問題の把握、問題の設定」としか提示おらず、問題発見をするだけと捉えがちである。しかし、オントロジーで指導方法を明示化したことにより、探究活動では問題発見から仮説を設定し、解決策の提案まで実施する必要があることを明らかにできたことを支持された。このことから、Problem フェーズを明示的に定義して可読性を高めたことにより、Problem フェーズには多くの段階が含まれていることが理解でき、多くの指導時間を要することが明らかとなった。したがって、構築したオントロジーによって、経験の浅い教員でも探究活動の年間指導計画を立てることが可能となり、Problem フェーズの指導時間数には考慮する必要があることがわかった。

Data フェーズにおいては、単にデータを取得するだけで留まらず、データ取得状態の質を高めるための整理整形を行った状態が存在すること、整理整形はPPDACサイクルの発展状態に位置するために全員に指導する必要はないが、探究活動が進んでいる生徒に対しては整理整形を行うように指導を進めればよいと指導方法

ことが示唆された。一方、「どちらともいえない」の回答が含まれている項目が項目 1-6, 2-3, 2-6, 2-7であった。可読性の観点では、Analysis 段階の生徒の理解状態の分類、適切性の観点では、Problem, Analysis, Conclusion に含まれる状態の定義について検討の必要性が示唆された。

大学教員の回答結果では、可読性・適切性における Problem, Data フェーズは肯定的な回答が 100%であるが、Plan, Analysis, Conclusion フェーズは「どちらともいえない」と「あまり」が含まれていた。

5 考察

専門家によるアンケート結果に基づいて、構築したオントロジーの評価と改善点について述べる。

表 2: アンケート調査結果

項目	全回答								高校教員								大学教員							
	1	2	3	4	5	平均	SD	1	2	3	4	5	平均	SD	1	2	3	4	5	平均	SD			
1-1	0	0	0	1	8	4.9	0.3	0	0	0	1	4	4.8	0.4	0	0	0	0	4	5.0	0.0			
1-2	0	0	0	1	8	4.9	0.3	0	0	0	0	5	5.0	0.0	0	0	0	1	3	4.8	0.4			
1-3	0	0	0	1	8	4.9	0.3	0	0	0	1	4	4.8	0.4	0	0	0	0	4	5.0	0.0			
1-4	0	0	1	2	6	4.6	0.7	0	0	0	1	4	4.8	0.4	0	0	1	1	2	4.3	0.8			
1-5	0	0	0	3	6	4.7	0.5	0	0	0	2	3	4.6	0.5	0	0	0	1	3	4.8	0.4			
1-6	0	0	2	1	6	4.4	0.8	0	0	1	0	4	4.6	0.8	0	0	1	1	2	4.3	0.8			
1-7	0	0	1	1	7	4.7	0.7	0	0	0	0	5	5.0	0.0	0	0	1	1	2	4.3	0.8			
可読性	0	0	4	10	49			0	0	1	5	29			0	0	3	5	20					
2-1	0	0	0	3	6	4.7	0.5	0	0	0	2	3	4.6	0.5	0	0	0	1	3	4.8	0.4			
2-2	0	0	1	4	4	4.3	0.7	0	0	0	4	1	4.2	0.4	0	0	1	0	3	4.5	0.9			
2-3	0	0	2	2	5	4.3	0.8	0	0	2	1	2	4.0	0.9	0	0	0	1	3	4.8	0.4			
2-4	0	0	1	1	7	4.7	0.7	0	0	0	1	4	4.8	0.4	0	0	1	0	3	4.5	0.9			
2-5	0	0	0	5	4	4.4	0.5	0	0	0	2	3	4.6	0.5	0	0	0	3	1	4.3	0.4			
2-6	0	1	1	4	3	4.0	0.9	0	0	1	2	2	4.2	0.7	0	1	0	2	1	3.8	1.1			
2-7	0	0	2	0	7	4.6	0.8	0	0	1	0	4	4.6	0.8	0	0	1	0	3	4.5	0.9			
適切性	0	1	7	19	36			0	0	4	12	19			0	1	3	7	17					
3-1	0	1	0	0	8	4.7	0.9	0	0	0	0	5	5.0	0.0	0	1	0	0	3	4.3	1.3			
3-2	0	1	0	1	7	4.6	1.0	0	0	0	0	5	5.0	0.0	0	1	0	1	2	4.0	1.2			
3-3	0	0	0	3	6	4.7	0.5	0	0	0	1	4	4.8	0.4	0	0	0	2	2	4.5	0.5			
3-4	0	0	0	0	9	5.0	0.0	0	0	0	0	5	5.0	0.0	0	0	0	0	4	5.0	0.0			
3-5	0	1	0	4	4	4.2	0.9	0	0	0	2	3	4.6	0.5	0	1	0	2	1	3.8	1.1			
3-6	0	1	1	2	5	4.2	1.0	0	0	0	2	3	4.6	0.5	0	1	1	0	2	3.8	1.3			
3-7	0	1	0	2	6	4.4	1.0	0	0	0	1	4	4.8	0.4	0	1	0	1	2	4.0	1.2			
有用性	0	5	1	12	45			0	0	0	6	29			0	5	1	6	16					

が明確化されると高評価を得た。このことから、質の高さの設定によって、学校の実態に応じて生徒の理解度に合わせた指導計画を設定することに有効であることを示したと考えられる。

構築したオントロジーは、探究活動の遷移状態がオントロジーによる体系化によって明確に表現されており、暗黙的な指導手順を明示していると知識の体系化に関する肯定的な意見があった。PPDAC サイクルオントロジーの有効性では、手続き的知識が構造化されたことにより、PPDAC サイクルの認識を高めることに対する良好な評価が得られた。大学教員からは探究活動の指導の全体像が表現されており、高校生段階では有効であると評価された。

一方、可読性と適切性に関して、Plan, Analysis は一定程度の評価は得ているが、十分とはいえず、検討する必要がある。Plan フェーズでは環境的要因が抽象的であること、Analysis フェーズでは他のフェーズと異なった状態遷移ではない表記となっていることが指摘された。したがって、知識の体系化による改善が必要であるが、本オントロジーによって、探究学習に関する知識が獲得されたこと、探究活動の状態が体系化されたこと、PPDAC サイクルに基づいて構造化されたことは良好に評価されていると判断した。

5.2 オントロジーの改善点

探究活動を進める上で、問題発見から問題解決まで進めていくための知識の獲得に不足はないかどうか、知

識の体系化において上位と下位の関係が逆ではないかという点について検討した。

可読性において、Plan, Analysis, Conclusion に肯定的ではない回答があった。Plan, Conclusion については、ラベルの用語が難しいという指摘があった。これは再度検討をするが、知識の獲得や体系化に関する問題点ではなかった。Analysis については、知識の不足の指摘があり、今後追加修正を行う。

適切性において、Data 以外のフェーズに肯定的ではない回答があった。Problem では現実状態調査について具体的な知識の体系化を加えること、Plan では環境的要因が抽象的であること、Analysis では分析状態のよりよくなる特殊化を追加すること、Conclusion では未考察状態のラベルの記載が伝わりづらいことという指摘であった。したがって、Problem, Analysis は知識獲得不足、Plan はプロパティ不足が指摘され、追加修正を行う。また、Conclusion についてはラベルについて再度検討を行う。

有効性において、大学教員から大学教育における探究活動、つまり研究指導では Problem や Analysis に関する手続き的知識は一方通行ではなく、試行錯誤を行いながら実行していくため、その表現が不十分であるという指摘であった。ただし、高等学校教育の探究活動を大学教育の研究活動の導入と捉えるならば、高等学校教育では十分に役に立つオントロジーであるという意見であった。このことから、大学での探究学習の指導という視点では有効性に肯定的ではないが、高等学校教育に限定すれば有効性は見られるという回答で

あった。

したがって、それぞれの懸念点はあるが、回答した要因を調査すると、本オントロジーが不適切と考えられる要因ではなく、概ね影響はないと判断し、今後、適切な修正を行う。

6 おわりに

本研究では、問題解決プロセスを用いた探究活動の指導方法に関するPPDACサイクルオントロジーを構築した。標準化ルーブリックに基づき、問題解決プロセスに関する知識を獲得し、体系化し、PPDACサイクルに対応した手続き的知識を構造的に記述した。構築したPPDACサイクルオントロジーは一定の評価を受けたことを示した。このことから、以下の4点のことを明らかにした。

- 問題解決型探究指導における指導に関する知識を獲得したこと。
- 知識の体系化によって、暗黙的であった指導方法が明示化されたこと。
- 本オントロジーによって、指導経験の浅い教員が指導計画を立てるサポートとなる可能性を示唆したこと。
- Problem フェーズの指導には多くの時間を要することが示唆されたこと。

今後の課題としては、専門家によるアンケート調査結果をもとにオントロジーを再構築すること、その他の探究学習の指導用の資料にどれだけ対応しているのか妥当性を検証することが挙げられる。

謝辞

本研究では、独立行政法人新エネルギー・産業技術総合開発機構(NEDO) JPNP18002の助成を受けた。

参考文献

- [1] 国立教育政策研究所:OECD 生徒の学習到達度調査 PISA2012 年問題解決能力調査-国際結果の概要-, 2014.
- [2] 文部科学省: 高等学校学習指導要領(平成30年告示)解説総合的な探究の時間編, 2018.
- [3] 文部科学省: 高等学校学習指導要領(平成30年告示)解説数学編理数編, 2018.
- [4] 文部科学省: 高等学校情報科「情報I」教員研修用教材第1章情報社会の問題解決, 2019.
- [5] 稲永由紀:「総合的な探究の時間」の指導を支える教員の学術経験-学士課程教育をめぐる状況と教員養成上の課題-, RcusWorkingPaper, No20, p. 1-9, 2020.
- [6] 文部科学省: 令和元年度学校教員統計調査, 2021.
- [7] 溝口理一郎: オントロジー研究の基礎と応用, 人工知能学会誌, Vol14, No6, p.978-988, 1999.
- [8] 溝口理一郎: オントロジー構築入門, オーム社, 2006.
- [9] 古崎晃司: ドメインオントロジーの構築と利用, 情報知識学会誌, Vol19, No4, p.296-305, 2009.
- [10] 武内雅宇, 林雄介, 池田満, 溝口理一郎: 実践・教育複合型協調学習場の設計支援に向けたオントロジー工学的アプローチ, 人工知能学会論文誌, 21 巻, 2 号 F, p.184-194, 2006.
- [11] Yusuke Hayashi, Seiji Isotani, Jacqueline Bourdeau, Riichiro Mizoguchi: Toward a Learning/Instruction Process Model for Facilitating the Instructional Design Cycle, Education and Technology for a Better World, p.138-147, 2009.
- [12] 笠井俊信, 山口晴久, 永野和男: オントロジー理論に基づく情報教育目標の体系的記述, 電子情報通信学会論文誌, 88 巻, 1 号, p.3-15, 2005.
- [13] 笠井俊信, 永野和男, 溝口理一郎: 教師の授業設計意図自動外化システムの開発とその有効性評価, 人工知能学会論文誌, 30 巻, 3 号 SP-E, p.570-584, 2015.
- [14] 西村 悟史, 大谷 博, 島山 直人, 長谷部 希恵子, 福田賢一郎, 来村 徳信, 溝口 理一郎, 西村 拓一: 現場主体の“知識発現”方法の提案, 人工知能学会論文誌, 32 巻, 4 号, p.1-15, 2017.
- [15] Satoshi, Nishimura and Yoshinobu, Kitamura and Munehiko, Sasajima and Akiko, Williamson and Chikako, Kinoshita and Akemi, Hirao and Kanetoshi, Hattori and Riichiro, Mizoguchi: CHARM as Activity Model to Share Knowledge and Transmit Procedural Knowledge and its Application to Nursing Guidelines Integration, Journal of Advanced Computational Intelligence and Intelligent Informatics, 17 巻, 2 号, p.208-220, 2013.
- [16] 西村 悟史, 笹嶋 宗彦, 来村 徳信, 中村 明美, 高橋 弘枝, 平尾 明美, 服部 兼敏, 溝口 理一郎: 目的指向の看護手順学習に向けた複数観点からの知識閲覧システム CHARM Pad と新人看護師研修への実践的活用, 人工知能学会論文誌, 30 巻, 1 号, p.22-36, 2015.
- [17] 飯野 なみ, 西村 悟史, 西村 拓一, 福田 賢一郎, 武田 英明: 領域オントロジーと手続き的知識を併用した知識構築の実践と分析, 人工知能学会論文誌, 35 巻, 1 号, p.1-12, 2020.
- [18] 菊野 慎太郎, 松元 新一郎: 中学校数学科における「データの活用」の指導 統計的な問題解決のサイクル (PPDAC) に基づいた単元の開発と実践, 静岡大学教育実践総合センター紀要, 30 巻, p.69-78, 2020.
- [19] 林 宏樹, 渡辺 博芳: 高等学校「情報I」における PPDAC サイクルの段階的指導法を用いたデータサイエンス教育の授業実践例, 情報処理学会研究報告コンピュータと教育 (CE), CE-167, 1 号, p.1-7, 2022.
- [20] 青山 和裕: ニュージーランドの統計教育から得られる示唆, 日本数学教育学会誌, 100 巻, 7 号, p.11-20, 2018.
- [21] 深澤 弘美, 櫻井 尚子, 和泉 志津恵: 統計的探究プロセスとその評価, 統計数理, 66 巻, 1 号, p.15-36, 2018.
- [22] 大貫 守: 高等学校での課題研究ルーブリック作成の取り組み—科学的探究の指導と評価を中心に—, 思考力・判断力・表現力育成のための長期的ルーブリックの開発 最終成果報告書(研究代表: 田中耕治), p.71-83, 2016.
- [23] 西岡 加名恵, 大貫守: スーパーサイエンスハイスクール8校の連携による「標準ルーブリック」開発の試み, 教育方法の探究, 京都大学大学院教育学研究科・教育方法学研究室, 23 巻, p.1-12, 2020.
- [24] 古崎 晃司, 来村 徳信, 佐野 年伸, 本松 慎一郎, 石川 誠一, 溝口 理一郎: オントロジー構築・利用環境「法造」の開発と利用, 人工知能学会論文誌, 17 巻, 4 号, p.407-419, 2002.

クラック分類タスクにおける LRP for Branch Networks を用いた視覚的説明生成

Visual Explanation Generation for Road Damage Classification by Using Layer-wise Relevance Propagation for Branch Networks

飯田 紡^{1*} 小槻 誠太郎¹ 平川 翼² 山下 隆義² 藤吉 弘亘² 杉浦 孔明¹
Tsumugi Iida¹, Seitaro Otsuki¹, Tsubasa Hirakawa²,
Takayoshi Yamashita², Hironobu Fujiyoshi², Komei Sugiura¹

¹ 慶應義塾大学
¹ Keio University
² 中部大学
² Chubu University

Abstract: 深層学習が幅広い分野に応用されている現代において、深層学習モデルの説明性は重要であるが、説明生成のためのモジュールを利用する場合、それ自体が複雑になってしまい透明性が低い。逆伝播により説明を生成する手法は透明性が高いものの、cyclic connection を持たないモデルのみに適用されている。そのため、ブランチ構造を持つモデルにおいては、複数層の寄与度が重複して反映されてしまう。そこで本論文では、ブランチ構造を持つモデルにおける逆伝播の計算方法を新たに提案する。そして、ブランチ構造を持つモデルに、逆伝播による説明生成手法を導入して拡張した、Layer-wise Relevance Propagation for Branch Networks (LRP-BN) を提案する。道路上のクラック有無を分類するモデルに対する視覚的説明を生成するタスクに焦点をあて、LRP-BN により理論的背景が明確で高品質な説明を生成する。実験の結果、提案手法は視覚的説明生成タスクにおける標準的な評価尺度である Insertion-Deletion Score においてベースライン手法を上回り、適切な視覚的説明の生成に成功することが示された。

1 はじめに

深層学習が幅広い分野に応用されている現代において、深層学習モデルの説明性は重要である [Shrikumar 17, Ribeiro 16]。例えば、理論が未解明な自然現象の予測に深層学習を用いた場合、視覚的説明による重要な部分の可視化を通して、理論の洞察を与えることができる。また、複雑な深層学習モデルにおいては、判断根拠を説明することが困難であり、誤った根拠をもとに分類しているかどうかを見分けることが難しい。この場合、クレバーハンス効果 [Pfungst 07] のように、モデルが本質的な特徴ではなく、無関係な特徴に基づいて分類を行い、汎化性能の低下をもたらす可能性がある。そのため、深層学習モデルの説明性を向上させることは有益である。

本論文では、モデルが分類結果を出力する過程に対する判断根拠の視覚的説明生成タスクを扱う。特に、道路

上のクラック有無を分類するモデルに対して視覚的説明を生成するタスクに焦点をあてる。この視覚的説明は、道路上のクラックに対するマスクと考えることもできる。この場合、セグメンテーションのマスクが ground truth として与えられず、クラック有無のラベルアノテーションのみを用いてマスクを生成するため、本タスクは image-level weakly supervised semantic segmentation タスクとみなすことができる。

視覚的説明生成タスクは各モデルにおいて、本質的に重要な領域を正確に抽出する必要がある困難なタスクである。実際、人間が作成した道路上のクラックのマスクと、標準的な説明生成手法である GradCAM [Selvaraju 17] が生成した説明との IoU は 0.16 程度しか達成できていない。また、本タスクには明確な正解が存在することが少ないうえ、モデルの特徴や構造によって適切な説明生成手法は異なる。そのため、本タスクは正解マスクを利用せずに過不足なく適切な領域に注目する必要のある、難しいタスクである。図 1 に Road Damage Detection Dataset [Arya 22] の画像例を示す。まず、本

*連絡先：慶應義塾大学理工学部情報工学科
〒 223-8522 神奈川県横浜市港北区日吉 3-14-1
E-mail: tiida@keio.jp

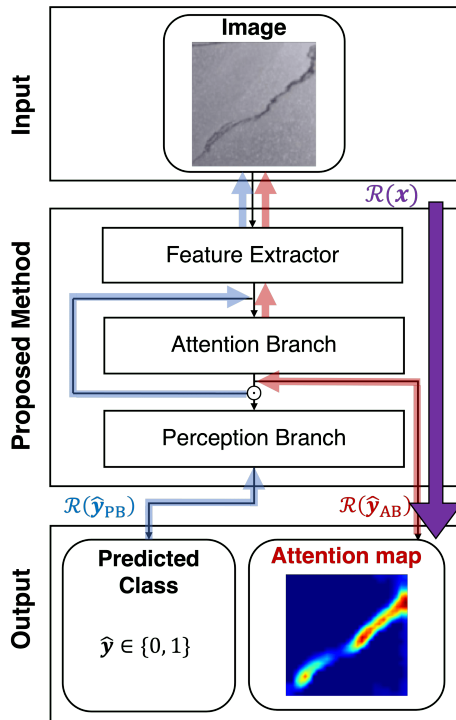


図 1: 提案手法の概略図

タスクにおいては図1中のInputに示すような道路画像をモデルに入力してクラックの有無を分類する。その過程で、図1中のAttention mapのような道路上のクラック領域に注目した判断根拠の視覚的説明を生成することが望ましい。

畳み込みニューラルネットワークを基盤とするモデルにおいて、視覚的説明の生成に関する研究は数多く提案されている [Selvaraju 17, Petsiuk 18, Zhang 21a]。これらの手法は、既定の計算方法により説明を生成する。このような手法はモデルの構造に依存しないが、複雑なモデル構造に特化した説明の生成が難しく、不適切な領域に注目する場合がある。また、説明生成のための専用モジュールをブランチとして組み込んだ手法として、Attention Branch Network (ABN) [Fukui 19] や Lambda Attention Branch Networks [Iida 22] などが存在する。しかし、説明生成専用のモジュール自体がブラックボックスになってしまい、透明性が低い。Layer-wise Relevance Propagation (LRP) [Bach 15] は、出力からの逆伝播を利用して説明を生成する手法である。各層での逆伝播の計算方法が定義されているため、拡張性が高い。実際、LSTM や Transformer に対する逆伝播の計算方法も提案されている [Arras 17, Ali 22a]。しかし、これらの方法は cyclic connection を持たないモデルのみに適用されているため、ブランチ構造や skip connection などの cyclic connection を持つモデルへの適用は、新たな計算方法の定義が必要である。

このような背景から、本研究では、ブランチ構造を持つモデルにおいて標準的な説明生成手法である ABN

に、説明生成の理論や計算過程が明瞭であり、高い透明性を有する LRP を導入して拡張する。これにより、理論的背景が明確で高品質な説明を生成する。既存研究との違いは、skip connection やブランチ構造を持つモデルにおける LRP の計算方法を新たに提案し、最も注目すべき領域を選択することで説明の品質を向上させる Choice 1 Component (C1C) を導入した点である。ブランチ構造や skip connection に対応した計算方法により、cyclic connection において、複数層の寄与度が重複して反映されてしまうことを防ぎ、適切な説明を生成することができる。また、最も注目度が高い画素を含む領域は、背景などの不適切な領域と連結していないことが多い。そのため、C1C により非連結な領域を除くことで背景を除去することができる。

本研究の独自性は以下の通りである。

- ブランチ構造や skip connection を持つモデルにおける LRP の計算方法を提案する。
- 生成した注目領域を元に、最も注目すべき領域を選択することで説明の品質を向上させる C1C を導入する。

2 関連研究

深層学習モデルの視覚的説明生成に関する研究は広く行われている [Bach 15, Selvaraju 17, Fukui 19, Ali 22a]。先行研究 [Das 20, Zhang 21b, Joshi 21, Ding 22] は、視覚的説明生成を含む深層学習モデルの説明生成に関して、包括的に調査し説明の生成方法ごとに各手法の分類・比較を行っている。視覚的説明生成タスクにおける標準的なデータセットとしては、ImageNet [Deng 09], CIFAR10, CIFAR100 等の標準的な画像分類データセットが使用されている。

視覚的説明生成の手法は、その生成方法によって Back Propagation (BP), Perturbation (PER) とその他に分類することができる。BP は逆伝播時の勾配に着目して説明を生成する。BP の手法として、LRP [Bach 15, Binder 16], Grad-CAM [Selvaraju 17], Integrated Gradients [Sundararajan 17], [Chefer 21] 等がある。[Sundararajan 17] は、感度と実装不変の2つの公理を満たすように、勾配を積分して説明を生成する手法である。[Ismail 21] は、重要でない領域の勾配をゼロに近づけることでノイズを減らす Saliency Guided Training を考案した。[Bach 15] は、出力からの逆伝播を利用して説明を生成する手法である LRP の基盤となる計算方法を定義した。また、LSTM や Transformer に対する逆伝播の計算方法も提案されている [Arras 17, Ali 22a]。

PER は入力に摂動を加えて、モデルの出力の変化から説明を生成する手法を指す。PER に分類される手法として、LIME [Ribeiro 16], Shapely Sampling [Lundberg 17], RISE [Petsiuk 18] 等がある。例えば、[Petsiuk 18]

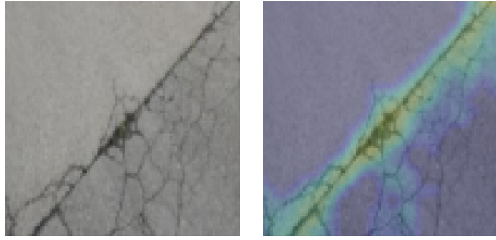


図 2: 視覚的説明生成タスクの例

は、マスクされた画像と出力の関係から説明を生成する手法である。

また、勾配や摂動以外から説明を生成する手法として ABN [Fukui 19], IA-CNN [Zhang 21c], IA-RED² [Pan 21] 等がある。ABN は、ブランチ構造として説明生成専用のモジュールを導入して説明を生成する拡張性が高い手法であり、Mask A3C [Itaya 21], PonNet [Magassouba 21], LABN [Iida 22] 等に応用されている。

サーベイ論文 [Cao 20, Ali 22b] は深層学習を用いた道路上のクラック検出タスクにおける各手法、標準データセット、標準評価尺度を包括的に紹介している。道路上のクラック検出には、Faster-RCNN [Ren 15] や SSD [Liu 16] 等の多くの物体検出モデルが応用されてきた [Yang 20, Yan 21]。[Yang 20] は、SSD に複数のカーネルサイズを持つ畳み込み層を含む Receptive Field を導入し、道路上のクラック検出に応用している。[Yan 21] は Deformable Convolution [Dai 17] を用いてクラックに沿った特徴抽出を行う Deformable SSD を提案している。道路上のクラック検出タスクにおける標準的なデータセットとしては RDD2022 Dataset [Arya 22] や Crack500 dataset [Yang 19] があげられる。

提案手法は説明生成専用のモジュール自体がブラックボックスである ABN とは異なり、ブランチ構造を持つモデルに透明性の高い LRP を導入する。また、cyclic connection に対応していない LRP とは異なり、skip connection やブランチ構造を持つモデルにおける LRP の計算方法を新たに提案する。

3 問題設定

本論文では、道路上のクラック有無分類タスクに対する判断根拠の視覚的説明生成を扱う。図 2 に道路上のクラック有無分類問題の例を示す。左図が入力であり、右図はモデルの注目領域を入力画像に重畳した画像である。本タスクでは、モデルの予測に貢献した画素に注目した視覚的説明が望ましい。

本論文では、画像から道路上のクラックを検出できることを前提とする。標準的な道路上のクラック検出手法は [Arya 22] にあげられている。本論文における用語を以下のように定義する：

- **クラック領域:** 画像における道路上のクラックを示す領域

本タスクの入力と出力はそれぞれ画像 $\mathbf{x} \in \mathbb{R}^{c \times h \times w}$ と \mathbf{x} がどのクラスに属するかの確率の予測値 $p(\hat{\mathbf{y}}) \in \mathbb{R}^C$ である。ここで、 C, c, h, w はそれぞれクラス数、入力画像におけるチャンネル数、縦幅、横幅を表す。また、視覚的説明として画像中の各画素に重要度を割り当てた attention map $\alpha \in \mathbb{R}^{h \times w}$ を利用する。

4 提案手法

提案手法はブランチ構造を持つ ABN [Fukui 19] に LRP [Bach 15] を導入して拡張した LRP for Branch Networks (LRP-BN) である。本論文においては、ABN をはじめとするブランチ構造を持つモデルに適用可能な LRP を扱う。本手法で行う拡張は、ブランチ構造を持つモデルにおける LRP の計算方法を定義したものである。そのため、ブランチ構造や cyclic connection をもつ手法一般に適用可能である。提案手法の新規性は以下の通りである。

- ブランチ構造や skip connection を持つモデルにおける LRP の計算方法を提案する。
- 生成した注目領域を元に、最も注目すべき領域を選択することで説明の品質を向上させる C1C を導入する。

4.1 モデル構造

図 3 に提案手法のモデル構造および入力における Relevance \mathcal{R} の計算方法の概略を示す。提案手法は、Feature Extractor (FE), Attention Branch (AB), Perception Branch (PB) の 3 モジュールから構成される。

FE は、モデルの注目領域の生成および予測に用いる特徴抽出をするためのモジュールで、畳み込み層、Batch Normalization 層、Max Pooling 層と B 個の Bottleneck 層から構成される。 f_{FE} の入力は \mathbf{x} で、画像特徴量 $\mathbf{h} \in \mathbb{R}^{c_1 \times h_1 \times w_1}$ を出力する。ここで、 c_1, h_1, w_1 はそれぞれ画像特徴量のチャンネル数、縦幅、横幅を表す。

AB は説明生成のための $f_{AB}^{(1)}$ と、説明と分類を関連づけるための $f_{AB}^{(2)}$ に分かれる。 $f_{AB}^{(1)}$ は Bottleneck 層、畳み込み層、Batch Normalization 層、Max Pooling 層から構成される。 $f_{AB}^{(1)}$ の入力は \mathbf{h} であり、出力は $\tilde{\alpha} \in \mathbb{R}^{w_1 \times h_1}$ である。また、予測に重要でない領域を削除して PB に入力するために、 $\tilde{\alpha}$ のうち、ハイパーパラメータ θ_α より小さな値を 0 とし $\alpha' \in \mathbb{R}^{w_1 \times h_1}$ とする。 $f_{AB}^{(2)}$ の入力は \mathbf{h} であり、出力は attention loss を計算するための確率の予測値 $p(\hat{\mathbf{y}}_{AB})$ である。 $f_{AB}^{(2)}$ は Bottleneck 層、畳み込み層、Batch Normalization 層、Max Pooling 層、Global Average Pooling 層から構成される。損失関数に $p(\hat{\mathbf{y}}_{AB})$ を加えることで、AB を分類に直接関連付けて学習させることができる。その結果、分類結果と関連する attention map を生成できる。

PB は \mathbf{h} と α の両方を用いて分類を行うモジュールである。PB は $N_B - B$ 個の Bottleneck 層と全結合層か

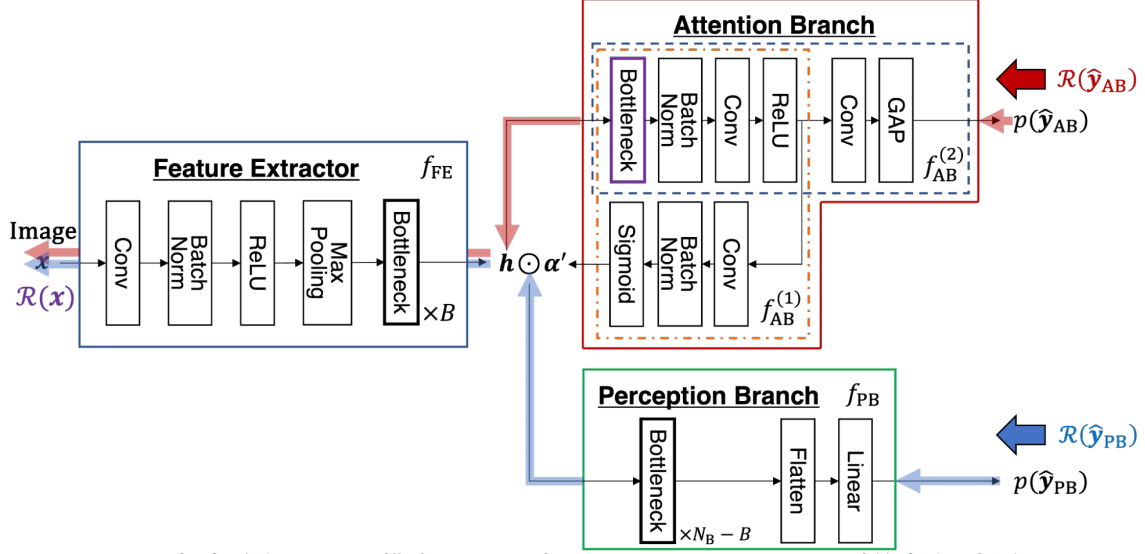


図 3: 提案手法のモデル構造および入力における Relevance \mathcal{R} の計算方法の概略図

ら構成される。ここで、 N_B はバックボーンネットワークの Bottleneck 層の数を表す。PB の入力は $\alpha' \odot h$ である。マスク処理をした α' と h を掛け合わせることで、予測に重要な領域を入力することができる。また、PB の出力はどのクラスに属するかの確率の予測値 $p(\hat{y}_{PB})$ である。

最終的なモデルの予測は以下の式で表される:

$$p(\hat{y}_{PB}) = f_{PB}(\alpha' \odot h) \quad (1)$$

$$p(\hat{y}_{AB}) = f_{AB}^{(2)}(h) \quad (2)$$

$p(\hat{y}_{PB})$ はどのクラスに属するかの確率の予測値であり、分類の予測結果を出力するために利用する。また、 $p(\hat{y}_{AB})$ は分類には直接用いないが、損失関数に導入することで説明の品質を向上させることができる。

4.2 LRP の計算方法

通常の構造、ブランチ構造、skip connection の 3 つの構造に分けて、提案手法における LRP の計算方法を説明する。

4.2.1 通常の構造における LRP の計算方法

通常の構造においては、標準的な LRP の z-rule [Bach 15, Binder 16] を適用して計算する。一例として、 $\mathcal{R}(z)$ を z の Relevance とすると、Linear 層における LRP の計算は以下で表される:

$$\mathcal{R}(z_i^{(1)}) = \sum_j \frac{\text{ReLU}(w_{ij} z_j^{(0)})}{\sum_k \text{ReLU}(w_{kj} z_j^{(0)})} \mathcal{R}(z_i^{(0)}) \quad (3)$$

ここで、 $z_i^{(1)}, z_i^{(0)}$ はそれぞれ Linear 層の入力、出力における i 番目の要素を、 w_{ij} は Linear 層の重みにおける (i, j) 要素を表す。しかし、上記の計算方法は cyclic connection を含まないモデルに対して提案されており、

ブランチ構造や skip connection などの cyclic connection をもつモデルには対応していない。そのため、本研究ではブランチ構造・skip connection における LRP の計算方法を提案する。

4.2.2 ブランチ構造における LRP の計算方法

本モデルにおけるブランチ構造では、 $p(\hat{y}_{AB})$ と $p(\hat{y}_{PB})$ それぞれから 2 つの Relevance $\mathcal{R}_{AB}, \mathcal{R}_{PB}$ が計算される。そのため、通常の構造とは異なる方法で計算する必要がある。

まず、 \mathcal{R}_{AB} の計算方法を考える。図 3 に示すように、 \mathcal{R}_{PB} の入力に用いる α' は $f_{AB}^{(1)}$ を通じて計算される。そのため、 \mathcal{R}_{AB} の計算方法として、 \mathcal{R}_{PB} を利用する方法と、 \mathcal{R}_{PB} とは独立に計算する方法が考えられる。ここで、 h の Relevance $\mathcal{R}(h)$ を計算する際、 \mathcal{R}_{AB} と \mathcal{R}_{PB} の両方を利用すると、前者では \mathcal{R}_{AB} を介して \mathcal{R}_{PB} の影響が二重に反映される可能性がある。そのため、図 3 に赤・青の矢印で示すように、 \mathcal{R}_{PB} とは独立に計算を行う。この計算方法において、ブランチ構造は α' を用いたゲート構造とみなすことができ、LSTM における LRP [Arras 17] において、ゲート構造を独立に計算する方法と一致する。

次に、 $\mathcal{R}(h)$ の計算方法に関しては、conservation [Bach 15] を考慮して $\mathcal{R}_{AB}, \mathcal{R}_{PB}$ の和を $\mathcal{R}(h)$ とする。AB の入力が h で、PB の入力が $\alpha' \odot h$ であるため、 $1 : \alpha'$ の重みを付けた和も考えられる。しかし、 α' の寄与は forward 計算時に既に含まれているため [Arras 17]、重複して寄与を考慮しないために \mathcal{R}_{AB} と \mathcal{R}_{PB} の和と定義した。以上より、 $\mathcal{R}(h)$ は以下の式で表される:

$$\mathcal{R}(h) = \mathcal{R}_{AB} + \mathcal{R}_{PB} \quad (4)$$

4.2.3 Skip connection における LRP の計算方法

Residual connection [He 15] はサイクル構造を持つが、z-rule により計算すると skip connection の影響が

考慮されない。また、residual block と skip connection の和を出力とする点で、並列に計算したアダマール積を利用する ABN のブランチ構造とは異なる。そのため、skip connection を考慮した LRP を提案する。

まず、residual connection の入出力をそれぞれ \mathbf{x}_s , \mathbf{y}_s と表し、residual block に z-rule を適用して計算した Relevance, 出力の Relevance をそれぞれ $\mathcal{R}(\mathbf{x})^-$, $\mathcal{R}(\mathbf{y}_s)$ と表す。 $\mathcal{R}(\mathbf{h})$ の議論と同様に、conservation を考慮すると \mathbf{x}_s の Relevance $\mathcal{R}(\mathbf{x}_s)$ は $\mathcal{R}(\mathbf{x})^-$ と $\mathcal{R}(\mathbf{y}_s)$ の加重和で表すことができると考えられる:

$$\mathcal{R}(\mathbf{x}_s) = \gamma \mathcal{R}(\mathbf{x}_s)^- + (1 - \gamma) \mathcal{R}(\mathbf{y}_s) \quad (5)$$

ここで、 γ は $\mathcal{R}(\mathbf{x}_s)^-$ と $\mathcal{R}(\mathbf{y}_s)$ の比率である。 γ は \mathbf{x}_s と \mathbf{y}_s を考慮して決定することもできるが、事前実験の結果良好な結果が得られたため、 $\gamma = 0.5$ とした。

4.3 Relevance と attention map を用いた視覚的説明の計算方法

提案手法においては、 \mathcal{R} と α' を組み合わせ、C1C を導入することで高品質な説明を生成する。本手法で説明として使用する α の計算方法を以下で述べる。

$\mathcal{R}(\mathbf{h})$ から FE の入力に対する Relevance を計算することで \mathbf{x} に対する Relevance \mathcal{R} が得られる。既存の LRP と同様に、この \mathcal{R} を説明として使用することも可能である。また、既存の ABN と同様に α' も説明として使用できる。しかし、単一の説明生成手法を利用した場合、不適切な領域に注目した説明が生成されることがあり、その際に修正の余地がない。一方、本手法では、LRP と ABN の双方が強く注目した領域をより強調し、高品質な説明を得るために \mathcal{R} と α' のアダマール積を説明に利用する。続いて、背景等の不適切への注目を防ぐため、C1C により最も注目すべき領域を抽出して α_{C1C} を得る。C1C においては、 $\mathcal{R} \odot \alpha'$ を 28×28 に縮小して細かいノイズや不要な情報を削除した上で、注目度が高い画素を含む連結領域を抽出する。 \mathcal{R} は多くの場合クラックに最も注目しており、最も注目度が高い画素を含む領域は背景などの不適切な領域と連結していないことが多い。そのため、非連結な領域を除くことで背景を除去することができる。最後に、 α_{C1C} を $w \times h$ に拡大して α を得る。

また、損失関数として、以下を使用する:

$$\mathcal{L} = \text{CE}(\hat{\mathbf{y}}_{PB}, \mathbf{y}) + \lambda \text{CE}(\hat{\mathbf{y}}_{AB}, \mathbf{y}) \quad (6)$$

ここで、 \mathbf{y} , CE, λ はそれぞれ正解ラベル, 交差エントロピー誤差関数, 損失関数の重みを表す。

5 実験

5.1 データセットと実験設定

本研究で扱う視覚的説明生成タスクのための標準データセットは我々の知る限り存在しない。視覚的説明生成

表 1: 実験で用いた設定

エポック	300	
バッチサイズ	64	
学習率	Feature Extractor	1.0×10^{-4}
	Linear	1.0×10^{-4}
	Attention Branch	1.0×10^{-3}
最適化	AdamW	

タスクにおいては、教師なしセグメンテーションタスクへの応用が可能で、データ数が十分であることが望ましい。そのため、セグメンテーションのマスクを含まず、人間によるマスク作成が可能でデータ数が十分な Road Damage Detection 2022 Dataset (RDD2022 Dataset) が最も適している。よって、RDD2022 Dataset の訓練集合から、画像選択・画像のクロップ・テスト集合作成の三段階の処理によって RDC Dataset を構築した。

RDC Dataset には、道路画像および、道路上のクラック有無が付与されたラベルが含まれている。RDD2022 Dataset には、日本・インド・チェコ・ノルウェー・アメリカ・中国の合計 6 カ国の道路画像が含まれる。このうち、中国以外の 5 カ国は車内から、中国はドローンやバイクから撮影された画像である。そのため、RDC Dataset においては画角の異なる中国を除いた 5 カ国を採用した。RDD2022 Dataset には、著者らが LabelImg と Computer Vision Annotation Tool によりアノテーションを行ったクラック領域情報が含まれている [Arya 20, Arya 22]。このアノテーションデータに含まれるクラック領域に基づき画像を切り出し、クラック有クラスの画像を作成した。また、クラック無クラスの画像はクラック領域を除いた領域からランダムに切り出して作成した。この際、クラック無クラスの切り出し後の縦・横幅はそれぞれアノテーションデータから計算したクラック領域の縦・横幅分布から選択した。最後に、縮小・標準化・二値化を行った後に画像間の XOR を計算することで類似画像を抽出し、類似画像のない画像群を選択してテスト集合を作成した。これらの処理により RDC Dataset を構築した。

RDC Dataset は日本・インド・チェコ・ノルウェー・アメリカの 5 カ国で撮影された道路画像を含む。クラック有クラスの画像は 47,513 枚、クラック無クラスの画像は 30,430 枚であった。訓練集合、検証集合、テスト集合はそれぞれ 66,641, 7,405, 3,897 サンプルを含む。本研究では、 \mathbf{x} を 224×224 にリサイズして、反転・回転・切り抜きによるデータ拡張を行った。訓練集合はモデルの学習に、検証集合はハイパーパラメータを調整するために使用した。また、テスト集合はモデルの性能評価に使用した。

表 1 に提案手法における設定を示す。提案手法のパラメータ数と積和演算数はそれぞれ 3200 万、92.1G で

表 2: 各手法における定量的結果

Method	Acc \uparrow	Insertion \uparrow	Deletion \downarrow	ID Score \uparrow
RISE [Petsiuk 18]	0.958 \pm 0.004	0.373 \pm 0.042	0.054 \pm 0.027	0.319 \pm 0.018
GradCAM [Selvaraju 17]	0.958 \pm 0.004	0.635 \pm 0.026	0.052 \pm 0.011	0.583 \pm 0.020
LRP [Bach 15]	0.958 \pm 0.004	0.528 \pm 0.117	0.301 \pm 0.111	0.227 \pm 0.010
ABN [Fukui 19]	0.957 \pm 0.004	0.358 \pm 0.035	0.090 \pm 0.013	0.268 \pm 0.039
Ours	0.957 \pm 0.004	0.804 \pm 0.005	0.069 \pm 0.006	0.735 \pm 0.007

あった。訓練にはメモリ 11GB 搭載 GeForce RTX 2080 Ti, Intel Core i9 9900K および 64GB の RAM を用いて、モデルの訓練時間および 1 サンプルあたりの推論時間は、それぞれ 3 時間および 1.3×10^{-3} 秒であった。検証集合における損失関数の値が 4 回連続改善しなかった場合に早期終了を行った。このとき、検証集合における損失関数の値が最も低いときのテスト集合における精度を、最終的な精度とした。

5.2 実験結果

ベースライン手法として、RISE [Petsiuk 18], GradCAM [Selvaraju 17], LRP [Bach 15], ABN [Fukui 19] を使用した。ABN をベースライン手法とした理由は、バックボーンネットワークとして ResNet を用いており、ブランチ構造を有する最も標準的な手法のためである。同様に、RISE・GradCAM・LRP は汎用的なモデルに適用可能な手法の中で標準的であるためベースライン手法とした。

本実験における評価尺度には、Accuracy, Insertion Score, Deletion Score, Insertion-Deletion Score (ID Score) を用いた。また、最も標準的な ID Score を主要評価尺度とした。Accuracy は分類タスクにおけるモデルの標準的な評価尺度であり、Insertion score, Deletion score, ID score は説明生成タスクの標準的な評価尺度であるため使用した。

Insertion Score, Deletion Score は Insertion 曲線, Deletion 曲線の AUC で計算される。また、ID Score は Insertion Score と Deletion Score の差で定義される。ここで、Insertion 曲線, Deletion 曲線はそれぞれ α を基に重要な領域を挿入, 削除した際の予測の変化を表す。詳細は以下で定義する。

まず、 α の要素を降順に $\alpha_{i_1, j_1}, \alpha_{i_2, j_2}, \dots, \alpha_{i_n, j_n}$ とし、集合 $A_n, \mathbf{i}_n, \mathbf{d}_n$ を次のように定義する。

$$A_n = \{(i_k, j_k) | k \leq n\} \quad (7)$$

$$(\mathbf{i}_n, \mathbf{d}_n) = \begin{cases} (x_{ij}, 0), (i, j) \in A_n \\ (0, x_{ij}), (i, j) \notin A_n \end{cases} \quad (8)$$

ここで、 n は挿入・削除するピクセル数を表す。 $\mathbf{i}_n, \mathbf{d}_n$ をモデルに入力した際の出力をそれぞれ $\mathbf{y}^{(\text{ins}, n)}, \mathbf{y}^{(\text{del}, n)}$ とする。このとき、 $(n, \mathbf{y}_c^{(\text{ins}, n)})$, $(n, \mathbf{y}_c^{(\text{del}, n)})$ をプロッ

トした曲線が、Insertion 曲線, Deletion 曲線である。ここで、 C は \mathbf{x} が属するクラスを表す。

表 2 にベースライン手法と提案手法との比較に関する定量的結果を示す。各手法につき実験を 5 回行い、その平均値および標準偏差を示した。また、表 2 中の太字は、統計的に有意な最良値を表す。表 2 より、主要尺度である ID Score において、RISE, GradCAM, LRP, ABN, および提案手法はそれぞれ 0.319, 0.583, 0.227, 0.268 および 0.735 であり、提案手法はベースラインの中で最も高い GradCAM と比較して 0.152 ポイント上回った。また、Accuracy においては RISE, GradCAM, LRP が 0.958, ABN と提案手法が 0.957 で同程度であった。主要尺度である ID スコアと Insertion スコアにおける性能差は統計有意であった ($p < 0.05$)。

図 4 に定性的結果を示す。(a) 列は元画像を示し、(c)-(e) 列はベースライン手法, (f) 列は提案手法によって生成した説明を元画像に重畳した結果を表す。図 4 の 1-3 行目は説明生成に成功した例で、4 行目は説明生成に失敗した例である。図 4(b) 列より、RISE によって生成された説明は道路上のクラックの周辺に注目領域を有するが、クラック以外の領域にも強く注目していた。また、(b), (e) 列より GradCAM, ABN によって生成された説明が強く注目していたのはクラックのうちの一部であった。(d) 列より LRP によって生成された説明は画像中のわずかな画素のみ強く注目しており、ほとんどの領域の注目度が等しく不適切である。一方で、(f) 列より提案手法は道路上のクラック全体に詳細に注目しており、クラック以外の道路の注目度は低く、適切な説明を生成している。

図 4 の 4 行目に示した失敗例について、(c), (e), (f) 列より、GradCAM, ABN, 提案手法により生成した説明は全て画像中の右側のクラックに強く注目しており、左側にあるクラックに注目できていない、また、(d) 列より LRP によって生成した説明は画像左下のクラックがない領域にのみ注目している。(b) 列より、RISE によって生成した説明は画像全体を注目している。しかし、中央左の注目度が低い領域にもクラックがあるため、全てのクラックを適切に注目できていない。上記より、全ての手法が道路上のクラックを過不足なく注目できていない。これは、道路が整備されておらず、クラックの無い道路とクラックの境界が曖昧になって

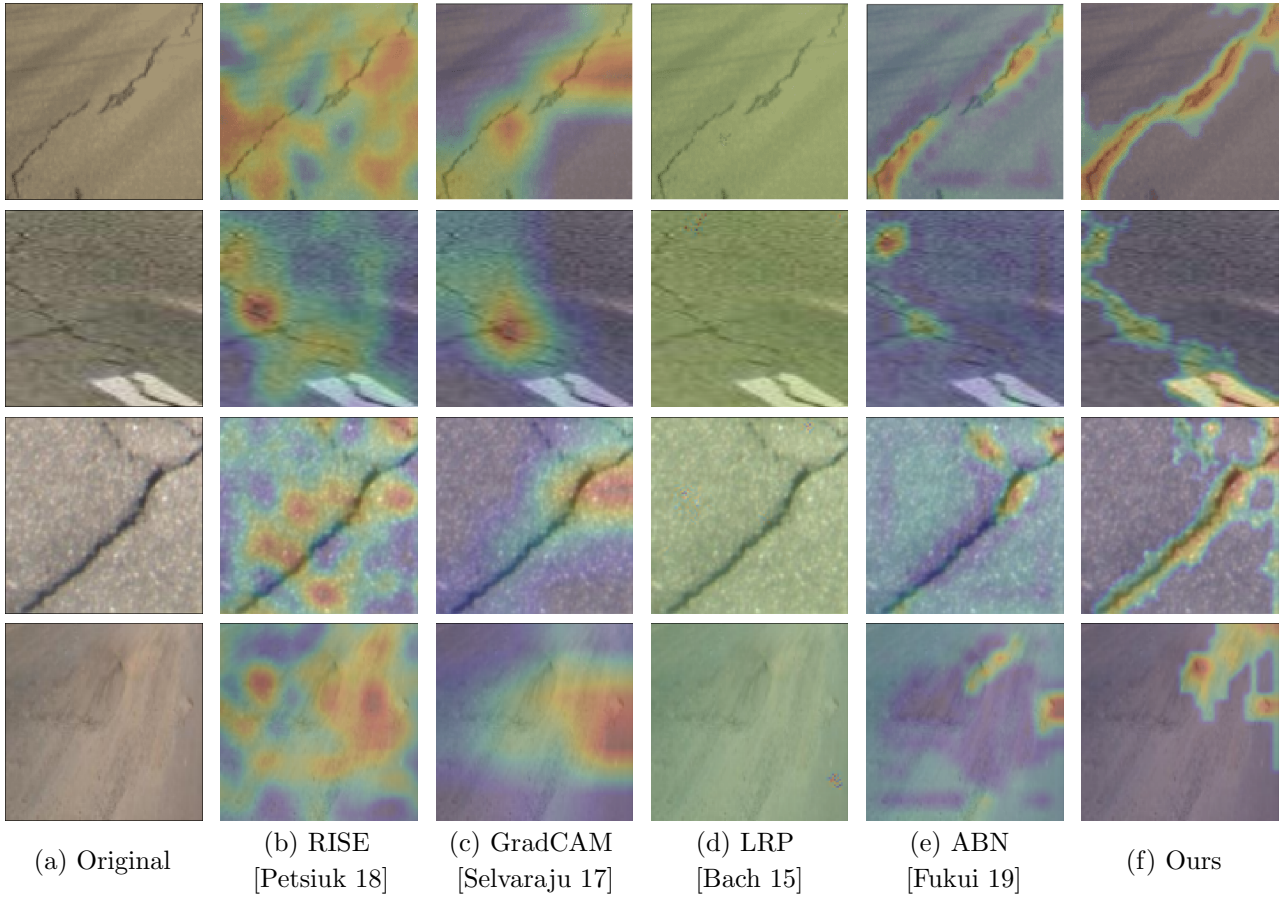


図 4: 各手法における定性的結果

表 3: 被験者が作成した正解マスクと各手法によって生成した説明との IoU 計測実験の結果

	RISE [Petsiuk 18]	GradCAM [Selvaraju 17]	LRP [Bach 15]	ABN [Fukui 19]	Ours
IoU ↑	0.167 ± 0.004	0.141 ± 0.002	0.111 ± 0.000	0.113 ± 0.107	0.184 ± 0.004

おり、判別が難しいことが原因だと考えられる。

最後に、被験者実験として、人間が作成したクラックのマスクと、ベースライン手法および提案手法が生成した説明の IoU を計測した。まず、被験者 4 人がそれぞれ異なる 50 サンプルについてクラック領域を示したマスクを作成し、合計 200 サンプルのマスクを得た。これを正解マスクとして、表 3 に、正解マスクと各手法によって生成した説明との IoU を示す。各手法につき実験を 5 回行い、その平均値および標準偏差を示した。また、表 3 の太字は最良値を表す。表 3 より、IoU において、RISE, GradCAM, LRP, ABN, および提案手法はそれぞれ 0.167, 0.141, 0.111, 0.113 および 0.184 であり、提案手法はベースライン手法の中で最も高い RISE と比較して 0.017 ポイント上回った。これらの結果より、提案手法が最も正解マスクと類似した説明を生成できていると示唆される。

6 おわりに

本論文では、道路上のクラック有無分類問題における判断根拠の視覚的説明生成を扱った。提案手法による貢献は以下である。

- Skip connection やブランチ構造を持つモデルにおける LRP の計算方法を提案した。
- 生成した注目領域を元に、最も注目すべき領域を選択することで説明の品質を向上させる C1C を導入した。
- 本タスクの標準的な評価尺度である Insertion Score, ID Score において、提案手法がベースライン手法を上回った。

謝辞

本研究の一部は、JSPS 科研費 23H03478, JST ムーンショット, NEDO の助成を受けて実施されたものである。

参考文献

- [Ali 22a] Ali, A., Schnake, T., Eberle, O., et al.: XAI for Transformers: Better Explanations through Conservative Propagation, in *ICML*, Vol. 162, pp. 435–451 (2022)
- [Ali 22b] Ali, L., Alnajjar, F., Khan, W., Serhani, M. A., et al.: Bibliometric Analysis and Review of Deep Learning-Based Crack Detection Literature Published between 2010 and 2022, *Buildings*, Vol. 12, No. 4 (2022)
- [Arras 17] Arras, L., Montavon, G., Müller, R., et al.: Explaining Recurrent Neural Network Predictions in Sentiment Analysis, in *WASSA*, pp. 159–168 (2017)
- [Arya 20] Arya, D., Maeda, H., Kumar, S., Toshniwal, D., et al.: Global Road Damage Detection: State-of-the-art Solutions, in *Big Data*, pp. 5533–5539 (2020)
- [Arya 22] Arya, D., Maeda, H., et al.: Crowdsensing-based Road Damage Detection Challenge (CRDDC ’2022), in *Big Data*, pp. 6378–6386 (2022)
- [Bach 15] Bach, S., Binder, A., Montavon, G., et al.: On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation, *PLOS ONE*, Vol. 10, No. 7, pp. 1–46 (2015)
- [Binder 16] Binder, A., et al.: Layer-Wise Relevance Propagation for Neural Networks with Local Renormalization Layers, in *ICANN*, pp. 63–71 (2016)
- [Cao 20] Cao, M.-T., Tran, Q.-V., Nguyen, N.-M., et al.: Survey on Performance of Deep Learning Models for Detecting Road Damages Using Multiple Dashcam Image Resources, *Adv. Eng. Inform.*, Vol. 46, p. 101182 (2020)
- [Chefer 21] Chefer, H., Gur, S., and Wolf, L.: Transformer Interpretability Beyond Attention Visualization, in *CVPR*, pp. 782–791 (2021)
- [Dai 17] Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., and Wei, Y.: Deformable Convolutional Networks, in *ICCV* (2017)
- [Das 20] Das, A. and Rad, P.: Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey, *arXiv preprint arXiv:2006.11371* (2020)
- [Deng 09] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database, in *CVPR*, pp. 248–255 (2009)
- [Ding 22] Ding, W., Abdel, M., Hawash, H., and Ali, A.: Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey, *Inf. Sci.*, Vol. 615, pp. 238–292 (2022)
- [Fukui 19] Fukui, H., Hirakawa, T., et al.: Attention Branch Network: Learning of Attention Mechanism for Visual Explanation, in *CVPR*, pp. 10705–10714 (2019)
- [He 15] He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, in *CVPR*, pp. 770–778 (2015)
- [Iida 22] Iida, T., Komatsu, T., Kaneda, K., et al.: Visual Explanation Generation Based on Lambda Attention Branch Networks, in *ACCV*, pp. 3536–3551 (2022)
- [Ismail 21] Ismail, A., Corrada, H., and Feizi, S.: Improving Deep Learning Interpretability by Saliency Guided Training, in *NeurIPS* (2021)
- [Itaya 21] Itaya, H., et al.: Visual Explanation using Attention Mechanism in Actor-Critic-based Deep Reinforcement Learning, in *IJCNN*, pp. 1–10 (2021)
- [Joshi 21] Joshi, G., Walambe, R., and Kotecha, K.: A Review on Explainability in Multimodal Deep Neural Nets, *IEEE Access*, Vol. 9, pp. 59800–59821 (2021)
- [Liu 16] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, Y., and Berg, A.: SSD: Single Shot Multi-Box Detector, in *ECCV*, pp. 21–37 (2016)
- [Lundberg 17] Lundberg, S. and Lee, I.: A Unified Approach to Interpreting Model Predictions, in *NeurIPS*, pp. 4765–4774 (2017)
- [Magassouba 21] Magassouba, A., Sugiura, K., et al.: Predicting and Attending to Damaging Collisions for Placing Everyday Objects in Photo-Realistic Simulations, *Advanced Robotics*, Vol. 35, No. 12, pp. 787–799 (2021)
- [Pan 21] Pan, B., Panda, R., Jiang, Y., et al.: IA-RED²: Interpretability-Aware Redundancy Reduction for Vision Transformers, in *NeurIPS* (2021)
- [Petsiuk 18] Petsiuk, V., Das, A., and Saenko, K.: RISE: Randomized Input Sampling for Explanation of Black-box Models, in *BMVC*, pp. 151–164 (2018)
- [Pfungst 07] Pfungst, O.: *Das Pferd des Herrn von Osten: der kluge Hans. Ein Beitrag zur experimentellen Tier- und Menschen-Psychologie*, Barth (1907)
- [Ren 15] Ren, S., He, K., Girshick, R., and Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, in *NIPS*, Vol. 28 (2015)
- [Ribeiro 16] Ribeiro, M., Singh, S., et al.: "Why Should I Trust You?": Explaining the Predictions of Any Classifier, in *KDD*, p. 1135–1144 (2016)
- [Selvaraju 17] Selvaraju, R., et al.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, in *ICCV*, pp. 618–626 (2017)
- [Shrikumar 17] Shrikumar, A., et al.: Learning Important Features Through Propagating Activation Differences, in *PMLR*, Vol. 70, pp. 3145–3153 (2017)
- [Sundararajan 17] Sundararajan, M., Taly, A., and Yan, Q.: Axiomatic Attribution for Deep Networks, in *ICML*, Vol. 70, p. 3319–3328 (2017)
- [Yan 21] Yan, K., et al.: Automated Asphalt Highway Pavement Crack Detection Based on Deformable Single Shot Multi-Box Detector Under a Complex Environment, *IEEE Access*, Vol. 9, pp. 150925–150938 (2021)
- [Yang 19] Yang, F., Zhang, L., Yu, S., Prokhorov, D. V., Mei, X., et al.: Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection, *IEEE trans Intell Transp Syst.*, Vol. 21, pp. 1525–1535 (2019)
- [Yang 20] Yang, J., Fu, Q., et al.: Road Crack Detection Using Deep Neural Network with Receptive Field Block, *Mater. Sci. Eng.*, Vol. 782, No. 4, p. 042033 (2020)
- [Zhang 21a] Zhang, Q., et al.: Group-CAM: Group Score-Weighted Visual Explanations for Deep Convolutional Networks, *arXiv preprint arXiv:2103.13859* (2021)
- [Zhang 21b] Zhang, Y., Tiño, P., Leonardi, A., and Tang, K.: A Survey on Neural Network Interpretability, *TETCI*, Vol. 5, No. 5, pp. 726–742 (2021)
- [Zhang 21c] Zhang, Z., Chen, Y., et al.: IA-CNN: A generalised interpretable convolutional neural network with attention mechanism, in *IJCNN*, pp. 1–8 (2021)

ChatGPT を用いた SNS ユーザのストレスを含むコメントの抽出とストレス推定

Extraction and Stress Estimation of Comments including Stress of SNS Users using ChatGPT

岡本 宗一郎^{1*} 砂山 渡² 服部 峻²
Soichiro Okamoto¹ Wataru Sunayama² Shun Hattori²

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: According to the Ministry of Health, Labour and Welfare's recent "White Paper on Measures to Prevent Deaths from Overwork, etc.," more than half of workers continue to experience strong anxiety, worry and stress related to their work and professional lives. There is a trend for people to post comments on social networking services on the Internet to gain sympathy and escape from reality in order to relieve the stresses that occur daily. Based on the above, it is important to estimate stress in order to detect persons in need of psychological support at an early stage. In this study, we aim to construct a system that can estimate the stress state of SNS users by extracting comments containing stress from SNS text information.

1 はじめに

昨今、健康食品やウェアラブル端末の普及により、人々の健康意識が高まっている。株式会社 JR 東海エージェンシーのビジネスパーソンの「健康意識」に関する調査 2017[1] では、20 歳以上のビジネスパーソン 1000 人のうち 82.0 %が健康を意識していると回答している。しかし健康意識が高まっている中、ストレスを感じている人も多くいる現状がある。そこで、精神的サポートが必要な人を早期に発見するために、ストレスを推定することは重要であると考えられる。また、総務省の『情報通信政策研究』に記載されている調査 [2] には、ネットアプリへの逃避の流れとして、日々生じるストレスによって共感による安心や現実的感覚の遮断が高まると考えられている。特に、共感による安心に着目すると、SNS 使用に関係するものが多いという調査結果が得られている。このことから、SNS のテキスト情報からストレス状態を推定できると考えている。既存研究では、テキストマイニングを用いて、ストレス分析を行う研究が行われている。しかし、これらの研究では、単語頻度や共起ネットワークの結果を用いて

ツール使用者が分析を行っている。

そこで本研究では、SNS のテキスト情報から、ストレスを含むコメントを抽出し、SNS ユーザのストレス状態を推定からストレス低減を目指す会話システムの構築を目的とする。POMS(気分プロフィール) というストレス尺度 [6] を利用し、ChatGPT によるストレス推定を行うことで、高い精度でのストレス分析が期待される。また、ChatGPT を用いた ChatBot との会話システムにより、ユーザのストレス低減が期待される。

2 関連研究

本章では、テキスト情報からストレスを分析することに関連する研究について述べる。

Twitter の投稿内容を収集し、KH Coder Ver.3 を用いて計量テキスト分析を行うことで新型コロナウイルス感染拡大に伴う不安やストレスの動向について調査する研究がある [3]。単語の関連性を可視化する共起ネットワークと対応分析を用いることで、SNS データを用いて人々の不安 やストレスの実態の推移を把握できる可能性を示している。しかし単語の頻度情報からは文脈情報を扱うことが難しい。そこで本研究では、ChatGPT を用いてストレス分析に特化したシステムを構築する

*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
岡本 宗一郎
〒 522-8533 滋賀県彦根市八坂町 2500
E-mail: ov23sokamoto@ec.usp.ac.jp

ことで、テキスト情報から自動的にストレス分析ができることを目指している。

他には、BERTを用いて一般的なツイートを含む集合からメンタルヘルスの状態の推測を行う研究がある [4]。SNS の投稿を収集しメンタルヘルス不調群と非メンタルヘルス不調者群に分けたデータセットを用いて BERT をファインチューニングすることで、SVC や LSTM よりも高い精度で分類できることを示している。しかし、メンタルに不調が有るか無いかの二分類である。本研究では、POMS というストレス尺度を用いて、項目ごとの分類によるストレス状態の推定を試みる。

3 ストレス反応とストレス推定システム

3.1 ストレス反応について

本節では、本研究で扱うストレス反応について定義する。文部科学省によると、ストレス反応は、長時間ストレスの刺激を受けた場合や、強いストレスを受けた時に生じる生体反応であり、ストレスに対する生体の自然な適応反応とされている [5]。また、ストレス反応は以下の三つに分類される。

- 「心理的ストレス反応」
情緒的反応として、不安、イライラ、恐怖、落ち込み、緊張、怒り、罪悪感、感情鈍麻、孤独感、疎外感、無気力などの感情が現れる。
- 「行動的ストレス反応」
怒りの爆発、けんかなどの攻撃的行動、過激な行動、泣く、引きこもり、孤立、拒食・過食、幼児返り、チック、吃音、ストレス場面からの回避行動などが現れる。
- 「身体的ストレス反応」
動悸、異常な発熱、頭痛、腹痛、疲労感、食欲の減退、嘔吐、下痢、のぼせ、めまい、しびれ、睡眠障害、悪寒による震えなど、全身にわたる症状が現れる。

本研究では、「心理的ストレス反応」を取り扱うこととする。

3.2 ストレス評価項目の設定

本節では、心理的ストレス反応を検出するために使用するストレス評価項目について定義する。本研究では、POMS(気分プロフィール)[6] という心理的ストレス尺度を用いる。POMS について研究によると、多様

な感情状態を呈するうつ病等において、その患者特有の感情状態を尺度化できるものと考えられるとしている。POMS で定義されている項目には、緊張-不安、抑うつ-落ち込み、怒り-敵意、活力、疲労、混乱の 6 項目がある。これらの項目の中で、活力以外の項目をストレス要素とし、抽出する。各項目の定義を以下に示す。

- 「緊張-不安」
緊張や不安の高まりが感じられる。神経の昂りや落ち着かないなどの特徴が見られる。
- 「抑うつ-落ち込み」
自分に価値がない、希望が持てないと感じている。罪悪感があるなど自信が喪失している状態が見られる。
- 「怒り-敵意」
怒りや他者への反感が感じられる。内心の腹立たしさや他人に意地悪したいなどの思いがある。
- 「疲労」
疲労感があると感じられる。意欲や活力が低下している状態が見られる。
- 「混乱」
当惑と認知効率の低さが感じられる。頭が混乱して考えがまとまらない状態が見られる。

3.3 SNS ユーザのストレスを含むコメントの抽出とストレス推定

本節では、ChatGPT を用いたストレスを含むコメントの抽出とストレス推定について述べる。本研究で用いるプロンプトを表 1 に示す。プロンプトとは、ChatGPT が応答を生成するための命令文である。プロンプトを調整することで理想的な回答を得られるようにしている。本研究では、OpenAI 社 [7] が提供する ChatGPT (gpt-3.5-turbo) を使用する。また、本研究では SNS として、X(旧 Twitter)[8] を利用する。

3.3.1 ストレスを含むコメントの抽出

本項では、ストレスを含むコメントの抽出について述べる。図 1 にストレスコメントを抽出するシステムの構成を示す。各ストレス項目に対応している数値を格納しているリストを用意している。ストレスが抽出された場合、該当するストレス項目の数値を 1、その他を 0 として、リストを出力として返す。

表 1: ストレス抽出とストレス推定に用いるプロンプト

システム	プロンプト
ストレスコメント抽出	<p>あなたは精神医学に詳しい心理学者として、入力された文章が、以下の6つの項目のいずれかに当てはまるかを、「項目名」で回答してください。いずれの項目にも当てはまらない場合は、「なし」と回答してください。</p> <p><項目名: 定義></p> <p>「緊張と不安」: 緊張や不安の高まりが感じられる。神経の昂りや落ち着かないなどの特徴が見られる。</p> <p>「抑うつと落ち込み」: 自分に価値がない、希望が持てないと感じている。</p> <p>罪悪感があるなど自信が喪失している状態が見られる。</p> <p>「怒りと敵意」: 怒りや他者への反感が感じられる。内心の腹立たしさや他人に意地悪したいなどの思いがある。</p> <p>「疲労」: 疲労感があると感じられる。意欲や活力が低下している状態が見られる。</p> <p>「混乱」: 当惑と認知効率の低さが感じられる。頭が混乱して考えがまとまらない状態が見られる。</p>
ユーザのストレス推定	<p>あなたは精神医学に詳しい心理学者として、入力された文章から、この発言を行った人の各項目の点数を0から4(4が最大)の数値で表した上で、JSON形式のデータ json-data で回答してください。</p> <p>各項目の定義は以下に示す。</p> <p><項目名: 定義></p> <p>「緊張と不安」: 緊張や不安の高まりが感じられる。神経の昂りや落ち着かないなどの特徴が見られる。</p> <p>「抑うつと落ち込み」: 自分に価値がない、希望が持てないと感じている。</p> <p>罪悪感があるなど自信が喪失している状態が見られる。</p> <p>「怒りと敵意」: 怒りや他者への反感が感じられる。内心の腹立たしさや他人に意地悪したいなどの思いがある。</p> <p>「疲労」: 疲労感があると感じられる。意欲や活力が低下している状態が見られる。</p> <p>「混乱」: 当惑と認知効率の低さが感じられる。頭が混乱して考えがまとまらない状態が見られる。</p> <p><点数について></p> <p>0: まったく当てはまらない</p> <p>1: 少し当てはまる</p> <p>2: まあまあ当てはまる</p> <p>3: かなり当てはまる</p> <p>4: 非常に多く当てはまる</p>

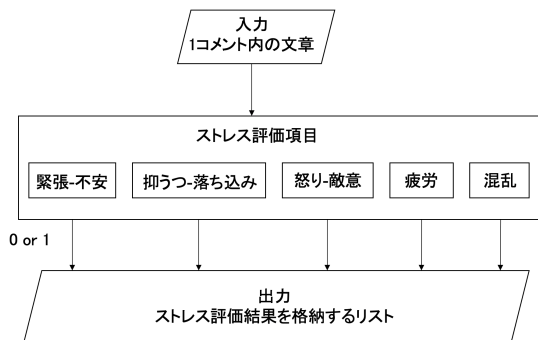


図 1: ストレスコメントを抽出するシステムの構成

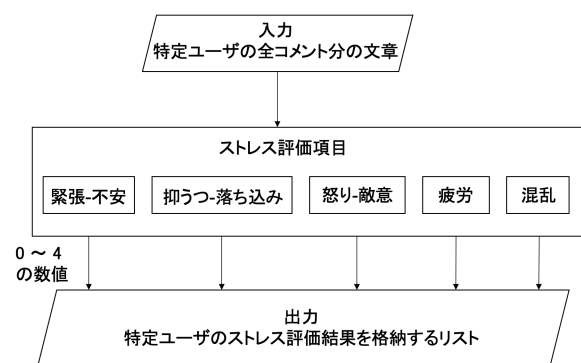


図 2: SNS ユーザのストレスを推定するシステムの構成

3.3.2 SNS ユーザのストレス推定

本項では、SNS ユーザのストレス推定について述べる。図2に SNS ユーザのストレス推定のシステムの構成を示す。X ユーザが投稿している直近のコメントをまとめて入力としている。入力された文章から、設定したストレス項目から X ユーザのストレス状態を推定する。各ストレス項目を POMS と同様に、0 から 4 の五段階で評価する。

- 0: まったく当てはまらない

- 1: 少し当てはまる
- 2: まあまあ当てはまる
- 3: かなり当てはまる
- 4: 非常に多く当てはまる

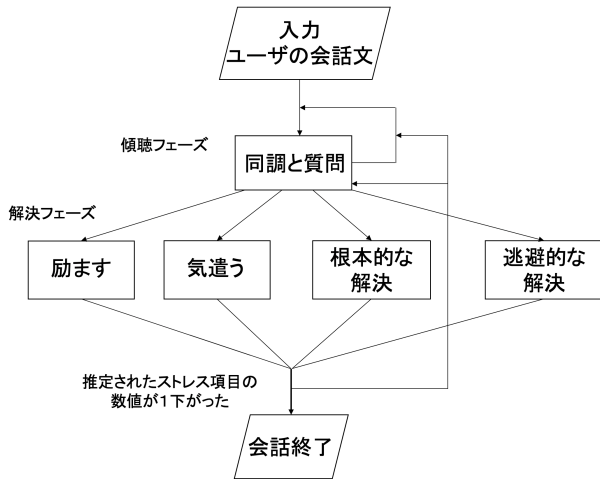


図 3: ストレス低減会話システムの構成

4 ChatGPT を用いたストレス低減会話システムの構築

本章では、ストレス低減会話システムについて述べる。図 3 にストレス低減会話システムの構成を示す。また、現在検討段階ではあるが、会話システムで用いるプロンプト案を表 2 に示す。会話文を生成するプロンプトの〇〇には各会話デッキの内容が入る。

4.1 会話デッキの選択

本節では、ストレス低減会話システムにおける会話デッキの選択について述べる。本研究で扱う会話デッキを以下に示す。

- 同調と質問：悩みを打ち明けてきた相手が話をしたくなるように聞き上手に徹する
- 励ます：相手の失敗を許容し、頑張りを認める
- 気遣う：相手の (精神的・身体的) 体調を気遣う
- 根本的な解決：相手のストレス要因を直接解決する提案をする
- 逃避的な解決：やりたいこと (希望) を尋ね、リフレッシュの手助けをする

傾聴フェーズで「同調と質問」を繰り返し行い、相手の悩みや考え、感情を聞き出す。会話相手の話がひと段落した時点で、解決フェーズへと移行する。解決フェーズでは、傾聴フェーズで聞き出した話の内容から ChatGPT により適切な会話デッキを選択する。

4.2 ストレス状態の再評価

相手が入力した会話文の履歴を ChatGPT への入力とし、ストレス状態の再評価を行う。ここで前章で述べた、POMS のストレス項目を用いる。評価結果により、推定されたストレス項目の点数が一段階下がった場合、ストレス低減成功となり会話が終了する。点数が下がらなかった場合は、解決フェーズに戻りその他の会話デッキを順番に実行していく。それでも点数が下がらなかった場合は、傾聴フェーズへと戻り再度会話を行う。

5 ストレス推定精度の評価検証

本章では、本研究で作成した SNS ユーザのストレスを含むコメントの抽出とストレスを推定するシステムの検証について述べる。X アプリ上でストレスを含むコメントを投稿しているユーザを 52 人を選出した。選出したユーザの最新ポストから順にコメントを取得し、本研究で作成したシステムに入力して分類精度を検証した。

5.1 検証に用いたデータ

X アプリ上で「不安」、「疲れ」、「憂鬱」、「怒り」などを含んだツイートをしているユーザを取得した。本研究では、その中から BOT や宣伝用のアカウントを除いた 52 人のユーザを使用した。選出したユーザから合計 5089 個のコメントを取得した。選出したユーザを本研究で定義したストレス項目を基に、著者の一人がストレスが高いユーザ 22 人と低いユーザ 30 人に分類した。取得したポストも同様にストレスを含むもの 2190 個と、含まないもの 2899 個に分類した。このデータを本検証での正解データとする。

5.2 検証の手順

本節では、実際に行った検証の手順を述べる。前節で述べたデータを扱う。取得した 52 人のユーザを 3.3.2 で述べているストレス推定システムに入力し、ストレスが高いユーザと低いユーザに分類する。システムから出力されたストレス項目で、3(かなり当てはまる)以上の項目が二つ以上のユーザをストレスが高いユーザに分類する。その他をストレスが低いユーザに分類する。前節で述べた分類結果を正解データとし、システムで得られた分類結果と比較する。また、取得したポストを一つずつ 3.3.1 で述べたストレスを含むコメントを抽出するシステムに入力し、ストレスを含むコメントと含まないコメントに分類する。システムから出力

表 2: ストレス低減会話システムで用いるプロンプト

システム	プロンプト
会話デッキの選択	あなたは精神医学に詳しい心理学者です。入力された相手の発話履歴に対して、以下の項目からストレスを低減させるのに適したものを一つ「項目名」で回答してください。 <項目名:項目の定義> 「励ます」:相手の失敗を許容し、頑張りを認める 「気遣う」:相手の身体的・精神的体調を気遣う 「逃避的解決」:やりたい事を尋ね、リフレッシュの手助けをする 「根本的解決」:相手のストレスの原因解決の手助けをする
会話デッキに対応する会話文の生成	入力された相手の発話履歴に対して、〇〇するコメントを50文字以内で作成してください。

表 3: ユーザごとのストレスの有無の推定結果

	man-有	man-無	合計	Recall	Precision
sys-有	13	4	17	0.59	0.76
sys-無	9	26	35	0.87	0.74
合計	22	30	52		

表 4: コメントごとのストレスの有無の推定結果

	man-有	man-無	合計	Recall	Precision
sys-有	1377	380	1757	0.63	0.78
sys-無	813	2519	3332	0.87	0.76
合計	2190	2899	5089		

されたリストで、一つでもストレス項目に当てはまれば、ストレスを含むものとして分類する。その他をストレスを含まないコメントとして分類する。その分類結果を前節で述べた分類結果とを比較する。

5.3 検証結果と考察

本節では、検証実験の結果と考察について述べる。ユーザの分類結果とストレスコメントの分類結果をそれぞれ表3、表4に示す。sys-をシステムの予想、man-を正解データとしている。

5.3.1 ストレスが高いユーザと低いユーザの分類結果の比較

高ストレスユーザ検出の適合率が0.76という値となった。このことから、システムが一定の高ストレスユーザを判断できていることが分かる。しかし、再現率の方が0.59とやや低い値となった。現段階では高ストレスユーザの見逃しが多いということが分かる。

5.3.2 ストレスを含むコメントと含まないコメントの分類結果の比較

ストレスコメント抽出の再現率、適合率がそれぞれ0.63, 0.78となった。前項のユーザ分類と同様に、再現率が低い結果となった。実際に抽出できているコメントをストレス項目ごとにまとめたものを表5に示す。

次に、ストレスコメントの見逃しが原因でユーザ分類の再現率が下がっていると考えられる。

見逃しが多かったストレスコメントを大きく分けると以下ようになった。これらに共通しているのが、辛い状態から逃げたいという気持ちが含まれているということである。

- A: 願望: 肯定文
- B: 願望: 否定文
- C: 不満や怒り

表6にそれぞれの例文を示す。Aの例について、仕事を辞めたいや帰りたいなどのしんどい状態から逃げたいという気持ちが含まれている文だと考えられる。Bの例については、働きたくないや動きたくないなど「～したい」に否定の表現が付属している。AとBの例に関しては、「～したい」という願望表現に着目して抽出できると考えられるため、願望表現を伴うストレスの例をプロンプトに含める改善案が考えられる。最後のCについては、他の事例に該当しないが、文の内容から不満や怒りが感じられるものである。もう少し実験を重ねて判定漏れが生じている例文を集めた上で、不満や怒りを含むストレス例についても、プロンプトに含めていければと考えている。

6 おわりに

精神的サポートが必要な人を早期に発見するために、テキスト情報から自動的にストレスを推定するシステムを構築した。構築したストレス推定システムが、実

表 5: 各ストレス項目に該当する抽出コメントの例

ストレス項目	例文
緊張-不安	最近ある事がきっかけでまた例の幼少期に言われた言葉を思い出してしまって。嫌な記憶ほど鮮明に。もう声を聴くと体がギョッてるほど限界なのよ、もう新人に関わりたくない、ほんとにしんどいの、何をどうすれば伝わるのかもわかんない
抑うつ-落ち込み	どうやって休もうかなと考えてる時点で多分病んでるかも 休みが終わる日の夜になるといつも死にたくなります 何かね、もう何も頑張る気が起きんのよ
怒り-敵意	患者に「看護師呼んできて」と言うけど、ワイも看護師なんだが?しばくぞ あー、イライラする。父の優柔不断さに 私はひらきなおって寝てるのだが腹立って眠れなくなった。あー、腹立つ
疲労	お弁当は、作ったけど、また、布団の中だるくて、おきあがれない 疲れが、どっどで夕飯つくれるか、材料は、きってるのに布団から、おきあがれない。 久しぶりにごりごり仕事してる..... 疲れた.....
混乱	今日ダメ... 浮き沈み激しくて... 無理... ずっと病んでるし何で精神不安定なのか分からないそれを考えて混乱してる 勝手に先輩が休みの日をずらしたから予定狂ったし帰るのが遅くなって混乱してるやることあったのに何も出来ない

表 6: システムが抽出できなかったストレスがありそうなコメントの例

項目	例文
A	学校辞めたい 会社辞めたいほんと早く 10 月になって欲しい あぁ... 仕事お帰りてえ 先生に『死にたい』って相談したら迷惑だよ
B	応援が、励ましが、期待が、苦しいキツイツライ聞きたくない 働きたくない働きたくない働きたくない
C	サマリー書く時の虚無感って何なんやろ適当にあああって書いて済ませやる 過眠治すってタイミング大事でしょ?日光に当たって、はい覚醒、リズム整うってそんな簡単な話じゃない

際に特定の SNS ユーザのストレス状態とストレスコメントを抽出できているかの検証を行った。その結果、それぞれの適合率が 0.76 と 0.78 という精度であった。一定のストレスコメントを判定できていることが分かった。しかし、再現率がどちらも 0.59, 0.63 という値となり、ストレスコメントの見逃しが見られた。考察から、辛い現実から逃げたいという気持ちを表す文が検出できていないことが分かった。今後の課題として、システムの精度の向上と、ChatGPT を用いたストレス低減会話システムを完成させた上で、その有効性を示す実験を行っていきたい。

参考文献

- [1] ビジネスパーソンの健康意識 2017: (URL) https://www.jrta.co.jp/pdf/research/business_person/person_vol21.pdf (2023 年 11 月 10 日確認)
- [2] 大野 志郎 : ネット逃避の現状—インターネットおよびアプリケーションの過剰使用者へのオンラインインタビュー調査より, 総務省学術雑誌『情報通信政策研究』, Vol.2, No.2, pp.1-17, (2019)
- [3] 四方田 健二:新型コロナウイルス感染拡大に伴う不安やストレスの実態 Twitter 投稿内容の計量テキスト分析から, 体育学研究, Vol.65, pp. 757-774, (2020)
- [4] 高須 遼, 中村 啓信, 岸本 泰士郎, 狩野 芳伸:大規模ツイートデータを用いたメンタルヘルス不調者の推測, 第 36 回人工知能学会全国大会, 1J4-OS-13a-04, (2022)
- [5] 第 2 章心のケア各論: (URL) https://www.mext.go.jp/a_menu/shotou/clarinet/002/003/010/003 (2023 年 11 月 10 日確認)
- [6] 赤林 朗, 横山 和仁, 荒記 俊一, 島田 恭子: POMS(感情プロフィール検査) 日本語版の臨床応用の検討, 心身医学, Vol.31, No.7, pp.577-582, (1991)
- [7] OpenAI: (URL) <https://openai.com/> (2023 年 11 月 10 日確認)
- [8] Twitter について: (URL) <https://about.twitter.com/ja> (2023 年 11 月 10 日確認)

マルチモーダル情報に基づく画像説明文の教師あり自動評価 Supervised Automatic Evaluation for Image Captioning Based on Multimodality

齋藤 大地* 和田 唯我 兼田 寛大 杉浦 孔明
Daichi Saito Yuiga Wada Kanta Kaneda Komei Sugiura

慶應義塾大学
Keio University

画像キャプション生成では、モデルが出力した生成文の品質を適切に評価することが重要である。しかし、 n -gram に基づく自動評価尺度は人間による評価との相関が低いことが報告されている。日本語の画像キャプション生成では JaSPICE がそれらに代わる自動評価尺度として提案されているものの、表層表現の不一致に対して適切に評価を行うことができない。また、COMET をはじめとする学習可能な自動評価尺度は、機械翻訳における自動評価タスクに最適化されており画像を考慮しないため、画像キャプション生成には適していない。そこで本論文では、画像キャプション生成に対する自動評価尺度 SuiSei を提案する。SuiSei は、画像特徴量と言語特徴量を扱うマルチモーダル特徴抽出機構および idf を考慮した文埋め込み機構を用いて人間による評価を回帰する。実験の結果、SuiSei はベースライン尺度と比較して人間による評価との相関係数が高いことを示した。

1 はじめに

画像キャプション生成は、視覚障害者支援 [Gurari 20, Ahsan 21, Dognin 22, Ghandi 23], 医療画像解析 [Pavlopoulos 19, Huang 21, Ayesha 21], ロボティクスにおける説明生成 [Kambara 22, Magassouba 20, Ogura 20] など、幅広い分野に応用されている。本研究分野では、画像キャプション生成モデルが出力した生成文の品質を適切に評価することが重要である。

先行研究では、 n -gram に基づく自動評価尺度は人間による評価との相関が低いことが指摘されている [Anderson 16]。そのため、JaSPICE [Wada 23] などの人間による評価との相関が高い自動評価尺度が提案されているものの、表層表現の不一致に対して適切に評価を行うことができない。また、COMET [Rei 20] をはじめとする学習可能な自動評価尺度は、機械翻訳における自動評価タスクに最適化されており、画像を考慮しないため画像キャプション生成における性能は不十分である。したがって、画像キャプション生成において人間による評価との相関が十分に高い自動評価尺度が構築されることが望ましい。

画像キャプション生成では、画像のどこに着目するか、画像をどのように自然言語で表現するかなどによって無数の正解文が存在し得る。そのため、生成文が正

解文と大きく異なっても画像に対して適切である場合があり、単なる文の類似度比較だけでは不十分な点において困難なタスクである [Yi 20]。

そこで本論文では、画像キャプション生成モデルに対する自動評価尺度 SuiSei¹ を提案する。SuiSei は、画像特徴量と言語特徴量を扱うマルチモーダル特徴抽出機構および idf (inverse document frequency, 逆文書頻度) を考慮した文埋め込み機構を用いた回帰を行うことで、人間による評価との高い相関を実現する。

例えば、正解文が {「少年がサッカーをしている」, 「男の子がサッカーの試合をしている」} であるような画像に対し、画像キャプション生成モデルが「子供がサッカーボールを蹴っている」という生成文を出力したとする。このとき、提案手法は画像と正解文に対して生成文がどの程度適切であることを示す評価値を計算する。

提案手法が既存手法と異なる点は、画像特徴量と言語特徴量を扱うマルチモーダル特徴抽出機構を自動評価尺度に導入する点、および idf を考慮した文埋め込み機構を導入する点である。マルチモーダル特徴抽出機構の導入により、正解文だけでなく画像に対しても生成文が適切であることを考慮することができる。また、idf を考慮した文埋め込み機構の導入により、より重要な単語に注目して評価値を予測することができる。

提案手法の新規性は以下の通りである。

*連絡先：慶應義塾大学理工学部情報工学科
〒 223-8522 神奈川県横浜市港北区日吉 3-14-1
E-mail: daichi-s@keio.jp

¹Supervised mUltimodal evaluatIon System for imagE captioning

- 画像キャプション生成に対する自動評価尺度に画像特徴量と言語特徴量を扱うマルチモーダル特徴抽出機構を導入した SuiSei を提案する.
- 画像キャプション生成に対する自動評価尺度に idf を考慮した文埋め込み機構を導入する.

2 関連研究

画像キャプション生成の研究は広く行われており [Li 20, Alayrac 22, Yu 22, Li 23], 鑑賞者の感情に基づいた絵画の説明文生成 [Ishikawa 23] や Transformer による生活支援ロボットの指示文生成 [Kambara 22] など, さまざまな分野に応用されている. 画像キャプション生成に関するサーベイ論文である [Ming 22] では, 画像キャプション生成モデル, 標準データセット, 評価尺度などについての包括的な総括がなされている.

画像キャプション生成における標準的な評価尺度としては, BLEU [Papineni 02], ROUGE [Lin 04], METEOR [Banerjee 05], CIDEr [Vedantam 15], BERT-Score [Zhang 20] などが挙げられる. また COMET は, 深層学習に基づいて人間による評価を回帰する学習可能な自動評価尺度であり, ルールベースの自動評価尺度と比較して人間による評価との相関が高いことが示されている [Rei 20]. 日本語の画像キャプション生成では, シーングラフに基づいて評価を行う JaSPICE も既存手法に比べ人間による評価と相関が高いことが報告されている [Wada 23]. 近年では, BERT [Devlin 19] に基づいて対照学習を行う UMIC [Lee 21] や, CLIP [Radford 21] を用いて画像と生成文の類似度を計算する CLIP-Score [Hessel 21] など, 正解文を使用しない自動評価尺度も登場している.

3 問題設定

本論文では, 画像キャプション生成に対する自動評価を扱う. 画像キャプション生成における自動評価尺度は, 人間による評価に近いことが望ましい. 具体的には, 自動評価尺度の評価値と人間による評価との相関係数が高いことが望ましい.

本タスクでは, 画像 \mathbf{x}_{img} , 生成文 \mathbf{x}_{cand} , N_t 個の正解文 $\{\mathbf{x}_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ を入力とし, \mathbf{x}_{img} と $\{\mathbf{x}_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ に対して \mathbf{x}_{cand} がどの程度適切であるかを示す評価値を計算する. また, 本論文では日本語の画像キャプション生成における自動評価を前提とし, 提案手法の評価には人間による評価との相関係数 (Pearson/Spearman/Kendall の相関係数) を使用する.

4 提案手法

本論文では, 人間による評価を回帰させる COMET を拡張した, 画像キャプション生成に対する自動評価

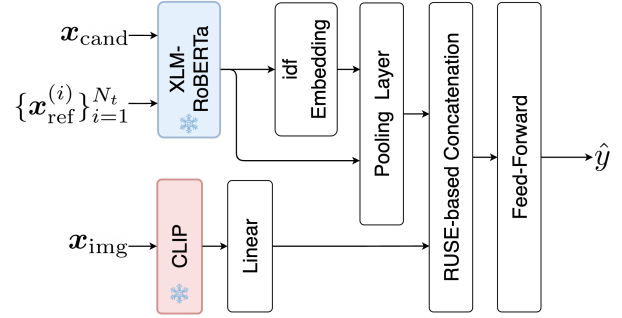


図 1: 提案手法のモデル構造.

尺度 SuiSei を提案する. 本手法における拡張は, 画像特徴量と言語特徴量を用いた人間による評価の回帰であり, 学習可能なパラメータをもつ自動評価尺度一般に適用可能であると考えられる.

提案手法の新規性は以下の通りである.

- 画像キャプション生成に対する自動評価尺度に画像特徴量と言語特徴量を扱うマルチモーダル特徴抽出機構を導入した SuiSei を提案する.
- 画像キャプション生成に対する自動評価尺度に idf を考慮した文埋め込み機構を導入する.

4.1 入力

図 1 に提案手法のモデル構造を示す. ネットワークの入力は以下の \mathbf{x} であり, 出力は \mathbf{x} に対する評価値の予測 \hat{y} である. ここで, \mathbf{x} は以下のように定義される.

$$\mathbf{x} = \left\{ \mathbf{x}_{\text{cand}}, \left\{ \mathbf{x}_{\text{ref}}^{(i)} \right\}_{i=1}^{N_t}, \mathbf{x}_{\text{img}} \right\} \quad (1)$$

ただし, $\mathbf{x}_{\text{cand}} \in \{1, 0\}^{V \times L}$, $\{\mathbf{x}_{\text{ref}}^{(i)}\}_{i=1}^{N_t} \in \{1, 0\}^{N_t \times V \times L}$, $\mathbf{x}_{\text{img}} \in \mathbb{R}^{3 \times H \times W}$ である. ここで, V は語彙サイズ, L は最大トークン長, N_t は 1 サンプルにおける正解文の数, H, W はそれぞれ \mathbf{x}_{img} の高さおよび幅を表す.

本モデルではまず, XLM-RoBERTa [Conneau 20] を使用して \mathbf{x}_{cand} および $\{\mathbf{x}_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ からそれぞれ言語特徴量 $\mathbf{h}_{\text{cand}} \in \mathbb{R}^{L \times d_{\text{XLM}}}$ および $\{\mathbf{h}_{\text{ref}}^{(i)}\}_{i=1}^{N_t} \in \mathbb{R}^{N_t \times L \times d_{\text{XLM}}}$ を得る. ここで, d_{XLM} は XLM-RoBERTa の出力次元数である. また, 事前学習済みの CLIP 画像エンコーダ (ViT-B/16) [Radford 21] を用い, \mathbf{x}_{img} から画像特徴量 $\mathbf{h}_{\text{img}} \in \mathbb{R}^{d_{\text{CLIP}}}$ を得る. ここで, d_{CLIP} は CLIP 画像エンコーダの出力次元数である.

4.2 文埋め込み機構

idf embedding では, \mathbf{h}_{cand} および $\{\mathbf{h}_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ から idf を考慮した文埋め込みを作成する. 文章の類似性の計算において, 文書集合の中でその語を含む文書数, すなわち文書頻度 (document frequency) が低い単語は, 文書頻度が高い単語よりも重要であることが知られて

いる [Zhang 20]. そこで本モジュールでは、文書頻度の低いトークンをより重要なトークンとして評価値の予測に反映させるために、トークンごとに idf を計算し、 \mathbf{h}_{cand} および $\{\mathbf{h}_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ に重みとして掛け合わせる。ここで、 N 個の正解文が与えられたとき、トークン w に対する文書頻度を $\text{df}(w)$ とすると、 w に対する idf は以下のように計算される。

$$\text{idf}(w) = \log \frac{N}{\text{df}(w)} \quad (2)$$

これを用いて、以下のように \mathbf{h}_{cand} から $\mathbf{h}'_{\text{cand}}$ を得る。

$$\mathbf{h}'_{\text{cand}} = \left\{ \text{idf}(\mathbf{x}_{\text{cand}}^{(l)}) \cdot \mathbf{h}_{\text{cand}}^{(l)} \mid l = 1, \dots, L \right\} \quad (3)$$

ただし、 L は最大系列長を表す。同様に、 $\{\mathbf{h}_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ から idf を考慮した文埋め込み $\{\mathbf{h}'_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ を作成し、 $\{\mathbf{h}'_{\text{cand}}\}$, $\{\mathbf{h}'_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ を得る。ここで、term frequency (tf) を使用しない理由は、処理する対象が単一の文であり、多くの場合 tf の値が 1 になるためである。

4.3 マルチモーダル特徴抽出機構

RUSE-based concatenation では、COMET と同様に RUSE [Shimamaka 18] に基づく手法を用いて各特徴量を結合し、Feed-Forward Network (FFN) により \mathbf{x} に対する評価値の予測 \hat{y} を得る。入力は $\{\mathbf{h}_{\text{cand}}\}$, $\{\mathbf{h}_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$, $\mathbf{h}'_{\text{cand}}$, $\{\mathbf{h}'_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$, \mathbf{h}_{img} 、出力は \hat{y} である。本モジュールでは、RUSE に基づく手法を用いて \mathbf{h}_{cand} および $\{\mathbf{h}_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ から以下のように \mathbf{h}_{RUSE} を得る。

$$\mathbf{h}_{\text{RUSE}} = \left\{ \left\{ \mathbf{h}_{\text{cand}} \odot \mathbf{h}_{\text{ref}}^{(i)} \right\}_{i=1}^{N_t}, \left\{ \left| \mathbf{h}_{\text{cand}} - \mathbf{h}_{\text{ref}}^{(i)} \right| \right\}_{i=1}^{N_t}, \mathbf{h}_{\text{cand}} \odot \mathbf{h}_{\text{img}}, \left| \mathbf{h}_{\text{cand}} - \mathbf{h}_{\text{img}} \right| \right\} \quad (4)$$

ここで、 \odot はアダマール積を表し、 \mathbf{h}_{cand} および $\{\mathbf{h}_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ には Pooling 層を適用した。最後に、 \mathbf{h}_{RUSE} , \mathbf{h}_{cand} , $\{\mathbf{h}'_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ を結合し、 \mathbf{h} を得る。同様に、 $\mathbf{h}'_{\text{cand}}$, $\{\mathbf{h}'_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ についても RUSE に基づく変換を行い、 \mathbf{h} と結合して FFN により出力 \hat{y} を得る。損失関数には、平均二乗誤差を用いた。

5 実験設定

本研究では、画像キャプション生成における人間による評価を学習するため、*Shichimi* データセット [Wada 23] および PFN-PIC-gen データセットを用いた。*Shichimi* データセットおよび PFN-PIC-gen は、日本語の画像キャプション生成に対する自動評価タスクにおいて最大規模のコーパスである。これらのデータセットにおける人間による評価は、画像に対して生成文が適切であるかを 5 段階で評価したものである。*Shichimi* デー

表 1: 提案手法の設定。

最適化手法	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
学習率	3.0×10^5
バッチサイズ	32
エポック数	9

タセットは、画像、生成文、正解文、および生成文に対する人間による評価で構成された大規模コーパスであり、500 人のアノテータによる評価値を含む。サンプル数は 103,170、語彙サイズは 6,269、平均文長は 12.4 文字である。また、PFN-PIC-gen は [Wada 23] で収集されたものであり、PFN-PIC [Hatori 18] の画像から生成したキャプションに対して人間による評価が付与されたデータセットを指す。サンプル数は 1,920、語彙サイズは 647、平均文長は 20.2 文字である。

提案手法の評価には、SuiSei が出力した評価値 $\{\hat{y}_i\}_{i=1}^N$ と人間による評価値 $\{y_i\}_{i=1}^N$ に対する相関係数を用いた。ここで、 N はサンプル数を表す。相関係数には、Pearson, Spearman, Kendall の相関係数を使用した。

本実験では、*Shichimi* データセットを訓練集合とテスト集合に分割し、それぞれ 51,988 サンプル、51,182 サンプルとした。また、PFN-PIC-gen は、ゼロショット性能を検証するために全てテスト集合として使用した。ただし、訓練集合をモデルの学習に用い、テスト集合を評価に用いて実験を行った。

表 1 に提案手法の設定を示す。ここで、提案手法における訓練可能パラメータ数は 1.84×10^8 であった。また、モデルの学習にはメモリ 24GB 搭載の GeForce RTX 3090 および Intel Core i9 12900K を使用し、訓練時間および 1 サンプルあたりの推論時間は、それぞれ約 1.2 時間および約 7.6ms であった。

6 実験結果

6.1 定量的結果

表 2, 表 3 に *Shichimi* データセットおよび PFN-PIC-gen における定量的結果を示す。本タスクでは、自動評価尺度の評価値と人間による評価との相関係数に基づいて自動評価尺度を評価する。ベースライン尺度には、日本語の画像キャプション生成において標準的な尺度である BLEU, ROUGE, METEOR, CIDEr, JaSPICE を用いた。また、提案手法は COMET を拡張したものであるため、COMET もベースライン尺度に採用した。

表 2 より、*Shichimi* データセットにおける提案手法は、Pearson, Spearman, Kendall の相関係数において、それぞれ 0.672, 0.644, 0.504 であり、JaSPICE と比較して 0.173, 0.113, 0.091 ポイント上回った。また、COMET と比較して、0.048, 0.071, 0.062 ポイント上回った。*Shichimi* データセットにおいて、人間による評価に対する Pearson, Spearman, Kendall の相関係数

表 2: *Shichimi* データセットにおける各自動評価尺度と人間による評価との相関係数.

	Pearson	Spearman	Kendall
BLEU	0.296	0.343	0.260
ROUGE	0.366	0.340	0.258
METEOR	0.345	0.366	0.279
CIDEr	0.312	0.355	0.269
JaSPICE	0.499	0.531	0.413
COMET	0.624	0.573	0.442
SuiSei	0.672	0.644	0.504
Human	0.759	0.750	0.669

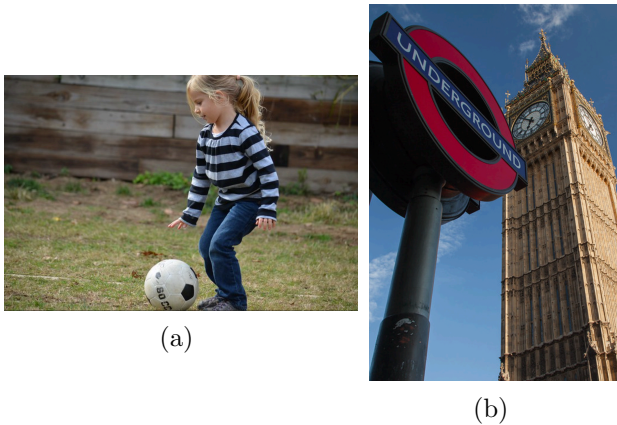


図 2: 成功例における画像.

はそれぞれ, 0.759, 0.750, 0.669 であった. 人間による評価に対する相関係数が 1.0 よりも小さい理由は, 人間による評価に完全な一貫性がなく, 同一サンプルに対する評価値が必ずしも一致しないためである. また, 人間による評価に対する相関係数は, 自動評価尺度の性能における上限値であると考えられる.

同様に, 表 3 より PFN-PIC-gen における提案手法は, Pearson, Spearman, Kendall の相関係数において, それぞれ 0.576, 0.590, 0.443 であり, JaSPICE と比較して 0.030, 0.017, 0.005 ポイント, COMET と比較して, 0.137, 0.155, 0.118 ポイント上回った.

6.2 定性的結果

図 2 に定性的結果における成功例を示す. 図 2 (a) は *Shichimi* データセットにおける結果の一つであり, \mathbf{x}_{cand} は「デニムパンツをはいた少女がサッカーボールを蹴ろうとしている」, $\{\mathbf{x}_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ は「少女がサッカーボールと戯れている」, 「ポーター柄のシャツを着た少女がサッカーボールで遊んでいる」} である. 図 2 (a) における人間による評価値 y と提案手法の予測値 \hat{y} はそれぞれ, $y = 5, \hat{y} = 0.974$ であった. テスト集合において, この入力に対する SuiSei の値は上位 3% の値であるため, 提案手法は図 2 (a) の例において人間による評価に近い評価値を出力していると言える.

表 3: PFN-PIC-gen における各自動評価尺度と人間による評価との相関係数.

	Pearson	Spearman	Kendall
BLEU	0.484	0.466	0.352
ROUGE	0.500	0.474	0.365
METEOR	0.423	0.457	0.352
CIDEr	0.416	0.462	0.353
JaSPICE	0.547	0.573	0.438
COMET	0.439	0.435	0.325
SuiSei	0.576	0.590	0.443

表 4: Ablation study の結果.

Model	\mathbf{x}_{img}	idf を考慮した 文埋め込み	Pearson	Spearman	Kendall
(i)		✓	0.532	0.536	0.402
(ii)	✓		0.517	0.515	0.386
(iii)	✓	✓	0.576	0.590	0.443

同様に図 2 (b) も *Shichimi* データセットにおける定性的結果の一つであり, \mathbf{x}_{cand} は「時計塔の前に道路標識が立っている」, $\{\mathbf{x}_{\text{ref}}^{(i)}\}_{i=1}^{N_t}$ は「ビッグベンの前に地下鉄のサインが出ている」, 「高い時計台と地下と表示された看板がある」} である. 図 2 (b) における人間による評価値 y と提案手法の予測値 \hat{y} はそれぞれ, $y = 5, \hat{y} = 0.950$ であった. テスト集合において, この入力に対する SuiSei の値は上位 6% の値であるため, 提案手法は図 2 (b) の例においても人間による評価に近い評価値を出力していると言える.

6.3 Ablation Study

次に, 以下の二つの条件を Ablation study に定めた. その結果を表 4 に示す.

Image Ablation 入力から \mathbf{x}_{img} を取り除くことによる性能への影響を調査した. その結果, Model (i) は Pearson, Spearman, Kendall の相関係数においてそれぞれ 0.532, 0.536, 0.402 であり, Model (iii) と比較して 0.044, 0.054, 0.041 ポイント下回った. この結果から, \mathbf{x}_{img} の自動評価尺度への導入が提案手法の性能向上に寄与していることが確認できた.

Sentence Embedding Ablation idf を考慮した文埋め込みを FFN の入力から取り除くことによる性能への影響を調査した. その結果, Model (ii) は Pearson, Spearman, Kendall の相関係数においてそれぞれ 0.517, 0.515, 0.386 であり, Model (iii) と比較して 0.059, 0.075, 0.057 ポイント下回った. この結果から, idf を考慮した文埋め込みの自動評価尺度への導入が提案手法の性能向上に寄与していることが確認できた.

7 結論

本論文では、画像キャプション生成に対する自動評価を扱った。本研究の貢献を以下に示す。

- 画像キャプション生成に対する自動評価尺度に、画像特徴量と言語特徴量を扱うマルチモーダル特徴抽出機構を導入した SuiSei を提案した。
- 画像キャプション生成に対する自動評価尺度に idf を考慮した文埋め込み機構を導入した。
- SuiSei はベースライン尺度と比較して、人間による評価との相関係数が高いことを示した。

謝辞

本研究の一部は、JSPS 科研費 23H03478, JST ムーンショット, NEDO の助成を受けて実施されたものである。

参考文献

- [Ahsan 21] Ahsan, H., Bhalla, N., Bhatt, D., and Shah, K.: Multi-Modal Image Captioning for the Visually Impaired, in *NAACL*, pp. 53–60 (2021)
- [Alayrac 22] Alayrac, J.-B., Donahue, J., Luc, P., et al.: Flamingo: a Visual Language Model for Few-Shot Learning, *NeurIPS*, Vol. 35, pp. 23716–23736 (2022)
- [Anderson 16] Anderson, P., Fernando, B., Johnson, M., and Gould, S.: SPICE: Semantic Propositional Image Caption Evaluation, in *ECCV*, pp. 382–398 (2016)
- [Ayesha 21] Ayesha, H., et al.: Automatic Medical Image Interpretation: State of the Art and Future Directions, *Pattern Recognition*, Vol. 114, p. 107856 (2021)
- [Banerjee 05] Banerjee, S. and Lavie, A.: METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments, in *ACL*, pp. 65–72 (2005)
- [Conneau 20] Conneau, A., Khandelwal, K., et al.: Unsupervised Cross-lingual Representation Learning at Scale, in *ACL*, pp. 8440–8451 (2020)
- [Devlin 19] Devlin, J., et al.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *NAACL-HLT*, pp. 4171–4186 (2019)
- [Dognin 22] Dognin, P., et al.: Image Captioning as an Assistive Technology: Lessons Learned from VizWiz 2020 Challenge, *JAIR*, Vol. 73, pp. 437–459 (2022)
- [Ghandi 23] Ghandi, T., Pourreza, H., and Mahyar, H.: Deep Learning Approaches on Image Captioning: A Review, *ACM Comput. Surv.*, Vol. 56, No. 3 (2023)
- [Gurari 20] Gurari, D., Zhao, Y., Zhang, M., and Bhat-tacharya, N.: Captioning Images Taken by People Who Are Blind, in *ECCV*, pp. 417–434 (2020)
- [Hatori 18] Hatori, J., et al.: Interactively Picking Real-World Objects with Unconstrained Spoken Language Instructions, in *ICRA*, pp. 3774–3781 (2018)
- [Hessel 21] Hessel, J., Holtzman, A., et al.: CLIPScore: A Reference-free Evaluation Metric for Image Captioning, in *EMNLP*, pp. 7514–7528 (2021)
- [Huang 21] Huang, J.-H., Wu, T.-W., et al.: Contextualized Keyword Representations for Multi-modal Retinal Image Captioning, in *ICMR*, pp. 645–652 (2021)
- [Ishikawa 23] Ishikawa, S. and Sugiura, K.: Affective Image Captioning for Visual Artworks Using Emotion-Based Cross-Attention Mechanisms, *IEEE Access*, Vol. 11, pp. 24527–24534 (2023)
- [Kambara 22] Kambara, M., et al.: Relational Future Captioning Model for Explaining Likely Collisions in Daily Tasks, in *IEEE ICIP*, pp. 2601–2605 (2022)
- [Lee 21] Lee, H., Yoon, S., Deroncourt, F., and Jung, K.: UMIC: An Unreferenced Metric for Image Captioning via Contrastive Learning, in *ACL*, pp. 220–226 (2021)
- [Li 20] Li, X., Yin, X., Li, C., Zhang, P., and Hu, X. o.: Oscar: Object-Semantics Aligned Pre-training for Vision-language Tasks, in *ECCV*, pp. 121–137 (2020)
- [Li 23] Li, J., Li, D., et al.: BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models, *ICML* (2023)
- [Lin 04] Lin, C.-Y.: ROUGE: A Package for Automatic Evaluation of Summaries, in *Text summarization branches out*, pp. 74–81 (2004)
- [Magassouba 20] Magassouba, A., Sugiura, K., et al.: Multimodal Attention Branch Network for Perspective-Free Sentence Generation, in *CoRL*, pp. 76–85 (2020)
- [Ming 22] Ming, Y., et al.: Visuals to Text: A Comprehensive Review on Automatic Image Captioning, *IEEE/CAA JAS*, Vol. 9, No. 8, pp. 1339–1365 (2022)
- [Ogura 20] Ogura, T., Magassouba, A., Sugiura, K., et al.: Alleviating the Burden of Labeling: Sentence Generation by Attention Branch Encoder–Decoder Network, *IEEE RAL*, Vol. 5, No. 4, pp. 5945–5952 (2020)
- [Papineni 02] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J.: BLEU: a Method for Automatic Evaluation of Machine Translation, in *ACL*, pp. 311–318 (2002)
- [Pavlopoulos 19] Pavlopoulos, J., Kougia, V., and Androutsopoulos, I.: A Survey on Biomedical Image Captioning, in *SiVL*, pp. 26–36 (2019)
- [Radford 21] Radford, A., Kim, J. W., Hallacy, C., et al.: Learning transferable visual models from natural language supervision, in *ICML*, pp. 8748–8763 (2021)
- [Rei 20] Rei, R., Stewart, C., Farinha, A. C., and Lavie, A.: COMET: A Neural Framework for MT Evaluation, in *EMNLP*, pp. 2685–2702 (2020)
- [Shimanaka 18] Shimanaka, H., et al.: RUSE: Regressor Using Sentence Embeddings for Automatic Machine Translation Evaluation, in *WMT18*, pp. 751–758 (2018)
- [Vedantam 15] Vedantam, R., Lawrence Zitnick, C., and Parikh, D.: CIDEr: Consensus-Based Image Description Evaluation, in *CVPR* (2015)
- [Wada 23] Wada, Y., Kaneda, K., and Sugiura, K.: JaSPICE: Automatic Evaluation Metric Using Predicate-Argument Structures for Image Captioning Models, in *CoNLL* (2023)
- [Yi 20] Yi, Y., Deng, H., and Hu, J.: Improving Image Captioning Evaluation by Considering Inter References Variance, in *ACL*, pp. 985–994 (2020)
- [Yu 22] Yu, J., Wang, Z., Vasudevan, V., Yeung, L., et al.: CoCa: Contrastive Captioners are Image-Text Foundation Models, *TMLR* (2022)
- [Zhang 20] Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., and Artzi, Y.: BERTScore: Evaluating Text Generation with BERT, in *ICLR* (2020)

複数回にわたる楽器指導の特徴分析 —ギターレッスンを対象として—

Analysis of Instrumental Instruction Across Multiple Times - A Focus on Guitar Lessons -

飯野なみ^{1,2*} 三浦寛也² 武田英明¹
Nami Iino^{1,2} Hiroya Miura² Hideaki Takeda¹

¹ 国立情報学研究所

¹ National Institute of Informatics

² 理化学研究所

² RIKEN Center for Advanced Intelligence Project

Abstract: In instrumental instruction, there exists a unique form of interaction where musical performance and verbal communication intersect. Particularly, when multiple lessons are conducted for a piece of music, there may be specific trends and patterns in the process in terms of the student's progress and the teacher's perspective. In this study, we analyzed audio data recorded from one-on-one classical guitar lessons to visualize the implicit aspects of instructional processes.

1 はじめに

本稿では、楽器指導における非明示的な指導過程の可視化を目的として、複数回にわたって実施されたギターレッスンの音声データを対象とした分析について述べる。通常、楽器指導(以降、レッスンと呼ぶ)は生徒の目的に合わせて選定された特定の楽曲に対して何度も行われる。そのレッスン過程では、生徒の上達や教師の着眼点などの側面で特徴的なパターンや変化が見られる可能性がある。特に、レッスン特有の演奏と発話によるインタラクションや教師の指導内容に着目することによって、レッスンの様相を明らかにすることが期待できる。

このような背景から、我々はこれまでクラシックギターの個人レッスンを対象として、実際のレッスン現場で録音された音声データの収集と分析を進めてきた[1, 2]。現在、音声データは対面レッスンとオンラインレッスンを合わせて100件を超えるが、簡易的に収録されたデータであるために、音声区間や発話の情報を適切に抽出することが難しいという問題がある。中でも人手による発話の書き起こしは人的コストが大きく、これまでの先行研究で扱ったデータ数は極めて少なかった。

そこで本研究では、データの拡充とレッスン内の状

態を可視化することを試みた。データの拡充では、教師と生徒のペアを増やし、既存の音声区間検出フレームワークや音声認識モデルを用いて分析データを作成した。レッスンの分析では、教師の発話に焦点を当て、指導内容の意味的要素を表現したラベルを定義・付与し、レッスン内で起こる発話と演奏の状態遷移を調査した。さらに状態遷移については、要素間の因果関係や包含関係を体系化するために、ISM法を適用した構造分析も実施した。結果として、レッスンのフェーズや楽曲によって様相が変化することが明らかとなった。

楽器の個人レッスンにおけるインタラクションに関する既存研究は、ピアノを対象としたものが多く見られる。グループレッスンに比べて個人レッスンは教師が支配的であり、生徒の相互作用は限定的であることを示した研究[3]や、教師と生徒によって演奏表現が協働的に構築されていくことを明らかにした研究[4]などがある。また、ジャズピアノの即興レッスンでは、教師によってフレーズが再構成されていくことを示されている[5]。このような教師と生徒のバランスについて、複数回のレッスンを考慮した研究は少ない。そのため、本研究は教師と生徒のインタラクションの変遷を明らかにするという点で新しく、かつ挑戦的な試みと言える。

*連絡先：国立情報学研究所
〒101-0003 東京都千代田区一ツ橋2丁目1-2
E-mail: nami-iino@nii.ac.jp

表 1: 分析対象であるレッスンデータの概要

ID	ペア	形態	回数	合計時間	曲名/作曲者	時代	フェーズ
1	T1-S1	対面	2	0:11:14	8 度のエチュード/A. Cano	古典派	初期
2			0:07:38	6 度のエチュード/F. Sor	古典派	初期	
3	T2-S2	対面	4	1:35:00	アメリカの遺言/M. Llobet	ロマン派	後期
4			0:57:47	魔笛の主題による変奏曲/F. Sor	古典派	後期	
5			2:00:49	スペインの城よりオリーテ/F. M. Torroba	ロマン派	初期	
6	T2-S3	オンライン	5	2:17:19	盗賊の歌/M. Llobet	ロマン派	後期
7			3	1:15:46	マリアルイサ/J. S. Sagreras	ロマン派	後期

2 分析用データの作成

本章では、クラシックギターの個人レッスンで収録された音声データから、音声区間や発話情報といった基本的な情報抽出を行う。さらに、指導の変遷を調査するために、抽出した発話情報に対して意味付加を行う。

2.1 音声区間検出と発話の書き起こし

表 1 に、データの概要を示す。1 曲につき 2 回以上のレッスンが実施された 7 曲を対象として、全 24 件の音声データである。各ペアの音声データは、T1-S1 は Youtube に公開されているレッスン動画¹から、T2-S2 は対面レッスン時に IC レコーダーで録音された音声ファイルから、T2-S3 は Zoom で記録された動画から、それぞれ収集した。ここでのフェーズとは、レッスンが実施された時期、あるいは生徒の楽曲に対する取り組み具合を意味し、初期は新しく取り組み始めた楽曲を扱ったレッスン、後期は生徒が過去に演奏したことがある楽曲のレッスン、あるいは本データ以前にも指導が行われていた場合のレッスンと位置付けた。

本研究ではまず、新たに追加した ID1, 2, 6, 7 の音声区間検出を行った。レッスンにおいて主要素である演奏区間を検出するために、CNN ベースの音声区間検出ライブラリである inaSpeechSegmenter[6] を用いて、演奏 (Music)、発話 (Speech)、ノイズ (Noise)、無音 (NoEnergy) の 4 つの区間を検出した。ただし、楽器レッスンのような演奏と発話が重複する場面においては、Speech が Music に分類されるという問題があった。

そこでより適切な Speech 情報を獲得すべく、次に発話の書き起こしと区間 (セグメント) 情報の抽出を行った。書き起こしの自動化を進めるために、予備調査として YouTube, Google Cloud Speech-to-Text, Whisper, Amazon Transcribe の精度を比較した。その結果、Whisper の精度が 80% を超え、それ以外は 5 割を下回っ

た。Whisper は、OpenAI が開発した音声認識モデルであり、セグメント情報の取得も可能である。そのため、本研究ではこのモデルを採用して書き起こしを実施した。ただし、話者分離には対応しておらず、かつ既存モデルと組み合わせた場合も精度が低かったことから、手動で教師と生徒の発話を分離した。結果として、総セグメント数は、教師 (T-Speech) は 1669 件、生徒 (S-Speech) は 981 件となった。

2.2 教師の発話に対する意味的要素の付加

教師は、演奏あるいは演奏者の現状を整理し、目標とする演奏に向けた改善点や練習方法を示す。我々は、このような教師の指導における思考過程を体系化すべく、指導コンテンツラベル (Instructional Content Label, ICL) と名付けたラベルを定義した。これは、既存研究 [7] に基づいて定義したものであり、以下の 6 つのラベルで構成され、発話の内容がどのような意味を持つかを分類できる。

- Giving Objective Information (GSI) : 教師の主観に基づく一般的、あるいは特定の概念の情報提供。感想や印象に加え、独り言や雑談を含む。
- Giving Subjective Information (GOI) : 楽譜や音楽関連の知識など、参照可能な事象や概念に基づく客観情報の提供。
- Asking Question (AQ) : 質問、予測、問いかけ。「ここから弾いて」などの指示も含む。
- Giving Feedback (GF) : 生徒の演奏に対する具体的な情報提供。くせなどに関する評価・分析・見立て。
- Giving Practice (GP) : 特定のパッセージの練習方法の提示。部分的な改善・改良に関するコメント。
- Giving Advice (GA) : 目標提示、アドバイス。特定の目標達成へ導くための具体的な意見や推奨。演奏全体へのコメント。

¹ 【クラシックギター レッスン】のえレッスンより Vol.3, 4
https://www.youtube.com/watch?v=ZPjGTsHYwho&list=PL_1N47mmXpTad57CmFKvXIZU9LJMqR0U79&index=7

表 2: 各種音声区間における発話と指導コンテンツラベル (ICL) の一例

種類	開始時間	終了時間	発話内容	ICL
Music	18:30	20:22		
T-Speech	18:35	18:43	あとねもうちょっとなんかこの、あの、はっきり音を出せない、欲しいな。	GSI
T-Speech	19:00	19:05	これもちゃんとあの4分音符分伸ばしてください。	GP
S-Speech	19:06	19:09	あそうか、伸びてない。	
T-Speech	19:13	19:14	そうそう。	GSI
T-Speech	19:56	20:00	今までより音確りクリアで良かったです。	GSI
T-Speech	21:32	21:34	変な音出さない。	GF
S-Speech	21:41	21:43	消えちゃうな。	
T-Speech	21:48	21:50	3ですよ。	GOI
T-Speech	22:00	22:19	もうちょいね、あの、調整を、していただければ。いやでも素晴らしい、いやだからあの、弾くのがどうこうじゃなくて、音楽的な部分、伝えなきゃいけないので、聴衆の人に、	GA
Music	22:08	22:19		

ICL をセグメントごとに付与した一例を表 2 に示す。セグメントの一部には、相槌だけなど明確に分類できない発話や、一つのセグメント内に複数のラベルが該当する発話があった。前者に関しては、前のトピックを引き継いでいたり、暗に示している場合があるため、全てのセグメントに対して前後を考慮してラベルを付与した。後者に関しては、ラベルの優先順位を GSI < GOI < AQ < GF < GP < GA と定め、優先されるラベルを付与した。そのほか、(a) GP は変化を促すものとして、「～してみるのもあり」といった表現の提案を含む、(b) 右手の弾き方は部分的な要素が強いため GP とする、などいくつかのルールを設けた。ラベルの付与は、本著者を含む 2 名が吟味し一意に定め、二者間のアノテーション一致度は 94.5% であった。

(GA) をする様子が多かった。また、1 回目のレッスンではラベルが短い区間で頻繁に変化する傾向が見られたが、レッスンが進むにつれてラベルの変化が少なくなった。このようなレッスンの回数に関する変化は、レッスンのフェーズに着目することで特徴を捉えることができると考えられる。

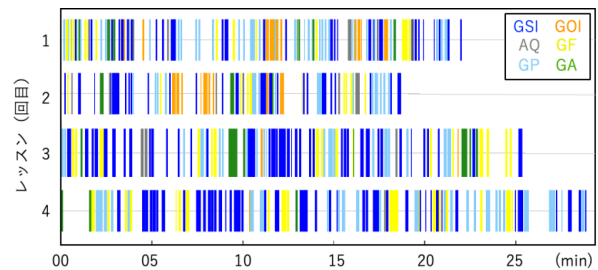


図 1: 教師の発話における ICL の時系列変化の一例

3 ギターレッスンの状態遷移の観察

前節で作成した分析用データを用いて、レッスンの様相を可視化する。

3.1 レッスン内の ICL の変化

図 1 は、ID3 における指導コンテンツラベル (ICL) の時系列変化を示している。例えば、GF (例: なってしまっている) と GP (例: ~したほうがよい) が隣接するパターンが多くみられた。また、レッスンの最初と最後は、GSI になる傾向があった。他のレッスンにおいても、生徒がまず楽曲を通して演奏することが多く、それに対して教師はいきなり具体的な指導 (GP) をするのではなく、パフォーマンス全体に関するコメント

3.2 演奏と発話の状態遷移

本節では、レッスンのフェーズにおける演奏および発話区間の共起関係を調査した。教師の発話 (T-Speech) を 6 種に分類した ICL に加え、演奏 (Music) および生徒の発話 (S-Speech) による合計 8 つの要素を対象として、bigram を抽出し、Python のライブラリである NetworkX を用いて有向グラフを生成した (図 2)。矢印に傍近の数値は、割合である。S-Speech については、発話の割合が少ないことや指導の意味的要素を含まないことから、ラベルによる分類は行わなかった。

初期では、Music, S-Speech, GSI の3つの要素間が強い共起関係にあり、次いで S-Speech と GF, Music と GP, Music と GF の結びつきが強かった。Music と S-Speech は、教師の発話のように分類されていないため、必然的に共起する割合が高くなる。しかし、GF との共起関係も強かったことや、GF と GP それぞれが連続する割合が高かったことから、初期のレッスンは演奏に対する問題提起 (GF) や解決方法の提示 (GP) が中心の指導であることが示唆された。

後期では、GF と他の要素が共起する割合が少なくなった一方で、Music, S-Speech, GSI, GP の共起関係が強くなった。また、Music 間の割合が 13.16% と高く、かつ Music を中心として上記の 3 要素との強い結びつきがあった。つまり、レッスンが進むにつれて、演奏 (Music) を主軸に細かな指摘 (GP) や印象評価 (GSI) の指導に変化することが明らかとなった。

以上のように、状態遷移の可視化によってレッスンの変遷を捉えることができた。しかし、有向グラフではレッスンの様相を適切に読み取ることが難しく、より直感的な表現方法が必要である。

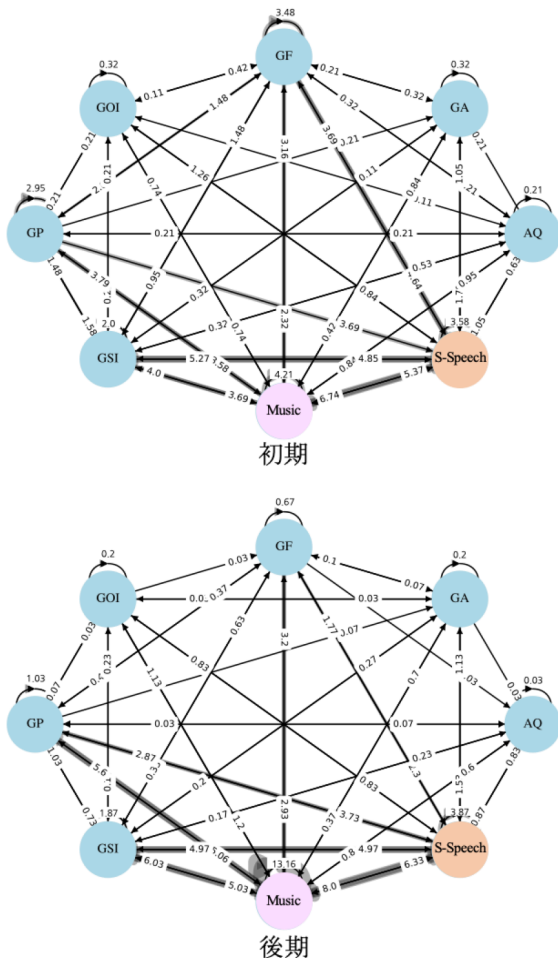


図 2: 教師の発話における ICL の有向グラフ

4 状態遷移結果の構造分析

本節では、前章で述べた指導コンテンツラベル (ICL) の状態遷移結果に対する構造分析をおこなう。我々は、ISM 法 (Interpretive Structural Modeling) による構造化モデリング手法を用いて、ラベル間の因果関係を客観的かつ視覚的に構造化することを試みた。ISM 法とは、グラフ理論の応用として、J.N.Warfield によって提唱された手法の 1 つであり [8]、システムのユーザビリティの問題解決手法や設計仕様の因果関係の分析などで利用されている。要素間の相互関係を一対比較による binary 値の置き換えにより、様々な要素間の相互関係が多階層有向グラフとして図示できる。

4.1 ISM 法の適用手順

ISM 法の適用手順は、(1) 要素の抽出、(2) 隣接行列の作成と行列計算、(3) 構造化である。本手法を用いて、教師の発話における要素間の因果関係や包含関係を可視化する。

(1) 要素の抽出

前章で生成した有向グラフの 8 つの要素 $S = \{ s_1, s_2, \dots, s_8 \}$ を対象として、状態遷移を表現する “binary relation” R を定義する。例えば、要素 s_i が要素 s_j に直接関係している場合、 $s_i R s_j$ と表現される。そして、任意の要素 s_i, s_j, s_k の間に、 $s_i R s_j$ and $s_j R s_k \Rightarrow s_i R s_k$ が成立する場合、関係 R は “transitive” である。グラフ理論を用いると、各要素は頂点 (vertex) に対応し、任意の要素間に関係 R が存在する場合は、辺 (edge) でこれを表現できる。ただし、8 要素の場合は有効辺の数が多いため、特徴的な関係性の正確な理解が難しい。そのため、本分析では状態遷移の割合に関する閾値を 5.0% と設定し、この閾値を満たす上位項目に対して ISM 法の適用した。

(2) 隣接行列の作成と行列計算

関係 R は、二値行列として表現できる。集合 S の各要素を行と列に持つ行列を考えると、 $s_i R s_j$ が成立していれば、行 i と列 j の交点に “1” を記入し、そうでない場合 “0” を記入する。閾値を設定した有向グラフ D に等価な二値行列 A は次の通りである。

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

この行列 A は、有向グラフ D の随伴行列 (adjacency matrix) と呼ばれる。隣接行列 A のべき乗 $A^2, A^3 \dots A^m$ から直接、間接的な因果関係の全てをあらわす可到達行列 T をブール代数演算則に従い、次式により求める。

$$T = \sum_{k=1}^n A^k + I, A^m = O$$

この時、 O は零行列、 I は単位行列、 m は n 次元の正方向行列で $A^m = O$ を満たす $n-1$ より小さい自然数である。

$$T = (A + I)^2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1^* & 1^* & 1 & 1^* & 1 & 1 \\ 0 & 0 & 1^* & 1^* & 1^* & 1 & 1 & 1 \\ 0 & 0 & 1 & 1^* & 1 & 1 & 1 & 1 \\ 0 & 0 & 1^* & 1^* & 1^* & 1 & 1 & 1 \end{bmatrix}$$

ここで、*印のついた“1”は、もとの行列 $A + I$ にはなかった要素を示している。例えば、要素 s_3 から s_5 には直接的な関係はないが、要素 s_7 を通じて間接的に関係しているため、その transitive な関係を示している。従って、行列 T には、もとの行列 A における全ての直接的関係と共に、関係 R の推移性によって得られる間接的關係も同時に表現される。

(3) 構造化

行列計算で得られた可到達行列 T の各項目について行和 d と列和 r を求め、関連度 ($d+r$) および影響度 ($d-r$) を算出する。関連度は他項目に与える影響 d と他項目から受ける影響 r の和であるため、当該項目と他項目との関係の深さを示す。一方、影響度は他項目に与える影響 d より他項目から受ける影響 r を差し引いたものであるため、当該項目が他項目に与える影響の強さを示す。これらを 2次元の散布図で示し、クラスター分析により、関連度と影響度に基づく類似性を分類する。最後に隣接行列の関係に従い有向枝で連結する。

4.2 分析結果と考察

前節で詳述した ISM 法の適用により、前章の状態遷移結果に対する多階層有向グラフを獲得した(図3, 4)。図3に示すように、グラフは階層構造を持ち、上位の階層 (Level 3) が他の要素へ影響をより強く与え、下位の階層 (Level 2, 1) が他の要素から影響をより強く受けることを表現している。具体的には、全要素の中で Level 3 の Music, S-Speech, GSI, GP が要素間で

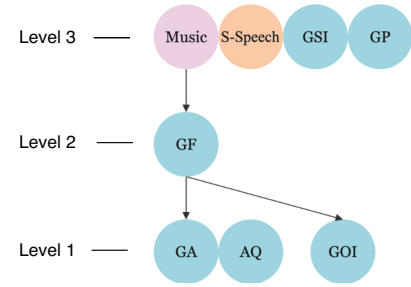


図 3: すべてのレッスンにおける多階層有向グラフ

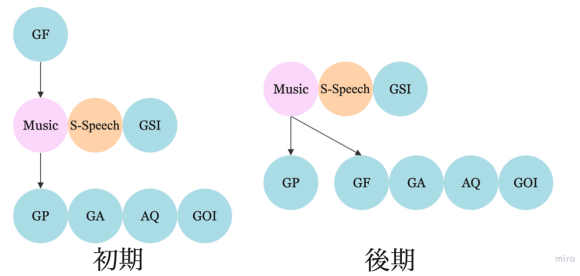


図 4: フェーズごとの多階層有向グラフ

互いに影響を及ぼし合い Level 2 の GF に影響を与え、さらに Level 2 から Level 1 の GA, AQ, GOI に影響を与えるという関係を示している。

図4のフェーズごとの構造を見ると、初期では GF が起点となり、Music、そして GP 他に影響を与えていることがわかった。そして、後期では Music が上位になり、GF は影響を受ける要素に変化した。また、Music, S-Speech, GSI の3つの要素は常に影響を及ぼし合うことが示された。これらは、前章の状態遷移結果から推察した結果と同じである。つまり、ISM 法を適用することによって、有向グラフだけでは不明瞭な要素間の関係を適切かつ直感的に構造化できることがわかった。

上記の結果を受けて、新たに作曲時代とペアによる構造化を試みた結果を図5, 6に示す。作曲時代に関しては、古典派では GSI と GP が上位であったが、ロマン派は教師の発話がすべて下位の階層に分類され、両者ともに Music と S-Speech が他の要素に影響を与えることが明らかになった。ロマン派の楽曲の場合は、より演奏中心のレッスンになることを示唆している。ペアによる比較では、3ペアとも類似しており、T1-S1 と T2-S3 においては同じ構造であった。これらの特徴として、S-Speech あるいは GP が起点となりレッスンが展開されることが示唆されるが、基本的にほとんどの要素が同一階層だったことから、ペアによる特徴的な傾向はないと言える。今後、データ数やバリエーションを増やした分析を実施することで、楽器の指導過程の定式化が期待できる。

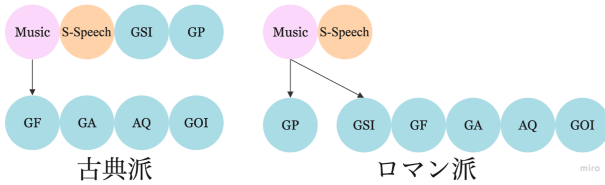


図 5: 作曲時代ごとの多階層有向グラフ

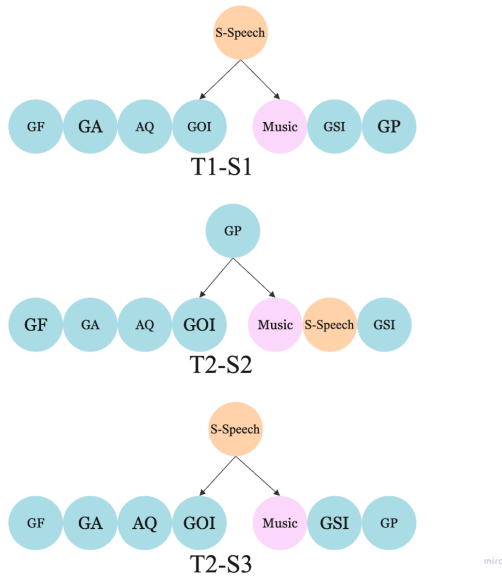


図 6: ペアごとの多階層有向グラフ

5 おわりに

本稿では、複数回にわたる楽器の指導過程の可視化を目的として、クラシックギターの個人レッスンを対象とした音声データの分析結果について報告した。収録音源から音声区間や発話情報を抽出し、さらに教師の発話に対する意味的要素を表現したラベルを付加した結果に基づいてレッスン内の状態遷移を観察したところ、レッスンのフェーズによって指導の様相が異なることが示された。また、構造化モデリング手法を用いた状態遷移結果の構造分析を試みたところ、レッスンの特徴をより直感的に表現できることがわかった。今後は、より詳細な議論を通じて、レッスンの振り返りや指導ノウハウの体系化につなげたい。

謝辞

本研究の一部は、公益財団法人柏森情報科学振興財団 K33 研 XXVI 第 598 号の助成を受けたものである。

参考文献

- [1] 三浦寛也, 飯野なみ: クラシックギターの指導現場における知識の体系化とその実現に向けた検討, 第 25 回インタラクティブ情報と可視化マイニング研究会 (2020).
- [2] Iino, N., Miura, H., Takeda, H., Hamanaka, M., and Nishimura, T.: Structural Analysis of Utterances during Guitar Instruction, *Proceedings of the 16th International Symposium on Computer Music Multidisciplinary Research*, pp. 692–695 (2023).
- [3] Daniel, R.: Exploring music instrument teaching and learning environments: video analysis as a means of elucidating process and learning outcomes, *Music Education Research*, Vol. 8, No. 2, pp. 191–215 (2006).
- [4] 山本敦, 古山宣洋, ピアノレッスンにおける演奏表現のマルチモーダルな協働的構築, *社会言語科学*, Vol. 23, No. 1, pp. 84–99 (2020).
- [5] 蓮見絵里, ジャズの即興演奏学習場面における演奏者の「音の協働探索」, *日本認知科学会第 33 回大会*, pp. 657–665 (2016).
- [6] Doukhan, D., Carrive, J., Vallet, F., Larcher, A., and Meignier, S.: An Open-Source Speaker Gender Detection Framework for Monitoring Gender Equality, *IEEE International Conference on Acoustics Speech and Signal Processing, ICASSP*, Vol. 7, pp. 103–122 (2018).
- [7] Matsubara, M., Kagawa, R., Hirano, T., and Tsuji, I.: Analysis of Usefulness of Critique Documents on Musical Performance: Toward better Instructional Document Format, *The 23rd International conference on Asia-Pacific Digital Libraries, ICADL*, pp. 344–353 (2021).
- [8] J.N.Warfield: *Societal Systems; Planning, Policy and Complexity*, John Wiley & Sons Inc (1976).

学習理論の体系に関する調査とオントロジーを活用したPBL演習のモデル化

Survey of Learning Theory and a Model of PBL Using Ontology

後藤千颯¹ 林宏樹¹ * 笹嶋宗彦¹
Chihaya Goto¹ Hiroki Hayashi¹ Munehiko Sasajima¹

¹ 兵庫県立大学
¹ University of Hyogo

Abstract: 生徒の主体性を育む方法の一つとして、グループでのPBL (Project Based Learning) を主体とした教育方法が注目されている。PBLの教育効果を客観的に評価するためには、PBLで行われている教育を教育理論に従ってモデル化し、適切な指標について計測する必要がある。本稿では、教育理論を体系化したオントロジーであるOMNIBUSオントロジーを参考に教育理論を分類し、兵庫県立大学社会情報科学部で行われているPBL演習を例として、学習理論の組み合わせとしてモデル化する方法について述べる。

1 はじめに

グローバル化、情報化、少子高齢化などの社会の急激な変化によって、社会の仕組みが大きく変容し、これまでの価値観が根本的に見直されつつある。2012年中央教育審議会では「新たな未来を築くための大学教育の質的変換に向けて」の答申において、学士課程教育の質的な変換が示され、想定外の事態に遭遇したとき、問題発見、解決するための道筋を見定める能力の育成が求められている。そのためには、学生が受動的となる従来の知識詰め込み型の教授法ではなく、学生が能動的に学修に参加する教授法を取り入れる必要がある。

2016年に示された第5期科学技術基本計画では、「課題の発見・解決に向けた主体的・協働的な学び（いわゆるアクティブ・ラーニング）」の視点からの学習・指導方法の改善を促進する方針が出された。高等学校教育も改革が進む中、大学教育として、教育方法について検討を行わなければならない状況である。

近年、データサイエンス学部や学科が設立され、データサイエンス教育が活発となってきた。渡辺[1]は、データサイエンス教育の目的である、新たな価値創造をする実践力を養う方策の1つは、受動的に学ぶ授業形態ではなく、グループでのPBL (Project Based Learning) を主体とした教育方法であると述べている。

一般に、PBLと呼ばれる科目には、Project Based Learning (プロジェクト学習、以下本稿ではPjBL)と

Problem Based Learning (問題解決学習、以下PBL)の2つがある。溝上[2]は、PjBLとPBLをアクティブラーニングの一種と捉えて、それぞれを次のように定義している。PjBLとは「問題や問い、仮説などの立て方、問題解決に関する思考力や協働学習等の能力や態度を身につける学習」である。PBLとは「実世界で直面する問題やシナリオの解決を通して、基礎と実世界とを繋ぐ知識の習得、問題解決に関する能力や態度等を身につける学習」である。溝上の定義に従えば、2つの学習の違いは教師が学習者に提示する問題の抽象度である。PjBLでは、教師が抽象的な形で提示した問題から、学習者が具体的な課題を設定して、問題解決に向けて取り組む。PBLでは、教師が具体的な課題を提示して、生徒はその課題の解決に向けて取り組む。

しかし、これら2つの概念について明示的な分類ができていないというのが現状である。具体的には、学習目的が課題発見力を身につける授業において、担当教員が誤って、具体的な課題を提示する「問題解決型」の学習方法を行ってしまったとする。この場合、学習者は教員が本来意図した教育効果を十分に得ることができない可能性がある。教授者がPjBLとPBLの意味を混同してしまい、誤った教授法を実施することで目的とした学習効果を得ることができないことは問題である。

西城[3]は、「教育方法を選択するということは、すなわち効果的学習を通じて、どのような学習者を育成したいかという選択なのである」と述べている。価値創造力を養うためには、適した教育方法を選択しなければ、目的とした教育効果を得ることができないといえる。この問題の根源に対して著者は、PjBLとPBLの相違点である、「問題」と「課題」の用語の解釈の違い

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

によって起きていると考えられる。

本研究は、一般の大学教育科目をモデル化し、効果を測定するための方法論を確立することを目指している。教育方法を一般的にモデル化する研究としては、教育理論に関する研究が長年行われており、各科目のモデル化には、それらを利用する。本稿では、教育理論に関する先行研究をサーベイし、教育理論を用いたPBL演習のモデル化と考察を行う。

2 関連研究

学習・教授理論を組織化したものとして、林ら [4] は OMNIBUS オントロジーを構築した。この研究により、多様な学習・教授理論が、同じ基盤のもとで体系化され、比較検討することが容易になった。

オントロジーとは、対象となる世界の概念を、同じレベルにおいて分類し、体系化することである。知識を構造化することによって、物事の裏に隠された暗黙的な違いを明示できることがオントロジーの強みである [5]。

OMNIBUS オントロジーでは、各理論において暗黙的であった学習理論・教授理論の繋がりを組織化することに成功した。学習理論とは、学習者が学習するために何らかの行為を行い、その結果として学習者の状態が変化した行動のことであり、教授理論とは、教授者が学習者に対して行う教授行為のことを指す。学習観は大きく3つに分かれ、行動主義、認知主義、構成主義である。これらの違いについては3章で詳しく述べる。

しかし、OMNIBUS オントロジーは、異なる学習理論や教授理論の間の違いや関係性を明らかにすることを主な目的としているため、個々の理論の詳細については、オントロジーとして定義されておらず、根拠となる論文などを示すに留まっている。本研究の目的は、OMNIBUS オントロジーが体系化している学習理論や教授理論を用いて、実際に行われている講義をモデル化することであるが、そのためには、各理論の詳細については、出典を調査する必要がある。本研究では、OMNIBUS オントロジーの体系を参照しながら、モデル化については、オントロジーだけでなく、その出典を確認しながら行う。

3 学習理論について

本章では、OMNIBUS オントロジーで示される概念の違いや、各学習観に含まれる学習理論についてである。

OMNIBUS オントロジーでは、学習理論を学習観によって分類されたのち、各理論が示されている。学習理論において学習観とは、学ぶという行動の場合、教育の主導権は誰が保持しているかに注目している。教育の主導権が教授者の場合は「行動主義」、学習者の場合は「認知主義」、教授者と学習者の双方である場合は「構

成主義」としている。行動主義とは、教授者から教えられた知識を、学習者が練習問題を解くなどの、なにかしらの行動することによって知識が定着することを目標とする。認知主義とは、学習者は教授者から、自らの欲する知識を獲得することで、構造的知識を認知することを学習行為とする。構成主義とは、学習者は社会的参加もしくは、実践的状况に参加し、教授者との対話などから知識を構成する学習行為を行う。以下の節では、各学習観に含まれる理論についてまとめる。本論文内での学習理論の名称は OMNIBUS オントロジー内での名称に準拠している。

3.1 行動主義

本節では学習理論を行動主義の枠組みから概観する。行動主義の視点から教育理論を成しているのは、ブルームのタキソノミー、クラウドラーの学習理論、マスターラーニング、古典的条件付け、オペラント条件付けの5理論である。

3.1.1 ブルームのタキソノミー

教育心理学者の Bloom が確立した学習理論である。学習者の学習に関わる状況を5つのカテゴリに階層化することに成功した。

5つカテゴリは、「知識」「理解」「応用」「分析」「評価」の階層であり、評価に近づくほど、学習者の知識は高次であるものとした。知識は、学習者が理解することのできた知識や概念である。理解は、習得した知識から、関係する具体的な例をあげられることである。応用とは、理解に加えて、問題を解決することのできる力である。分析は、複数の知識を構造化し、物事の関連性を見出すことである。最後に評価は、分析した情報を内省することである [6]。

これにより、学習者の学習状況が、学習プロセスのどのフェーズに立っているかを明確化し、知識の構築がどれほど成されているかを客観的に評価することができるとした。

3.1.2 クラウドラーの学習理論

この理論は、既存のプログラム学習に追加して、枝分かれ構造を導入した。心理学者の Crowder が確立したプログラム学習とは、学習者それぞれの学習進度に応じて、プログラムのようにフロー化された学習内容が提供されるカリキュラムである。本理論ではプログラム学習の構造に加えて、学習者が間違った経路を進んだ場合にも、それに応じた経路で学習を進めることができるように、学習フロー上に分岐を用意したプログラムとなっている [7]。プログラム学習に枝分かれ構造を導入したことにより、学習者各々に対してパーソナライズされた学習教材を提供することができるため、学習者によって異なる学習状況に合わせて学習を進めることが可能となる。

3.1.3 マスタリーラーニング

この理論は、学習者一人ひとりの学習ペースに合わせて教材や、学習進度を変化させるようにした。教育心理学者の Bloom が確立した学習理論は、学習者一人ひとりの学習ペースに合わせて、教材や学習進度を変化させることで学習者の学びを最適化させる方法である [8]。学習目標を明らかにし、学習者が達成すべきノルマを定め、学習状況の現在地を明らかにする。ノルマを達成できない場合に与えるべき教材や指導を事前に準備する必要がある。このように、学習者各々に対してパーソナライズされた学習教材を提供するため、学習者の成長スピードに合わせた効果的な学習を進めることが可能となる。

3.1.4 古典的条件付け

生理学者である Pavlov が確立した学習理論は、人間の生理的応答を応用することで効果的な学習を進めようとした。生理的応答を及ぼす刺激と同時に、生理的応答を及ぼさない物体を何度も与えることで、生理的応答を及ぼさない物体に対しても、あたかも刺激物であるかのように行動変位をしてしまうことである。教育の現場では主に幼児者の教育に使用されるが、二瓶ら [9] が「学習心理学的知見と臨床心理実践との乖離は未だに大きい」と称しているように、教育現場においては使うことが難しいとされる教育理論である。

3.1.5 オペラント条件付け

心理学者である Skinner が確立した学習理論である。学習者の自発的な行動に対して、報酬を与えることで学習者の行動を評価し、自発的行動の生起確率の増加を目指す理論である [10]。オペラント条件づけは、能動的に強化因子を取得する。学習者が能動的に活用するため、教育現場で広く利用されており、学習者の恒常的な学習意欲を促すことが可能である。

3.2 認知主義

本節では学習理論を認知主義の枠組みから概観する。認知主義の視点から教育理論を成しているのは、ガニエの 5 分類、自己調整学習、観察学習、ACT*理論、認知的柔軟性、グレイサーの学習理論、包摂理論、Rumelhart と Norman の学習形態の 8 理論である。本稿では、次の 5 つの理論について述べる。

3.2.1 ガニエの 5 分類

教育心理学者の Gagne が確立した学習理論である。学習課題の種類を 3 つに分類し、それぞれに対応する知識の定着方法を明らかにした [11]。学習課題が、頭にまつわる課題を認知領域、体にまつわる課題を運動領域、心にまつわる課題を情意領域に分類されており、分野に応じた必要な学習方法を明確にすることで効果的な学習が可能である。

3.2.2 自己調整学習

教育心理学者である Zimmerman らが確立した学習理論である。学習者自らの目標に対して、最良の学習方法を学習者自らが考え、実行する学習方法 [12] である。自己調整学習を行うためには、学習者が習得すべき 3 つの要素が存在し、各要素が互いに影響を及ぼし合うことで自己調整学習が成立する。動機づけによって学習の目的を明確化し、学習方略を選び、自分自身の学習行動を客観的にメタ認知する。このように、自己調整学習では学習者自らに沿った最良の学習行為を行うことができるため、習得した知識を構造化しやすいといった側面がある。

3.2.3 観察学習

心理学者である Albert Bandura が確立した観察学習とは、他人の行動を観察・模倣するだけで学習が完了される [13] とした理論である。学習の強度を高めるために、注意・保持・複製・動機づけの 4 つの手順が含まれていることが求められる。ここでの「注意」とは、学ぼうとしているモデルに対して、学習者の注意の視線を払わせることである。「保持」とは、学習者の注意の視線を保持することで、モデルの行動を記憶することであり、「複製」とは、保持した記憶を模倣に移すことである。最後に、「動機づけ」とは、保持した記憶を行動に移させるためにやる気を促すことである。観察学習は他人の行動を模倣することを学習と捉えているため、演習型の講義において広く普及していることがわかる。

3.2.4 認知的柔軟性

Spiro J Rand らが確立した学習理論である。特定の環境の変化に対して考え方を柔軟に変化させることができる能力を養うこととしている [14]。学習者の外的環境の変化が起こった場合においても、認知構造を柔軟に変容することができる能力のことである。

3.2.5 包摂理論

心理学者の David Ausubel が確立した学習理論である。新しい知識を自分の認知構造に階層的に取り込むことで、迅速な学習と知識の定着を図った理論である。認知構造が盤石でなければ新しい知識を得ることができず、相関的包摂と副次的包摂の 2 種類に分類した [15]。このように、包摂理論では知識と知識の組み合わせによって学習が進められると考えたため、教育現場において広く使用されている理論である。

3.3 構成主義

本節では、学習理論を構成主義の枠組みから概観する。構成主義の視点から教育理論を成しているのは、状況学習、ピアチュータリング、社会認知学習、発生的認識論、bruner の学習理論、探求学習、進歩学習教育、構築主義、誘導発見学習の 9 理論である。本稿では、次の 5 つの理論について述べる。

3.3.1 状況学習

社会実践的な活動の中から学習を進めることを提唱した理論である。Lave と Wenger によって確立した状況学習を達成するためには、実践共同体と正統的周辺参加 (LPP) の 2 点を求められる。実践共同体とは、社会における学びを実践することができる場所のことで、正統的周辺参加とは、学習者が新参者としてコミュニティに参加することで、熟練者の行動を見よう見まねで学ぶことである。学習者は十全参加をすることで知識を完全に習得することを目標としている [16]。社会参加することで学習を行うことができるため、教育現場においては実現することが難しい理論である。

3.3.2 ピアチュータリング

学習者同士が助け合い、教え合いながら学ぶ学習理論である。生徒同士が教え合いながら学ぶことにより、教授側の生徒は知識の体系化が可能に、学習側の生徒は新たな知識を得ることができる。より学習経験のある学生が、1 人以上の学習者の学習を支援する仕様が一般的である [17]。このように、生徒同士が教え合って学ぶ形式のため、教師の教育負荷が軽いことがわかる。その一方で、生徒はある程度の知識を事前に保持している必要性が生じるため、教師が生徒の習熟度をコントロールすることが難しいといった側面がある。

3.3.3 社会認知学習

最近接領域の考えから学習を成立させる理論である。心理学者の Vygotsky が確立した最近接領域とは、学習の途中まで熟練者が学習者へ補助をすることで、学習者の発達を迅速に促し、学習者の学習が習熟するに従って、熟練者の手助け無しで問題を処理できるようになるとしている [18]。教師は生徒が問題を解決するプロセスの初期段階に介入するため、問題の意図を外れた提案がなされることが少なく、学習効果が高いことが考えられる。

3.3.4 Bruner の学習理論

学習者が学習するときは、新しい事実を発見するように学ぶべきであるとした。Bruner は学習者の自主性こそが教育の核心であると考えたため、学習者が学んだことをさらに発展させ、学習者に対して新しい知識を次々に発見させることのできる、スパイラル上のカリキュラムを構築すべきであるとしている [19]。このようなプロセスをスパイラルカリキュラムといい、大学講義での基礎・演習・発展科目の関係性はこの理論に基づいている。

3.3.5 探求学習

自ら問題を設定し、その問題を解決するための情報を収集し、分析を行うことで一定の答えにたどり着く、という一連の流れを学習と捉えた。課題を解決するためには、課題の現状を様々な方法から調査を行い、情報

の収集を行うこと。集められた情報を分析し、整理することで課題を解決するための施策の立案を行うことの 3 要素のサイクルを繰り返すことで、課題解決方法をまとめ、他者に表現することで学習が成立するとしている [20]。探求学習はビジネス問題を解決するプロセスと類似しているため、企業のインターンシップから高校教育にまで広く普及している。

3.4 分類無し

本節では、学習理論中の各学習観から分けることのできなかった教育理論を概観する。分類分けすることができなかった教育理論は経験学習、多重知能理論、ダブルループ理論の 3 理論である。

3.4.1 経験学習

学習プロセス内の経験を重要視する学習理論であり、学習といった行為は経験に基づいて行われるとした。経験を学習に変容するためには、経験・内省的観察・抽象概念化・積極的実験のプロセスを踏まえるべきであるとした。学習プロセスとして始めに、実際に何らかの経験から内省的観察を行う。さらに、内省的観察から得た知識を抽象化し、他の場面でも適用するための抽象概念化の力を養う。最後に、これまでに得られた知識を積極的に実践し、知識の有効化を図ることができるとした [21]。この理論は、経験を学習に転用することが求められるため、学習者が内省的観察をどの程度行うことができるかが重要であると考えられる。

3.4.2 多重知能理論

各個人が保持している能力を分類し、その能力にあった教育を行うべきであるとした。人間が保持することのできる能力を 8 つに分類し、各個人がいずれかの能力に対して秀でていると考えた。8 つの知能は言語学知能・論理学知能・視覚的空間知能・身体的空間知能・音楽観知能・対人関係知能・内省的知能・博物学知能である [22]。本理論は、個人が保有する能力に対して、対応する効果的な学習方法を提案することに留まっている。これより、実際の教育の現場において応用することは難しいと考えられる。

3.4.3 ダブルループ理論

既存の枠組みを超え、新しい土壌の中で新しい考え方を取り入れる学習理論である。既存の考え方から学習する理論をシングルループ学習と命名したのに対し、組織の更なる向上のために、新しい価値観を積極的に取り入れる学習方法をダブルループ学習とした [23]。この理論は、常に成長を必要とする企業やビジネスにおいて重要視される学習方法である。

4 授業のモデル化

本章では、これまでに紹介した学習理論を用いて、授業のモデル化を行う。

2022年度に行われた兵庫県立大学社会情報科学部の1年次必修科目である「PBL 演習 I」を対象とする。本講義は、PjBL の考え方に基づいた演習型のプログラムとなっている。教員から問題を与えられ、1グループ5～6人ほどの数で、データを分析することから課題を発見し、課題を解決するための施策をそれぞれが考えるカリキュラムとなっている。扱われるデータは、株式会社マクロミル様から提供して下さっている QPR(消費者購買履歴データベース)と、株式会社光洋様から提供頂いている売り上げデータであり、これらを基にデータ分析を行い、課題を発見し施策の提案を行う。

演習型の授業においては、各授業回において学習の仕方が異なることが予想される。例えば、序盤では課題の説明であったり、分析の仕方を学ぶ、いわゆる座学型の講義を行っているが、回が進むにつれ、生徒の自主性を重んじて、思考力を問うワーク型の講義を展開するように変化することが考えられる。このように、授業回によって学習者の学び方について違いがあると考えたため、笠井ら [24] の研究と同様に、「授業の展開をいくつかの場面（本稿では以下、“Step”と呼ぶ）に分割」することとした。これより、過去に行った PBL 演習 I の授業を内省し、授業を分割し、それぞれの step に対応する学習理論を当てはめることで、授業のモデル化を試みる。

授業の初回にはオリエンテーションと分析使用ツールの説明を行う。この回では、授業全体の流れの説明であったり、分析使用ツールを使う際の注意点について教員から生徒に述べられる。このように、生徒は教員の話をもとに聞くことで知識を習得しているため、行動主義内のブルームのタキソノミーが行われているといえる。しかしながら、ブルームのタキソノミーでは知識の習得後、理解・応用・分析・評価と、知識の汎化を目指さないとはいけませんが、この授業においては知識を与えているだけの状態であるため、Step1 では、ブルームのタキソノミーにおける知識部分の学習のみが行われているといえる。

次に、個人での QPR チュートリアルと課題を行う。学習者各自が、教員から与えられた分析課題に対して、教員が作成した基本利用操作資料から情報を取得しながら課題を解いていく段階である。この回では、生徒が QPR の使い方について習得することを目的に設けている。このように、課題に対して必要な情報が記載されている場所を学習者が探し、教授者の実演資料を模倣する形で問題を解決しているため、Step2 では、認知主義内の観察学習が行われているといえる。

QPR のチュートリアルが終了すると、グループ課題

の公表と、課題の進め方の説明が行われる。これはオリエンテーションと同様に、教員が課題の説明を行い、生徒は教員の話をもとに聞くことで知識の習得が行われている。これより、Step3 では、ブルームのタキソノミーにおける知識部分のみの学習である。

グループ課題では始めに、抽象的な問題を与えられ、抽象的な問題から課題を各自設定し、課題を解決するための施策について検討する。課題設定は生徒主導で行われるが、分析初等者である受講生は、抽象的な問題から課題を巧みに設定することを困難に感じる事が考えられる。こういった課題から、教員が生徒の課題設定を途中まで補助する形でサポートをしている。このような、生徒と教員の対話から学習をサポートする学習理論は、構成主義内の社会認知学習であるといえる。よって Step4 は、社会認知学習である。

課題設定が終わると、課題を解決するために、データの裏付けに基づいた施策の提案を行う。この回では、生徒が購買データを自分なりの視点から比較検討し、各グループで決定した課題に対して、必要となる情報を取捨選択し、課題を解決する施策について考える。ここで、班員によってデータの見方は異なるため、それぞれの考え方についてグループの中で教え合って学ぶことを行っている。これより Step5 では、構成主義内のピアチュータリングが行われていると考えた。

兵庫県立大学の PBL 演習 I では、課題解決を手助けする方法の1つとして、対象店舗の現状の視察を兼ねた、実店舗見学を行っている。実店舗見学では、現場社員の声を直接聞くことや、現在の商品の陳列方法について調べることができる。このように、経験を通して知識を学び、獲得した知識を構造的にまとめ、課題解決に使える形に変化させることで知識の汎用化を目指す学習を行っている。よって Step6 は、認知主義の Rumelhart と Norman の学習形態が行われていると考えた。

この授業では、全体講義の最後に各グループの施策について発表を行う。各班が決めた課題に対しての施策を発表できる形にまとめ、班ごとに有識者や受講生の前で発表するプログラムが最後に用意されている。受講生は他のグループの施策について聞き、自らの案との違いであったり、質問することによって知識を習得している。このように、生徒同士で教え合う学習プロセスは、構成主義のピアチュータリングの学習理論である。

5 結果と考察

以上の PBL 演習 I の授業をモデル化したものを以下の表 1 に示す。

表 1 の結果より、兵庫県立大学社会情報科学部の PBL 演習 I の授業は、7つの授業 Step から構成されており、2つの行動主義、2つの認知主義、3つの構成主義から構成されていることが確認できた。その中でも、他の理

論と比べて、構成主義の考え方から学習を行っている回数が多いことが確認できた。つまり、PBL 演習 I では、人と人との社会的な対話によって学習行為を進める、といったことを重要視している、と推測を立てることができた。

表 1: PBL 演習 I の学習理論を用いたモデル化

Step	授業内容	学習理論
1	オリエンテーション・分析使用ツールの説明	ブルームのタキソノミーの知識部分
2	個人での QPR チュートリアルと課題	観察学習
3	グループ課題解説	ブルームのタキソノミーの知識部分
4	グループでの課題設定	社会認知学習
5	課題取り組み	ピアチュータリング
6	実店舗見学	Rumelhart と Norman の学習形態
7	施策の発表	ピアチュータリング

これより、PBL 演習 I に代表される PjBL の一例として、3つの教育間をバランスよく配置している中でも、構成主義の割合をやや多く取り入れることで、授業を構成しているという特徴を持つということがわかった。

本研究での授業のモデル化においては、実際の授業内容を内省することで、学習者の学習の変化が起きたとされる場面を Step 化し、それぞれの授業 step において学習理論を当てはめた。よって、どのような授業が行われていた時に、どのような教育理論が当てはまるのかという、モデルの一般性については今後の展望とする。

6 おわりに

本研究では、OMNIBUS オントロジーを参考に学習理論をまとめ、兵庫県立大学社会情報科学部の PBL 演習 I の授業を、学習理論の集合としてモデル化することを提案した。本研究の特長として、既存の授業を教育理論の観点からモデル化することで、各授業 Step において、よりよい学習を行うための提案を教育理論の考え方から行うことができる。さらに、教員がすべきアプローチについても、教育理論を用いることで可視化することができるため、授業の改善に役立てることができると考えた。

今後は、OMNIBUS オントロジー内の教授理論についても本研究と同様にまとめること。そして、様々な分野の PjBL と PBL の授業をモデル化することで、両者の間に存在する授業構成の違いについて可視化することを目指す。

参考文献

[1] 渡辺 美智子 (2022): AI 社会におけるデータサイエンス教育～データと ICT を活用した問題解決力の育成～, 工学教育, 70 巻, 1 号, p.30-35.
 [2] 溝上 慎一 (2016): アクティブ・ラーニングとしての PBL と探究的な学習, 東信堂.

[3] 西城 卓也 (2013): 医学教育における効果的な教授法と意味のある学習方法①, 医学教育, 44 巻, 3 号, p.133-141.
 [4] 林 雄介, Jacqueline Bourdeau, 溝口理一郎 (2009): 理論の組織化とその利用への内容指向アプローチ - オントロジー工学による学習・教授理論の組織化と Theory-aware オーサリング-, 人工知能学会, 24 巻, 5 号, A, p.351-375.
 [5] 溝口理一郎, 古崎晃司, 來村 徳信, 笹島 宗彦 (2006): オントロジー構築入門, オーム社.
 [6] Bloom, B.S.(1956): Taxonomy of Educational Objectives, Handbook 1, Cognitive Domain. New York, David McKay.
 [7] Owen, R.S., & Aworwu, B. (2011): Contemporary Instructional Design. In Instructional Design: Concepts, Methodologies, Tools and Applications (p.95-100). IGI Global.
 [8] Block, J.H., & Airasian, P.W., & Carroll, J.B., & Bloom, B.S. (1971): Mastery learning: Theory and practice. Holt, Rinehart and Winston.
 [9] 二瓶 正登, 田中恒彦, 澤幸祐 (2019): 不安と関連する障害における古典的条件づけの役割と意義 - 古典的条件づけの諸現象と連合学習理論の臨床的应用-, 不安症研究, 11 巻, 1 号, p.13-23.
 [10] Staddon, J.E., & Cerutti, D.T.(2003): Operant conditioning. Annual review of psychology, 54(1), p.115-144.
 [11] Gagne, R. M. (1984): Learning outcomes and their effects: Useful categories of human performance. American psychologist, 39(4), p.377-385.
 [12] Boekaerts, M. (1999): Self-regulated learning: Where we are today. International journal of educational research, 31(6), p.445-457.
 [13] Boeree, C.G.(2006): Albert Bandura. Personality theories. (Retrieved 24 October, 2023 from <http://webspaceship.edu/cgboer/bandura.html>).
 [14] Spiro, R. J. (1988). Cognitive Flexibility Theory: Advanced Knowledge Acquisition in Ill-Structured Domains. Technical Report No.441.
 [15] Adhikari, K. (2020): Ausubel's learning Theory: Implications on Mathematics Teaching. Research Gate.
 [16] 石山恒貴 (2023), 特集, 現在の労働時間を考える上で改めて読んでおきたい文献: レイブ=ウエンガー『状況に埋め込まれた学習-正統的周辺参加』, 日本労働研究雑誌, p.54-57.
 [17] Mynard, J., & Almarzouqi, M.(2006): Investigating peer tutoring. ELT Jornal, 60(1), p.13-22.
 [18] Mcleod, S.A.(2015): Lev Vygotsky. Retrieved from www.simplypsychology.org/piaget.html.
 [19] Bruner, J. S. (1966): Toward a theory of instruction. Harvard University Press.
 [20] Pedaste, M., & Mäeots, M., & Siiman, L.A., & et. al,(2015): Phases of inquiry-based learning: Definitions and the inquiry cycle. Educational Research Review, 14, 47-61. <https://doi.org/10.1016/j.edurev.2015.02.003>
 [21] Kolb, D.A.(1984): Experiential Learning: Experience as the Source of Learning and Development. New Jersey, Prentice Hall.
 [22] Gardner, H.(1987): The theory of multiple intelligences. Annals of dyslexia, 37, p.19-35.
 [23] Argyris, C.(1977): Double loop learning in organizations. Harvard business review, 55(5), p.115-125.
 [24] 笠井俊信, 永野和男, 溝口理一郎 (2015): 教師の授業設計意図自動化システムの開発とその有効性評価, 人工知能学会, 30 巻, 3 号, p.570-584.

確率的経路生成モデルの規格化条件の導出と実験による評価

Derivation of Normalization Conditions for Probabilistic Route Generative Model and Experimental Evaluation

柴田祐樹^{1*} 高間康史¹
Hiroki Shibata¹ Yasufumi Takama¹

¹ 東京都立大学大学院システムデザイン研究科
¹ Graduate School of System Design, Tokyo Metropolitan University

Abstract: This paper derives normalization conditions for the probabilistic route generative model (PRGM). Variety of uncertainty must be dealt with in the sightseeing route recommendation. PRGM is appropriate as the model of it because PRGM intrinsically expresses the uncertainty, and can employ well studied methods such from statistical physics. After the proposal of PRGM, its formulation still lacks generality, i.e., each node in a graph of road map must be visited only once or less than once. It is extended so that any node can be visited infinitely and shows conditions the model satisfies mathematically and its validity by numerical experiments.

1 はじめに

本稿では、確率的経路生成モデルの確率分布が規格化されるための目的関数が持つべき条件の導出を行う。また評価実験によりその妥当性を示す。本研究はインタラクティブ観光経路推薦システムに関するものである。

インタラクティブ観光経路推薦のための確率的経路生成モデルが提案されている [1]。文献 [1] は、不確実性の多い状況で利用が想定される観光経路推薦システムに対して、確率モデルによる推薦問題の定式化の有用性を主張している。一般のグラフ上での順回路（経路）を扱い、巡回 Salesman 問題（Traveling Salesman Problem）[9]と同様に一つ経路で各節点は一度のみ通過可能であるとしている。本研究では、このモデルを拡張し、何度でも節点を通過可能であるとする。この仮定は、もと来た道を引き戻さなくてはならない事態の発生しうる、インタラクティブな経路推薦を行う上で必要となる。文献 [1] はインタラクティブに実際の状況に応じて経路を再構成することを想定しているモデルであるが、この仮定を満たしていないため、起こりうる経路の経路長さが有界でなくなるといった問題が発生する。より形式的に言えば、経路の状態空間の大きさが [1] で仮定されていた有限で

はなく無限になるという問題がある。無限の状態空間に対して定められる確率分布は、各標本に対する確率の和が 1 となることが自明ではないため、規格化の条件を別途定める必要がある。本研究はこの問題に取り組む。

不確実性とは、天気の変化、交通網上で事故、予定の変更、移動速度のばらつきなどの観光を行う上で発生しうる要因のことである。観光経路推薦における不確実性を取りあつかう方法は、[1] のような確率分布による方法以外にも数多く研究されている [2,3]。扱う不確実性の粒度はそれぞれの研究で異なるが、これらはどれも経路を求める問題を TSP、および Selective TSP (STSP) [9]をもとにした最適化問題として定式化し、最適な経路を利用者へ提示することを想定している。実際の不確実性への対処方法として、[3] はユーザからのフィードバックを得られるたびに再計算を実行するシステムを提案している。[2] は問題の定式化に着目している一方で実時間での状況変化へ対処法については言及がされていない。このように不確実性を適切に取り扱う試みが最適化問題の枠組みでなされているが、依然としてさらなる一般化の余地が残されていると言える。

[1] は経路の確率生成モデルを観光経路推薦に用いることを提案している。確率分布を用いることで、不確実性はモデルに内在するため、追加の定義なしに表現することが可能となる。また、近年統計物理学の発展により得られている、Markov Chain Monte Carlo (MCMC) 法 [7] や、Exchange Monte Carlo 法 (EMC) [6, 10] と

* 連絡先：東京都立大学システムデザイン学部情報科学科
〒191-0065 東京都日野市旭が丘 6-6
E-mail: hshibata@tmu.ac.jp

言った多くの先進的な標本抽出法を用いることができる。Bayes 推論の枠組みを適用可能であることから、先端の機械学習の成果を取り込むことが容易である。確率モデルの標本抽出法は単一解の最適化法に比べると計算量が大きいので、従来は実問題への適用が難しくあったが、近年の計算機の発展により、実用化の可能性が見られる。これらの背景から、[1] の提案は近年の背景に即したモデル化となっていると考える。このモデルを発展させることで、今後の経路推薦問題の定式化と推薦システムの性能向上が期待できるため、本研究は [1] のモデル化に着目する。具体的に、本稿では節点の訪問回数に制限をなくす一般化を行う。

本稿は上記一般化を行う際に必要となる確率分布を規定する目的関数が満たすべき規格化条件の導出を行う。また、実際にその規格化条件を満たす目的関数を複数提案し、評価実験によりそれぞれの特性を調査する。評価実験から、規格化条件を満たす目的関数を用いた確率分布からは、経路長さの期待値が収束し、そうでない場合は収束しないことを示す。結果について理論から想定されないことに関し考察を行う。インタラクティブな推薦手法をモデルの性質から考慮し、確率的経路生成モデルの定式化を行う研究は、著者の知る限り存在しないため、本研究の創造性は高いと考える。

1.1 定義と記法

\mathbb{R}_+ を正の実数、 \mathbb{N} を自然数全体の集合とする。自然数 n に対して、 $[n] = \{1, 2, \dots, n\}$ とする。集合を $\{ \}$ で $\{1, 2, \dots\}$ のように書き、族 (列) を $(1, 2, \dots)$ のように $()$ で表す。集合 A と自然数 N に対して、 A^N は直積集合を表し、その要素 $a = (\in A^N)$ は $\forall i \in N [a_i \in A]$ を満たす列である。例えば $x \in \mathbb{R}^{[n]}$ は $x = (x_1, x_2, \dots, x_n)$ で表される数ベクトル x の全体を意味する。 $\#A$ は A の要素の数を示す。標本空間を Ω で表す。確率変数 \mathbf{x} が与えられたとき、 $\mathbf{x}(\Omega)$ は \mathbf{x} の実現値の集合 (状態空間) を示す。 \mathbf{x} の確率分布を $p(\mathbf{x})$ 、特定の値 x に対する確率を $p(\mathbf{x} = x)$ と表す。この他の記法について [13] に従う。

2 関連研究

2.1 Traveling Salesman Problem

Traveling Salesman Problem (TSP) [9] は観光経路推薦問題の基礎的な定式化を提供する。TSP は最短順回路を求める問題であるが、これと類似の定式化で経路長さ制限が存在する元で、線形和で表される順回路の

利得 (価値) を最大化するという問題の Selective TSP (STSP) が実際には良く用いられる [1, 5, 9]。これについて説明する。 N を節点の集合、 E を辺の集合として、完全グラフ $G = (N, E)$ を考える。STSP は各 $n \in N$ に価値を割り当てる $b: N \rightarrow \mathbb{R}$ を定義する。これは観光経路における、スポット (Point of Interest) の価値に相当する。経路を $x \in N^{\mathbb{N}}$ で表し $L_x (\leq \#N)$ を x の長さであるとする*1。 $w: E \rightarrow \mathbb{R}_+$ を辺の重み、 $f: N \rightarrow \mathbb{R}_+$ を節点での滞在時間とする。ある $C \in \mathbb{R}_+$ に対して、以下の P_1, P_2, ϕ を定義する。

$$\phi(x) = - \sum_{i=1}^{L_x} b(x_i), P_1(x) : \Leftrightarrow i \neq j \Rightarrow x_i \neq x_j$$

$$P_2(x) : \Leftrightarrow \sum_{i=2}^{L_x} [w(x_{i-1}, x_i) + f(x_i)] < C, \quad (1)$$

このとき、STSP は、

$$P(x^*) \wedge \phi(x^*) = \min_{P_1(x) \wedge P_2(x)} \phi(x)$$

となる x^* を求める問題と定式化される*2。

C は時間制約 (time budget) などと呼ばれ、全体の旅行にかかることのできる総時間である。この定式化では問題のグラフは完全グラフでなくてはならない。そのため、地図情報を用いるには前処理を要する。多くの研究では、地図中のうちスポットに該当する地点を、最大で 100 程度選び、それらスポットで N を定義し b, f を設定する。また、スポット間を結ぶ最短経路を w として設定する。 $\#N \approx 100$ 程度である理由は主に計算負荷の削減のためである。近年ではより高性能な計算機が利用可能となっていることから、[1] ではその計算能力を活用可能な $\#N > 1000$ の場合を想定したモデル化と計算手法を提案している。これについて節 2.3 で後述する。

2.2 最適化法と標本抽出法

節 1 で述べたとおり、本稿が研究対象とする観光経路推薦問題は、ほとんどの研究で最適化問題として定式化される。例外として [1] は標本抽出による確率分布の近似問題として定式化する。本稿の研究と関連する最適化手法と [1] で用いられる標本抽出法をここで紹介する。

1 循環定義を避けるために経路は無限列 $N^{\mathbb{N}}$ の要素であるとしている。これに形式的に長さを定めるためには、 N に実在しない特殊な節点 n^ を用意し、 $n^* = x_i$ となった i を L_x とすれば良い。

*2 普通 $\phi(x)$ の符号は本稿と逆であり \max により定義されるが、本稿の他の定式化と一貫性を取るために、 \min による定義を用いている。

観光経路推薦問題は、Non-deterministic Polynomial Hard (NP 困難) に属するため、最適化手法は主に局所最適化と Meta-Heuristics (HM) を元に近似解法として構成される。よく用いられるのが、k 最適化法 [11] と Genetic Algorithm (GA) や Simulated Annealing (SA) といった HM の組み合わせである。また、GA や SA ほどの認知度が無いと見られるが、近年発展の目覚ましいものに、Exchange Monte Carlo (EMC, Parallel Tempering とも言う) が挙げられ、TSP に対する適用事例 [6] や、STSP に対する適用事例 [1] が存在する。

k-最適化法は、最適化途中の経路 x の k 個の辺 $((x_i, x_{i+1}; i \in K), \#K = k, K \subset N$ を選び、これらについてのみ式 (1) の ϕ を最小化するように最適化を行うことを集合 K を変更しながら繰り返す。これは貪欲法であり、容易に局所最適解へ早期収束する問題があるため、MH と組み合わせて通常用いられる。

SA は目的関数 ϕ を指数部に持つ、次の Boltzmann 分布を元に構成される。

$$p(\mathbf{x} = x) = Z(T)^{-1} \exp(-T^{-1}\phi(x)) \quad (2)$$

ここで、 $\mathbf{x} : \Omega \rightarrow N^N$ は経路の確率変数、 Z は確率分布を規格化するための定数、 $T (\in \mathbb{R}_+)$ は温度と呼ばれる分布の平滑さを制御する制御変数である。式 (2) より、 $T \rightarrow 0$ の極限において、

$$\phi(x) = \min_y \phi(y) \Rightarrow p(\mathbf{x} = x) = \frac{1}{\#\text{argmin}_y \phi(y)},$$

$$\phi(x) \neq \min_y \phi(y) \Rightarrow p(\mathbf{x} = x) = 0$$

となる。極限の分布から標本抽出を行うことは困難であるため、MCMC 法などの平衡分布に標本が近づく手法により標本抽出を繰り返しながら、徐々に T を 0 へ近づけ、極限分布から標本、すなわち最適解を抽出することを期待する方法である。最適解が抽出できない場合でも、無作為に選んだ場合に比べれば最適解に近い解を得られることが期待できる。ここで、MCMC では、提案分布と呼ばれる状態の無作為な変更方法が定義される必要があるが、ここに k-Opt 法を用いることができる。ただし、k-Opt 法で ϕ を最小化する変更は必ずしも適用されず、MCMC の棄却基準により適用の選択がされる。詳細は省くが T が小さいほど選択基準は貪欲的になり、逆の場合探索的になる。

SA では T の変化方法の定義に自由度があり、この点に問題がある。一度設定した T の降下計画が終われば、そこで最適化が終わってしまうからである。これは、最適化に使える時間を見積もることが難しい、例えば real-time 性の高い場面で問題となる。Exchange Monte

Carlo 法 (EMC) [6, 10] はこの温度計画を不要にし、SA と同様に低い温度の分布から近似最適解を抽出可能な性質を保ちつつ、定常的に最適化を続けることができる方法である。さらに、EMC は任意の温度に対する $p(\mathbf{x})$ を近似する標本を抽出することができる。これにより、確率分布を近似することが可能となる。MCMC でも確率分布は近似できるが、多峰性の分布に対して性能が低い。一方で EMC は高い T を持つ分布を利用することで、高い T の分布は分布の山が平坦になるため MCMC での遷移が容易となることから、他の低い T の分布を精度良く近似できるという特徴を持つ。

EMC について少し定式化を説明する (詳しくは [1] の節 3 を参照されよ.)。EMC では、SA の温度計画の代わりに温度の列 $(T_m; m \in M)$ を用意する。これらに対して、 T のみが異なる式 (2) と同様の分布 $p(\mathbf{x}_m; T_m)$ に対して MCMC 法で標本抽出を繰り返す。これら分布を複製 (Replica) と呼ぶ。得られた標本を、任意の間隔で、2 つの複製間で交換し MCMC を継続する。この交換は MCMC と同様の基準で確率的に行われ、常に発生するとは限らない。この基準による確率的交換により、それぞれの $p(\mathbf{x}_m)$ の分布に悪影響を与えること無く、その標本の精度を向上させることができる。

2.3 インタラクティブ観光経路推薦

インタラクティブな観光経路推薦手法の提案が行われている [1-3]。ここでは、確率分布から標本を抽出することによりインタラクティブ観光経路推薦問題を定式化する [1] に着目する。着目する理由は節 1 で述べたとおり、[1] は機械学習分野や統計物理学の手法を適用可能な程度に一般性が高く発展させる価値があるからである。[1] の定式化を Probabilistic Route Generative Model (PRGM) と本稿では呼ぶ。

PRGM の定式化を説明する。PRGM では目的関数 $\phi : N^N \rightarrow \mathbb{R}$ を以下の通り定義する。この関数の値が小さいほど良い経路であるとする。

$$\phi(x) = F(\alpha(x) - C) - \beta(x), \quad (3)$$

$$\alpha(x) = \sum_{i=2}^{L_x} w(x_{i-1}, x_i), \beta(x) = \sum_{i=2}^{L_x} b(x_{i-1}, x_i)$$

ここで、 $F : N^N \rightarrow \mathbb{R}$ は制約関数であり STSP の制約を表現する。実際に $F(x) = \lim_{a \rightarrow \infty} \exp(ax)$ とすれば、この目的関数を最小化する問題には STSP の制約が含まれる。その他の制約については、式 (1) の $P_1(x)$ を満たすものとする、つまり、同じ節点を複数回訪れる経路は生成されない。

$w : N^2 \rightarrow \mathbb{R}_+, b : N^2 \rightarrow \mathbb{R}$ はそれぞれ辺の負荷と価値である。これは完全グラフを前提とした定式化であるが、PRGM では完全グラフでない一般のグラフに対応可能とするために、仮想辺と呼ばれる辺が導入される。仮想辺は、MCMC の状態変化において経由されることで、非完全グラフの E_0 に対する局所変化を完全グラフに対する方法と同様に適用可能とするものである。元のグラフを N, E_0 として、 $e \in E_0$ を実辺、 $e \notin E_0$ を仮想辺と呼ぶ。このとき、 $e \in N^2$ に対し、次のとおり w, b を定義する。

$$\begin{aligned} e \in E &\Rightarrow [w(e) = w^0(e), b(e) = b^0(e)], \\ e \notin E &\Rightarrow [w(e) = s(e), b(e) = -Bs(e)], \end{aligned}$$

ここで、 w^0 は実辺の負荷であり、 $b^0(e)$ は実辺の価値である*3。 s は仮想辺の負荷であり、 $e = (n, n')$ としたとき以下の通り定義される。

$$\begin{aligned} s(n, n') &= \min_{x' \in N^N} \sum_{i=1}^{L_{x'}} w^0(x'_i, x'_{i+1}), x'_1 = n, \\ x_{L_{x'}} &= n', \forall i \in [L_{x'} - 1] (x_i, x_{i+1}) \in E_0. \end{aligned} \quad (4)$$

$B \in \mathbb{R}_+$ は罰則係数であり、ハイパーパラメータとなる。 $e \notin E_0$ の場合、 e は実際の道に対応しない辺であり、最終的に生成される経路に含まれてはならない。 $Bs(e)$ の値を大きくすることで、この辺を持つ経路 x の $\phi(x)$ は比較的高い値を持つため、生成されにくくなる。よって、 $Bs(e)$ は大きな値となるように $B, s(e)$ を定めるべきであるが、MCMC の効率を低下させることになるため際限なく大きな値とすることはできない。PRGM ではこれに対し式 (4) のように $e \notin E_0$ に対しての E_0 上の辺で構成される最短経路の負荷で $s(e)$ を定義する。これにより $\forall e \in E_0 b^0(e) = 0$ のグラフに限るが、遷移が効率的に行われ、貪欲法であっても最終的に経路から仮想辺が消滅することが示されている。

実際の経路生成では、目的関数 ϕ を指数部に持つ式 (2) により確率分布を定義し、EMC により標本抽出を行う。標本平均により描かれた経路の確率分布をユーザに提示する。ユーザが確率分布の値の高さをその道の推薦度とみなし、経路を自ら描くという用途を期待する。単一の経路を一つ提示する従来の最適化に基づく推薦に比べ、ユーザの意志を不必要に制限する必要がなくなり、また、EMC によりユーザが選択した以外の経路を、ユーザの選択を制約として、持続的に再計算が可能であ

るため、インタラクティブ性に優れることが主張されている。

制約関数 F について STSP に対するもの他には、 $F(x) = \text{ReLU}(x)$ などがある。PRMG では ReLU と同等関数により急激な目的関数の変化を防ぎ、最適化を容易にしているとある。しかし、その効果について実証はされていない。次節以降 F について議論する。

3 提案

本稿は PRGM [1] の定式化のうち式 (1) の P_1 の制約、つまり経路長さの制限をなくす一般化を行う。PRGM の定式化では、来た道に戻ることができず、その状況が起こった場合に別途処理が必要であるという問題があった。実際にそういう場合は、実時間でインタラクティブな操作を想定したときユーザの行動が予想できないため十分に考えられる。よって、一般化を行うが、このとき確率分布が規格化されることが自明でなくなる。式 (2) の分母 Z の規格化定数は、分子の x の状態空間 $\mathbf{x}(\Omega)$ についての総和であり、この状態空間の大きさが無限となるとき、一般に総和が有限の値とはならないからである。

3.1 規格化条件の導出

経路が定義されるグラフとして完全グラフを仮定する。この仮定は議論の一般性を失わない。また頂点の数は有限で M とする。 $M < \infty$ であるが、このグラフ上で生成される経路長は同じ節点を通過することが際限なく可能であるため ∞ となりうる。重みは、場合の数が最も大きくなり、最も厳しい場合の議論ができるようにすべて 1 であると仮定する。式 (3) における価値は 0 とする。

経路 \mathbf{x} の長さ (重みの合計) を $\alpha(\mathbf{x})$ と表す。 $\alpha(\mathbf{x}) = n$ である場合の数は、ほぼ M^{n-1} である (始点と終点固定される分の自由度が減るため M^n ではない)。今、目的関数は (3) の定義から経路の長さ n のみに依存するため、これを ϕ_n で表す。このとき、

$$\begin{aligned} p(\alpha(\mathbf{x}) > k) &\sim \sum_{n=k+1}^{\infty} M^{n-1} \frac{1}{Z} \exp\left(-\frac{1}{T} \phi_n\right) \\ &\leq \sum_{n=k}^{\infty} \exp\left(n \log M - \frac{1}{T} \phi_n\right) \end{aligned}$$

であるが、 p は確率分布であるから任意の k について

$$p(\alpha(\mathbf{x}) > k) \leq 1 \quad (5)$$

*3 辺に価値が割り当てられておりこれは STSP の定式化である節点に価値を割り当てる定義と異なるが、ここでは簡素化のため、また PRGM と同様の定義とするためにこの定義を用いる。

でなければならない。このための必要十分条件は、

$$\exists C \in \mathbb{R}, \sum_{n=1}^{\infty} \exp\left(n \log M - \frac{1}{T} \phi_n\right) < C \quad (6)$$

である。また式 (6) の必要条件は、

$$\lim_{n \rightarrow \infty} \exp\left(n \log M - \frac{1}{T} \phi_n\right) = 0 \quad (7)$$

である。よって、任意の $M, T (\in \mathbb{R}_+)$ について成立させるためには、

$$O(n) \subset O(\phi_n) \wedge O(n) \neq O(\phi_n) \quad (8)$$

となることが式 (5) の必要条件である。例えば、 $\phi_n = n^2$ などは式 (8) を満たす。またこれは、式 (6) も成立させるため、 $\phi_n = n^2$ は式 (5) の十分条件となる。

3.2 目的関数の提案

式 (8) を満たすものとして、前節最後の n^2 の他には、ある $a > 0$ に対する式 (9) や式 (10) がある。

$$F(x) = x^{1+a} \quad (9)$$

$$F(x) = x \log(x+1) \quad (10)$$

上記はさらに、式 (6) も満たすため、確率分布が規格化されるための十分条件式 (5) を満たす。

ここで、式 (7) には温度 T があるため、これに依存して 0 への収束速度が決まると考えられる。収束速度が小さい場合、より長い経路が生成されやすくなるが、長い経路は計算負荷が高いため問題となる。これについて検証する。 $\phi_x = F(x)$ を代入して、指数部が負となる場所を考える。式 (9) の場合、指数部は、

$$x \log M - \frac{1}{T} x^{1+a}$$

となるが、これが極大値を取る $x = x_*$ は、

$$x_* = \left(\frac{T \log M}{1+a}\right)^{\frac{1}{a}} \quad (11)$$

である。この値より $x > x_*$ の領域で確率の値は x の増加に伴い減少を始めるため、経路長さの期待値は x_* 付近となると考えられる。実問題において M には実効的に $O(1)$ 程度になると考える。

4 評価実験

式 (9) と式 (10) を用いた場合に確率分布が規格化されている、つまり経路長さの期待値が収束するかどうか

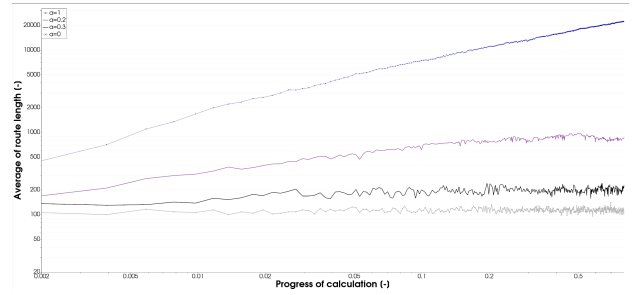


図1 Comparison of convergence of route length on maximum temperature $T = 4$ for each, x^{1+a} , $a = 0.0, 0.2, 0.3, 1.0$.

か検証する。またその期待値が (11) で予測されるとおりとなるか調べる。数値実験では [1] と同様の 24×24 の格子状グラフを用い、 $e \in E_0$ について $w^0(e) = 1$ とし $b^0(e)$ は $0, 0.25, 0.5$ の中から無作為に設定した。 $\#N = 576, \#E_0 = 1104, C = 96, B = 1$ である。EMC を用いて標本抽出を行い、 $T_m = 0.1d \exp(m/(M-1)0.9d)$, $m \in [128]$ の温度を持つ複製を用いた。 $d = 3, 4$ を図 1, 図 2 の実験にそれぞれ用いた。MCMC を 2^{10} 回行うごとに EMC の状態交換を行った。これを 2^{10} 回行い、2 回ごとに統計量の計算に用いる標本を記録した。乱数生成に Mersenne Twister [12]、実数の近似に 64 bit 浮動小数点数を用いた。

図 1 に経路長さとの計算の進捗の関係を示す $F(x) = x^{1+a}$, $a = 0.0, 0.2, 0.3, 1.0$ について両対数軸により示す。縦軸は $T = 4$ の状態の経路長さである。 $a = 0$ の場合を除いて収束している。これにより、式 (6) が十分条件として妥当であるとわかる。 $a = 0$ の結果について、これは典型的な Random walk の状態であるため、取りうる状態に制限が与えられていないと推測できる。すなわち、確率が規格化されておらず、(8) が必要条件として妥当であるとわかる。

図 2 に目的関数の違いによる平均経路長さとの温度の関係を示す $F(x) = x^{1+a}$, $a = 0.08, 0.1, 0.12, 0.14, 0.2, 0.3, 0.5, 1$ および $F(x) = x \log(x+1)$ の場合について両対数軸により示す。図 2 から、高い T で経路が長くなる傾向はどの場合も認められ、 a の値が小さいほど、大きいとわかる。 $T = 1$ を堺にして増加が始まり、 a に依存せずすべて同程度の増加速度となっている。式 (11) では、 a の値により増加率が変わるため、この結果は式 (11) の予想に反する。 $T = 1$ を堺に増加が始まっていることについて、実験で用いたグラフの値 ≈ 0.5 が影響していると考えられる。 T が値を超えたときに状態が仮想辺を通過することが可能となり始めるため、このような遷移が現れていると考え

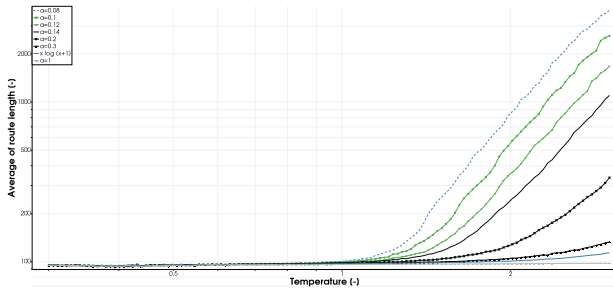


図2 Comparison on various constraint function, x^{1+a} , $a = 0.08, 0.1, 0.12, 0.14, 0.2, 0.3, 0.5, 1$ and $x \log(x + 1)$.

られる。 a による増加率が予想に反することについて、これも価値が影響を及ぼしていると考えられる。今回一般的な条件とするために価値を与えたグラフを実験で用いたが、価値を与えないグラフを用いた実験を行った場合に違った傾向が観測されると考える。

5 おわりに

本稿では、確率的経路生成モデルの確率分布が規格化されるための目的関数が持つべき条件の導出を行い、また評価実験によりその妥当性を検証した。数値実験より、導出した条件の妥当性を示した。一方で、予測される確率分布の裾の広がり、実験で示される裾の広がり一致しなかった。この点についてさらなる調査が望まれる。

インタラクティブな推薦手法をモデルの性質から考慮し、また、確率的経路生成モデルの定式化を行う研究は、著者の知る限り存在しないため、本研究の創造性は高いと考える。

謝辞

本研究の一部は JSPS 科研費 JP22H03698 の助成を受けたものである。

参考文献

[1] 柴田祐樹, 高間康史: インタラクティブ観光経路推薦手法の構築に対する検討, 第 27 回 SIGA, pp. 19–24 (2021).
 [2] Hasuike, T., Katagiri, H., Tsubaki, H., Tsuda, H.: Interactive multi-objective route planning for sightseeing on Time-Expanded Networks under various conditions, *Procedia Computer Science*,

Vol. 22, pp. 221–230 (2013)

- [3] Fu, C., Hu, M., Lai, J., Wang, H., Wu, J.: TravelBuddy: Interactive Travel Route Recommendation with a Visual Scene Interface, *MultiMedia Modeling*, pp. 219–230 (2014)
 [4] Hang, L., Kang, S., Jin, W., Kim, D.: Design and Implementation of an Optimal Travel Route Recommender System on Big Data for Tourists in Jeju, *Processes*, Vol. 6, No. 8 (2018)
 [5] 倉田 陽平: CT-Planer 3: Web 上での対話的な旅行プラン作成支援, 観光科学研究, No. 5, pp. 159–165 (2012)
 [6] Wang, C., Hyman, J., Percus, A., Cafisch, R.: Parallel Tempering For The Traveling Salesman Problem, *International Journal of Modern Physics C*, Vol. 20, No. 4, pp. 539–556 (2009)
 [7] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A.: Equation of State Calculations by Fast Computing Machines, *The Journal of Chemical Physics*, Vol. 21, No. 6, pp. 1087–1092 (1953)
 [8] Kirkpatrick, S., Gelatt, C., Vecchi, M.: Optimization by Simulated Annealing, *Science*, Vol. 220, No. 4598, pp. 671–68 (1983)
 [9] Feillet, D., Dejax, P., Gendreau, M.: Traveling Salesman Problem with Profits: An Overview, *Transportation Science*, Vol. 39, No. 2, pp. 188–205 (2005)
 [10] Hukushima, K., Nemoto, K.: Exchange Monte Carlo Method and Application to Spin Glass Simulations, *Journal of the Physical Society of Japan*, Vol. 65, No. 6, pp. 1604–1608 (1995)
 [11] Sohl-Dickstein, J., Mudigonda, M., DeWeese, M.: General k-Opt Submoves for the Lin-Kernighan TSP Heuristic, *Mathematical Programming Computation*, Vol. 1, No. 2–3, pp. 119–163 (2009)
 [12] Matsumoto, M., Nishimura, T.: Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudorandom Number Generator, *ACM Transactions on Modeling and Computer Simulation*, Vol. 8, No. 1, pp. 3–31 (1998)
 [13] 松坂和夫: 集合・位相入門, 岩波書店, 第 59 版 (1968)