Stack Overflow におけるリンク分類とベストアンサーの関係分析

瀬澤 勇斗　槇原 絵里奈　近藤 将成　吉田 則裕　井上 克郎
楊 詩語

近年のプログラミング・IT 教育では，学習者が適切な情報を選択・整理して他者に分かりやすく説明する「回答力」の育成が重要となっている．回答力は，自ら考えて取り組む能動的な学習や，人に教えることで自身の理解が深まる効果の観点からも，回答力の育成は重要であると考えられる．本研究は，この回答力育成の設計指針を得ることを目的に，多くの技術者が知識を共有する Q&A サイト Stack Overflow（SO）を対象として,「良い回答」がどのように情報を参照しているかを実証的に分析する. 具体的には，質問者自身が役に立った回答として選ぶベストアンサーに着目し，回答内のリンクを「外部のみ」「内部のみ」「両方」「リンクなし」の 4 種類に分類したうえで，回答者の経験・質問の特性との関係を分析した．約 1,300 万件の質問・回答ペアを分析した結果，ベストアンサー採択率は回答者の経験量よりもリンクの使い分けと強く関連し，特にソースコードを含む質問ではサイト内の既存議論への内部リンクを伴う回答が採択されやすい一方，リンクを持たない回答は採択率が低いことがわかった．以上から，有用な回答は情報量の多さではなく，質問の文脈に応じて「どこを参照し，どこを自分で示すか」を適切に選択・構成することによって特徴づけられると考えられる．本知見は，参照の仕方に着目した回答力育成の教育設計に活用できると期待される．

1　はじめに

大規模言語モデル（LLM）の普及に伴い，プログラミング教育やソフトウェア工学教育では，学習者に対するヒント生成，フィードバック，問題生成，情報探索支援など，生成 AI を活用した学習支援が広く検討されている．これらの支援は，文法エラーの解消や不足した知識の補完を容易にし，学習者が疑問を解決するための障壁を下げる一方で，学習者自身が得た知識を整理し，他者の問題解決に利用できる形へ構造化する機会は十分に扱われていない．Active Learning や Learning by Teaching に関する研究が示すように，学習内容を自ら説明し，他者に伝える過程は，知識の整理や理解の深化に寄与すると考えられる．特にソフトウェアはチームで開発されることが多いため，ユーザは単に正しいコードを書く能力だけでなく，問題の文脈や相手の背景知識を踏まえ，複数の情報源から必要な情報を選択・統合し，問題解決につながる説明を構成する能力が重要だと考える．本研究では，このような能力を回答力と呼び，回答力を向上させる支援に注目する．

プログラミング演習やソフトウェア開発 PBL（Project-Based Learning）では，大学院生や若手研究者が Teaching Assistant（TA）として学習者の質問に回答する場面が多く，質問者の理解度に応じて授業資料，既存の議論，公式ドキュメントなどを適切に提示することが求められる．このような回答力を学習初期から育成することは，将来的な TA 活動だけでな

An Empirical Study of Link Classification and Accepted Answers on Stack Overflow

Yuto Sezawa, 立命館大学大学院情報理工学研究科, Graduate School of Information Science and Engineering, Ritsumeikan University.
Erina Makihara, Norihiro Yoshida, Katsuro Inoue, Shiyu Yang, 立命館大学情報理工学部システムアーキテクトコース, System Architect Course, College of Information Science and Engineering, Ritsumeikan University.
Masanari Kondo, 九州大学大学院システム情報科学研究院, Graduate School and faculty of Information Science and Electrical Engineering, Kyusyu University.

く，コードレビュー，デバッグ，技術的な意思決定など，より高度なソフトウェア開発活動にもつながると期待される．しかし，既存の AI を用いた教育支援の多くは学習者への回答や知識補完に焦点を当てており，学習者自身に有用な回答を構成させるための知見は限定的である．

そこで本研究では，回答力を育成する教育支援の設計に向けて，技術的な質問に対する有用な回答がどのように情報源を選択・利用しているかを明らかにするため，ソフトウェア開発に関する大規模 Q&A サイトである Stack Overflow（SO）に着目する．SO では，回答中に他の SO 投稿への内部リンクや，公式ドキュメント，GitHub 等への外部リンクを提示することで，既存の議論の参照，前提知識の補足，具体例の提示などを行うことができる．また，質問者自身が問題解決に最も役立った回答をベストアンサー（Accepted Answer）として選択できるため，コミュニティ全体による Voting とは異なり，質問者が持つ個別の問題コンテキストに適応した回答を分析するための指標として利用できる [4].

先行研究では約 1,300 万件の SO 上の質疑を対象として，回答に含まれるリンクを Internal Link（内部リンク），External Link（外部リンク），Both（両方），No Link（リンクなし）に分類し，各分類の特徴を調査した [11]．本研究ではさらに質問と回答の関係性に注目し，回答者の経験や質問の特徴とリンク利用，およびベストアンサーとの関係を分析した．その結果，回答経験の多さよりもリンクの使い分けが回答の受容と関係しており，単純に多くの情報源を提示するだけでは Accepted Answer になりやすいとは限らないと分かった．特に，ソースコードを含む質問では内部リンクを含む回答がベストアンサーとなる傾向がみられ，質問の特徴に応じて既存の議論や外部情報を取捨選択することの重要性が示された．これらの結果は，有用な回答の作成には情報量だけでなく，質問者の文脈に応じた情報の選択と構造化が重要であることを示している．本知見を教育支援に応用することで，LLM を単に学習者へ答えを提示する回答者として用いるのではなく，学習者に質問を提示し，必要な情報源の選択や回答の構成を促す質問者としても活用することが可能になると考える．

2 提案手法

2.1 リンク分類

本研究では SO の投稿に含まれるリンクを元に，回答を以下 4 つに分類する．

Internal Link(Internal) SO 内部の質問，回答，コメントへのリンクのみを含むもの

External Link(External) Github や言語ライブラリなど SO 外部へのリンクのみを含むもの

Both Internal Link と External Link を含む投稿

No Link どのリンクも含まない投稿

リンクを用いることの効果を測定するため，リンクを含まない No Link も分類に加える．また，投稿によってはリンクが複数含まれる場合もあるが，分類ではリンクの数には注目せずリンク先にのみ注目する．リンク数は回答の情報量を反映する可能性があるが，本研究では質問者が情報源をどの範囲から取捨選択したかという構造的特徴に注目するため，リンク数ではなくリンク先の種別のみで分類する．

SO で質問を行う際，質問内容の再現性や自己完結の点から，自身の体験や試行錯誤を記述することが望まれている [†1]．すなわち，質問時に他サイトへの URL を貼るだけの行為は，リンク先にアクセスせねば情報が分からず，既存研究ではリンク切れによる質問の品質低下が指摘されている．

回答時も同様の理由でリンクのみの投稿はコミュニティガイドラインにおいて推奨されておらず，あくまで補助的な使用に留めるよう指摘される [†2]．一方既存研究より，適切にリンクを用いて回答した場合，質問者や読者の学習支援となり，コアとなる知識の特定にも繋がることが示唆される [3]．以上より，質問時のリンクの使用は再現性や質問者の背景知識が共有されない点から推奨されない一方，回答やコメントで補足的にリンクを用いることで，質問者や閲覧者のプログラミング理解を促すことができる．それぞれのリンク種別がベストアンサーとして採用される傾向の違いを明らかにすることで，質問者の問題解決に有用

†1 https://stackoverflow.com/help/how-to-ask

†2 https://meta.stackexchange.com/q/8231

な回答の構造的特徴を分析する．

2.2 Dataset

図 1 にデータセット作成までの概要を示す．本研究では，SO のデータを取得するため，Google BigQuery 上の公開データセットである `bigquery-public-data.stackoverflow` を用いた．BigQuery にクエリを与えることで，SO の質問・回答・ユーザデータを取得した．

SO は流行のツールや言語，モデレーションの影響を受けやすく，単年調査では傾向把握が困難なため，本調査ではまず長期間かつ大規模なデータ収集を行った．RQ1 と RQ2 に回答するためのデータはそれぞれ異なった時期に取得したが，いずれも対象は 2008 年 7 月 31 日から 2021 年 12 月 31 日までのデータである．以下に手順の概要を示す．

Step1: Data Collection

Sub-step1: 質問および回答の基本情報の取得

まず，質問と回答投稿に関する基本情報（e.g., ID），リンクを含む回答およびコメント，全体の単語数情報などを取得する．ここで述べる基本情報とは，BigQuery に保存されている，SO から直接取得可能な質問や回答の内容などを指す．

使用したテーブルは以下の 3 つである．

- `posts_questions` ：質問投稿の基本情報を格納
- `posts_answers` ：質問に付随する回答
- `comments` ：質問・回答に付随するコメント

主な処理内容として，リンク付き要素については `ARRAY_AGG` 関数で配列化し，各テキストの単語数は `LENGTH` 関数で計算し，質問・回答・コメントごとに合計した．最終的な出力では，リンクを含む回答およびコメントについては配列形式で保持し，文字数はすべての回答・コメントを対象に合計した．取得データは年ごとに分割し，ローカル環境に JSON 形式で保存した．なお，紙面の都合上使用したクエリは省略する．

Sub-step2: 単語数情報の補完

次に，HTML タグや Markdown 記法を除去したうえで，投稿やコメントの単語数を計測するため，以下に示す処理を追加で行った．これらの処理により，質問本文，回答本文，質問コメント，回答コメントの単語数を取得した．

- HTML タグの除去
- Markdown リンク変換
- 単語分割

最終的にそれぞれのサブクエリを結合し，質問 ID ごとの総単語数を集計した．

表 1　質問データに含まれる主な項目

項目名	説明
id	質問の識別子
title	質問のタイトル
accepted_answer_id	採用された回答の ID
accepted_answer_score	採用回答のスコア
creation_date	質問の投稿日時
score	質問のスコア
view_count	質問の閲覧数
answer_count	質問に対する回答数
answer_info	リンクを含む回答の情報
unlinked_answers	リンクのない回答の情報

上記 sub steps で取得したデータの概要を表 1 に示す．Step2 で示したとおり，質問と回答の本文は HTML 形式で保存されており，単語数の算出時には HTML タグを取り除いた上でカウントした．質問や回答に対するコメントは Markdown 形式 で記述されているため，Markdown 構文を除去した上で単語数を計算した．

上記の手順により合計 15,003,031 件の質問，25,728,871 件の回答，2,615,012 件のユーザ情報を取得した．回答の内訳は，Internal: 829,059，External: 8,949,915，Both: 534,678，No Link: 15,415,219 件である（図 1 の Raw data）．以降，本データを D_{raw}，質問と回答のペアを QA インスタンスと定義する．

Step2: 包含基準および除外基準

調査を行うにあたり，分析精度を高めるため表 2 に示すデータの包含基準および除外基準を設定した．まず，第三者に評価されている質問および回答であるこ

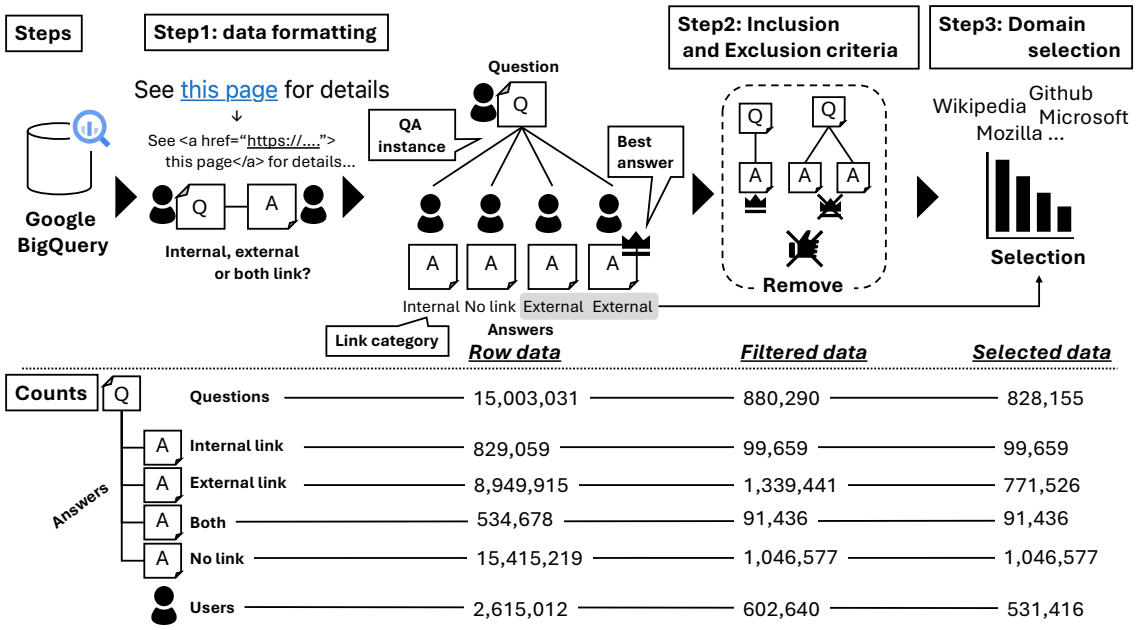

図 1　Overall workflow of creating dataset

包含基準	
1)	質問および回答の評価数（Voting）が 1 以上，かつ回答数が 2 つ以上である
2)	ベストアンサーが採用済みである
除外基準	
1)	External Link のドメインが上位 5% に含まれない

表 2　包含基準と除外基準

と，さらに回答数が 2 つ以上あることを条件とした．理由は，第三者も評価する回答が複数あるということは，質問者が良い回答を取捨選択できる環境であったことだと考えたためである．また，Voting が 0 の質問・回答は SO ユーザによる評価が行われていないとみなし，最低限の評価が行われている 1 以上の質問・回答を対象とした．

さらに，ベストアンサーが決められていることを条件に加えた．ベストアンサー自体は回答数が 1 だったとしても，質問者が有用だと判断すれば採用することができる．一方，我々の調査では，ユーザは回答に含まれる何らかの手がかりを元に，その回答を有用だと判断すると仮定している．単体の質問に対しても回答の有用性を評価することはできるが，複数の質問を見比べたうえで最も有用だと評価された回答の方が，我々の調査対象にふさわしいと考えた．

上記 Inclusion criteria によるフィルターの結果，880,290 件の質問，2,577,113 件の回答，602,640 件のユーザ情報が残った．回答の内訳は，Internal: 99,659，External: 1,339,441，Both: 91,436，No Link: 1,046,577 件である（図 1 の Filtered data）．削減率は Internal: 87.98%，External: 85.04%，Both: 82.90%，No Link: 93.21% であり，全体として 89.98% のデータが削減され，一定以上の相互作用および評価が存在する QA インスタンスが残った．以降，本データを D_{filtered} と定義する．

Step3: Exclusion of External Links

特性上，External Link は Internal Link に比べて参照先ドメインの種類が豊富であり，分析対象数も

大規模になる可能性がある．実際，前節でも Internal が 99,659 件，Both は 91,436 件に比べ External は 1,339,441 件と 10 倍以上の大きさである．

また既存研究において，External Link で自身のブログへ誘導するアフィリエイトのような不適切な External Link の利用が報告されている [7]．我々が対象とする External Link の使用は，言語ドキュメントや Github のソースコードなどを参照とした，知識の拡張や補填としての使用である．よって，全ての External Link を対象とすると，アフィリエイトやリンク切れのドメインなど，ノイズとなるデータが多く含まれる恐れがある．

そこで，不適切と思われる External Link を除外し，主要な External Link の利用を調査するため，External Link におけるドメインに注目した．対象の External Link のドメインを全て調査し，それぞれのドメイン発生回数を降順に並べた．結果，合計のドメイン数は 72906 件であった．紙面の都合上，ドメインの上位 50 件の結果を図 2 に示す．

このドメインの集合空間を S と定義する．図 2 からも読み取れるとおり，外部ドメインの分布はロングテイルである．先行研究から，外部ドメインには公式ドキュメント，Wikipedia，ツールや API，ライブラリ多く含まれていると分かっている [6]．先行研究と本研究では調査した年数が異なるものの，先行研究の結果は図 2 の結果とも一致する．

累積 50% を超える最小のパーセンタイルを調べたとこ，5% であったそこで上位 5% のドメイン $S_{\text{top5\%}}$ に注目したところ，合計 24 件のドメインが対象となり，全体の 51.64% の External Link を網羅すると分かった．すなわち，$S_{\text{top5\%}}$ でも全体の半数以上の External Link を占めることから，External Link の傾向を知るには十分な数であると判断した．以上より，External Link における上位 5% のドメインに含まれない External Link を，フィルター後のデータから除外し（Exclusion criteria1），$S_{\text{top5\%}}$ 空間を構築した．

Exclusion criteria1 はリンク単位のフィルタリングであり，QA インスタンス単位の包含判定とは独立して適用する．具体的には以下の方針をとる．

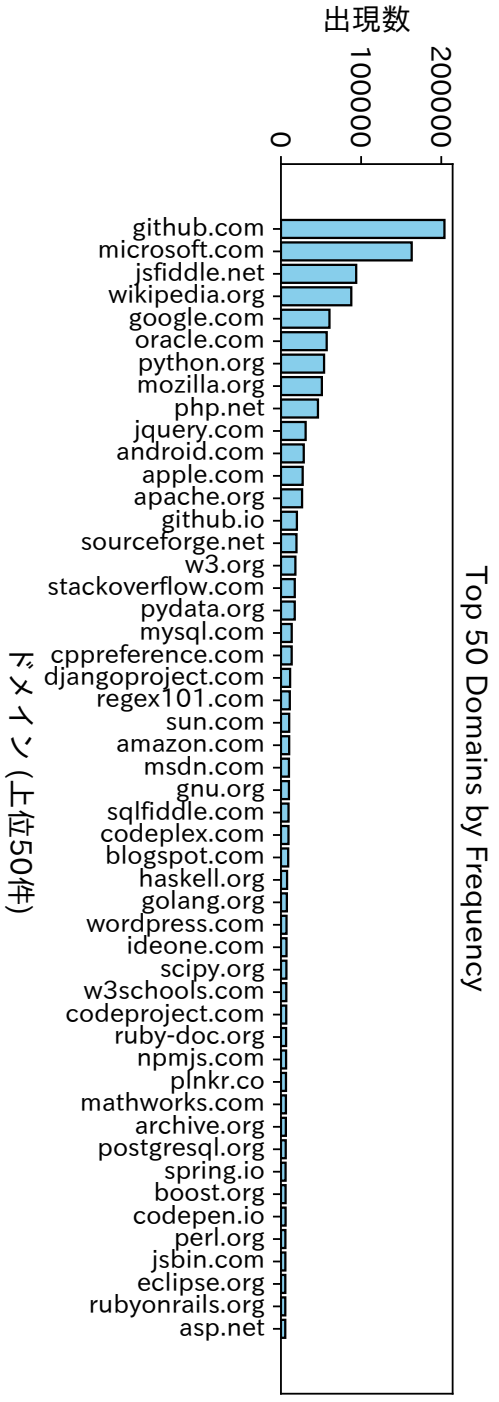

図 2　ドメインの上位 50 件

- ベストアンサーに採用された回答に含まれる External Link は Exclusion criteria1 の対象外

とし，除外しない．ベストアンサーは本研究の主要な分析対象であり，その External Link を除外すると分析結果にバイアスが生じ，ベストアンサーの実態を正確に反映できなくなるためである．

- Both カテゴリは External Link が除外された場合も Internal への再分類は行わず，Both として扱う
- Exclusion criteria1 の適用結果として Inclusion criteria1 を満たさなくなる QA インスタンスが生じる場合があるが，当該インスタンスはすでに D_{filtered} の段階で有効な議論が行われたものとして包含されているため，再評価は行わない

以上の結果，External Link は 1,339,441 件から 1,076,075 件が対象として選定された．また，質問は 880,290 件，ユーザ数は 560,925 名となった（図 1 の Selected data）．

2.3 Research Question

質問に対する適切な回答の調査に向けて，リンク分類とベストアンサーの観点から調査を進める．そのため以下の Research Questions(RQs) を設定する．

RQ1: ユーザの回答経験とリンク使用傾向に関連はあるか

RQ2: 質問の特性と回答におけるリンク分類にどのような関係があるか

RQ1 で回答経験が豊富なユーザはどのようにリンクを使い分けるかをまず調査する．これは，回答経験が豊富なユーザはリンクを多用することで情報が豊富な回答を行っていると仮定したためである．もし回答経験が豊富なユーザほど Internal と External Link を併用するなどの特徴がある場合，ユーザの回答力は使用するリンク分類で評価可能であることを示唆する．一方，個人によってリンク使用傾向が異なる場合，回答経験とリンクの使用には関係がない可能性がある．

次に RQ2 で，ユーザは質問のどのような要素をヒントにリンクを使い分けるかを調査した．これは，ベストアンサーは回答者が役に立ったと感じたものが選べる．したがって，回答単体のメトリクス（文字数やソースコードの有無）で評価ではなく，リンク使用は質問と照らし合わせ評価する必要があったと考えたためである．

まず，質問に含まれる特徴が回答のリンク分類に与える影響を調査するため，質問特徴を説明変数，回答のリンク分類を目的変数とした回帰分析を行う．単純なリンク分類とベストアンサーになった場合それぞれで回帰分析を行うことで，ベストアンサー独自の特徴も調査する．次に，回答内容に注目し，コードの有無やコード行数，回答本文量とリンク分類との共起・相関関係を分析する．これにより，質問側の要因と回答側の情報提示方法の両面から，SO におけるリンク使用傾向とベストアンサーの関係を明らかにする．

3 結果

3.1 RQ1: ユーザの回答経験とリンク使用傾向に関連はあるか

まず，回答数の中央値と標準偏差を調査した結果，中央値は 1.0，標準偏差は 22.7 となった．よって，回答活動は少数の高活動ユーザに集中しており，回答者間で活動量に大きな偏りが存在することが確認された．この結果はリンク使用傾向もユーザ活動量によって異なる可能性があることを示す．

そのため，ユーザごとのリンク分類傾向および質問特徴との関係を分析する．回答回数が上位のユーザ 50 名の回答回数を図 3 に，およびリンク分類を図 4 に示す．なお，ユーザ ID は本稿では匿名化しており，実際の SO におけるユーザ ID とは異なる．また，ベストアンサーとして選ばれた割合を図 5 に示す．

図 4 から，回答が活発なユーザでもリンク使用戦略は均一ではないことがわかる．ユーザによっては極端に External や No Link が多くなっており，これはユーザによって回答時にリンク使用の戦略が異なると示唆される．すなわち，回答が活発になることと，リンクの使用戦略は一致しないと言える．したがって，回答の回数は回答の有用性を図るメトリクスとして不適切であり，より詳細な調査が必要となる．

ここで，例えば ID：57695 や 69083 のユーザは No Link の回答が多く，またベストアンサー率も極端に低い．そこで，各リンクカテゴリのベストアンサーの採用率および「リンクの有無（No Link か否か）」が

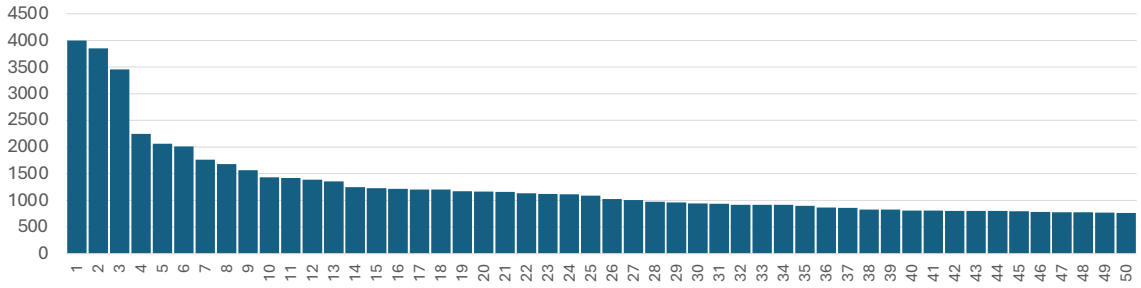

図 3　上位のユーザ 50 名の回答回数

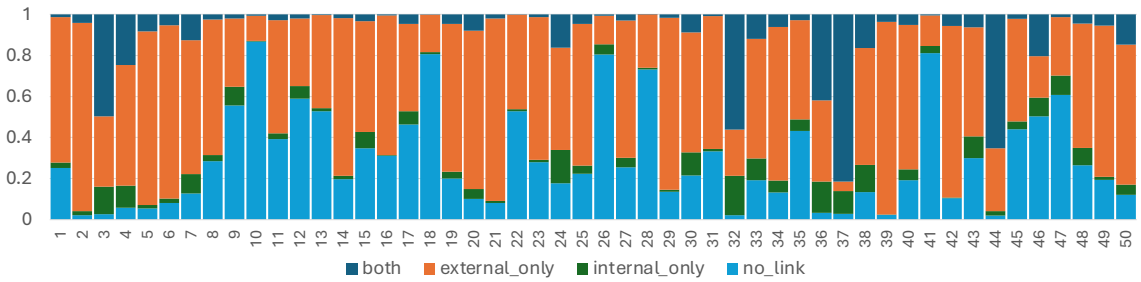

図 4　回答回数が上位のユーザ 50 名のリンクカテゴリ

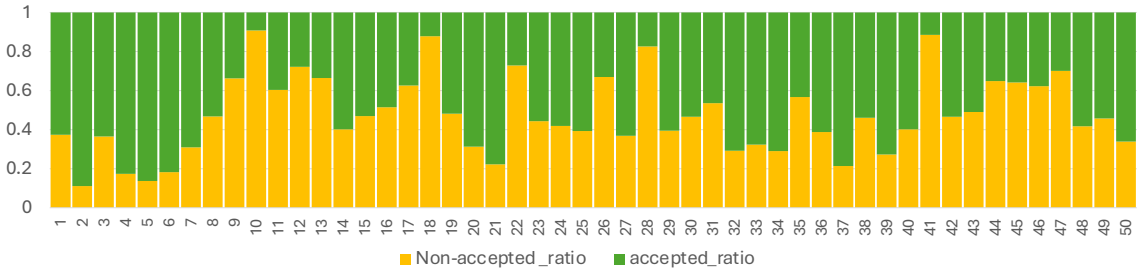

図 5　回答回数が上位のユーザ 50 名のベストアンサーの割合

表 3　各リンクカテゴリのベストアンサー採用率

Link Category	All Answers	(%)	Accepted Answers	Accepted Rate (%)
External Only	1,076,075	46.51	744,655	69.20
No Link	1,046,577	45.23	21,306	2.04
Internal Only	99,659	4.31	58,076	58.27
Both	91,436	3.95	56,253	61.52
Total	2,313,747	100.00	880,290	38.04

「ベストアンサーの選出」に与える影響をカイ二乗検定および相関分析を実施した．各リンクカテゴリのベストアンサー採用率を表 3 にまとめる．

結果，有意差検定の p 値は 0.01 を下回り，リンクを含まない回答とリンクを含む回答との間で，ベストアンサーの選出に有意な差を確認した．また，点双列相関による相関分析を行ったところ，相関係数は-0.6741 となった．すなわち「リンクを含まないこと」と「ベストアンサーに選出されること」の間に，強い負の相関が存在する．

以上の結果から，回答の有用性は回答活動量のみでは十分に説明できず，リンクを用いた情報提示戦略が

ベストアンサー選出に影響している可能性が示唆される．したがって次節では，Internal，External，Both といったリンク分類ごとの特徴に注目し，回答時の情報の取捨選択の戦略と回答の有用性との関係をさらに詳細に分析する．

Answer to RQ1

回答経験が豊富なユーザであってもリンク使用戦略は均一ではなく，External Link を多用するユーザや No Link が多いユーザなど，ユーザごとの差異が見られた．また，リンクを含まない回答である No Link カテゴリとベストアンサー選出の間には強い負の相関が確認された．以上より，ユーザの回答経験とリンク使用傾向に十分な関連は見られない．さらに，リンクを用いた情報提示戦略がベストアンサー選出と関連している可能性が示唆された．

3.2 RQ2: 質問の特性と回答におけるリンク分類にどのような関係があるか

RQ1 の結果よりリンクの使用がベストアンサー採用の重要な要因であることが示された．RQ2 ではリンク分類とベストアンサーの関係をさらに詳細に調査するため，No Link を含む全 4 分類を対象に分析を行う．

質問の特徴とリンク分類，およびベストアンサーの関係を調査するため，質問の特徴を説明変数，回答のリンク分類を目的変数とした多項ロジスティック回帰分析を行った．目的変数は Both，External Only，Internal Only，No Link の 4 カテゴリであり，基準カテゴリは No Link とした．すなわち，各係数は No Link と比較した場合の各リンク分類の選択されやすさを示す．RQ1 の結果でリンクが有る方が採用されやすい傾向が分かっているが，リンクを含むリンクカテゴリ間での比較のため，No Link を基準として用いる．

説明変数には，回答スコア（score），質問スコア（question_score），対象の質問に付随する全回答数（answer_count），質問へのコード有無（has_code）を用いた．全回答モデルではさらにベストアンサーフラグ（is_accepted）を加えた 5 変数を使用した．なお，ベストアンサーのみのモデルでは is_accepted が全て 1 となり変数として機能しないため，除外した 4 変数を使用した．各説明変数は StandardScaler により標準化したうえでモデルに投入したため，係数の大きさはカテゴリ間で直接比較可能である．External Only と No Link で全体の約 92% を占めるデータの不均衡に対処するため，class_weight='balanced' を設定した．モデルの実装には scikit-learn の LogisticRegression を使用し，solver に lbfgs，multi_class に mnomial を指定した．

表 4 に回答全体を対象にした計数，表 5 にベストアンサーとなった回答のみを対象とした計数を示す．表 4 と表 5 の結果を，各リンクカテゴリでまとめる．数値が大きいほど質問の特徴と，回答および採用されたベストアンサーの回答の関連の強さを示す．一方後述する各リンクカテゴリでのデータ数の不均等から，本稿では係数の絶対値には注目せず，係数の正負および相対的な大小の比較を主に議論する．

Both: ベストアンサーに関係なく，回答および回答の対象となる質問のスコアが高い．また，ベストアンサーに採用される場合，回答対象の質問には他の回答も多く行われており，議論が盛んな質問おいて用いられベストアンサーに採用されやすい．

External Link: ベストアンサー採用率が高く，回答の対象となる質問にはコードが含まれないことが多い．ベストアンサーに採用される場合，質問のスコアとの関連は低い．

Internal Link: 回答の対象となる質問にコードが含まれていることが多い．また，ベストアンサーとなる場合は，質問にコードが含まれており，かつ回答の対象となる質問の評価が高いことが多い．

全回答モデルおよびベストアンサーのみのモデルの Accuracy はそれぞれ 0.75，0.85 であった．一方，Macro F1 はそれぞれ 0.41，0.23 であり，External および No Link に比べ Both と Internal の分類精度が低い．これは External と No Link が全体の約 92% を占めるクラス不均衡に起因するものである．本分

表 4　全ての回答に関する多項ロジスティック回帰係数 (参照: No Link)

Link Category	score	question_score	answer_count	has_code	is_accepted
Both	0.143	0.034	−0.033	−0.042	0.493
External Only	0.077	0.005	−0.074	−0.120	0.658
Internal Only	0.016	−0.002	0.023	0.087	0.435
No Link (ref.)	0.000	0.000	0.000	0.000	0.000

表 5　ベストアンサーの回答に関する多項ロジスティック回帰係数 (参照: No Link)

Link Category	score	question_score	answer_count	has_code
Both	0.020	0.030	0.079	−0.020
External Only	0.017	−0.017	0.035	−0.090
Internal Only	−0.136	0.072	0.067	0.124
No Link (ref.)	0.000	0.000	0.000	0.000

析の目的はリンク分類の予測モデル構築ではなく，各説明変数とリンク分類の関連における正負および相対的な大小の把握である．Exclusion criteria1 により External Link のドメインには制限をかけたが，なお External Link が多数を示すことに変わりは無い．したがって，本結果の係数の正負および相対的な大小の比較は参考とできるが，絶対値の大きさの解釈には限界がある．ランダムサンプリングを用いたサンプル数の平等化および分析は今後の展望とする．

Answer to RQ2

質問の特徴に応じ，回答のリンク分類およびベストアンサーとして採用される回答のリンク分類に関係が見受けられた．Both カテゴリは議論が盛んな質問において使用された際，ベストアンサーに採用されやすい．External Link はベストアンサー率は高いが，質問のスコアとの関連は低く，コードを用いた質問ではあまり用いられない．Internal Link は External Link と対照的に，質問がコードを含む場合利用され，ベストアンサーにも採用されやすい．以上より，質問の特性に応じてベストアンサーに採用される回答のリンク分類は異なっており，質問の性質を見極めたリンクの使い分けが重要である．

4　考察

4.1　リンク分類と回答力

多項ロジスティック回帰分析により，質問の特徴と，それに対する回答およびベストアンサーで採用される回答の特徴が判明した．特に，質問にコードが含まれる場合，Internal Link が回答で用いられやすく，ベストアンサーにも採用されやすいことが分かった．一方，ベストアンサーにおいて回答の score が負になったことから，回答スコアが高くても必ずしも Internal Link を含む回答をベストアンサーに選ばないと示唆される．すなわち，質問者が特定のソースコードを用いて具体的な質問をする際は，言語ドキュメントや Github など外部の情報を与えても問題解決には繋がりにくい．すでに SO に存在する回答をまとめて，構造化し伝える能力が重要である．

また，Both カテゴリは議論が盛んな質問において使用された際，ベストアンサーに採用されやすい．すなわち，情報を検索した際様々な回答がヒットするような質問は，既存の議論や外部の情報などを混ぜ合わせ，情報量の高い回答が好まれると考えられる．

その一方，External Link は回答全体としては score の値が高いが，ベストアンサーに採用された回答に限ると score の値が低くなる．これは，External Link を用いた回答は，SO 全体からは評価される一方，質問者にとってはスコアの高さはベストアンサー選択

にあまり影響しないことを示唆する．また，質問にコードが含まれないことがおおいため，具体的なコードがない概念的な質問で External Link が用いられると考えられる．また，全体では question_score が 0.005 だったが，ベストアンサーに採用された時の質問の question_score は-0.017 である．これは，SO ではあまり評価されないニッチな質問に対し，External Link を用いた回答が好まれることを示唆する．

以上より，ベストアンサーに採用される回答のリンク使用は，質問の特性に応じて異なることが示された．質問にコードが含まれる具体的な問題には Internal Link が，コードを含まない概念的・ニッチな質問には External Link が，議論が盛んな質問に対しては内外のリンクを組み合わせた Both が好まれる．また，一方，コミュニティ全体の評価（score）と質問者の評価（ベストアンサー）の間にズレが存在した．

Takeaway message
質問者は回答の評価の高さよりも自身の問題解決への適合性を重視してベストアンサーを選択すると示唆される．すなわち，有用な回答とは情報量の多さではなく，質問の文脈に応じた適切なリンクの選択と構造化にあると考えられる．

4.2 妥当性の脅威

内部妥当性への脅威：表 4 と表 5 に示すとおり，本分析対象のデータは External Link と No Link が過半数を占めており，Internal Link と Both の係数計算が不安定である．class_weight=’balanced’ で影響が最小限になるよう修正しているが，Macro F1 は低くなっている．したがって，本結果の特に Internal Link と Both に関しては，係数の絶対値ではなく大小や他カテゴリとの相対的な比較にとどめている．

また，ベストアンサーは質問者の主観的なものであり，本分析で得られた結果は全ての質問者に対して平等に適用できるわけではない．一方，本分析は大規模且つ年度を問わない分析となっており，特定のユーザの性質に偏っている可能性は低い．

加えて，いつどのリンクを使ったかなど，回答回数の増加とリンク戦略の時系列は，今回の分析において考慮していない．さらに，本分析ではコメント数や単語数など，今回用いなかった質問メトリクスも取得しており，これらのメトリクスがベストアンサーに寄与する可能性もある．時系列解析や異なるメトリクスも含めた分析は今後の課題である．

外部妥当性への脅威：本調査は 2008 年から 2021 年のデータを対象としている．近年の生成系 AI の影響を鑑みると，本分析の結果は現在の SO コミュニティやユーザの傾向と異なる可能性がある．

また，本調査ではタグなどを用いて質問のトピックを分けていない．具体的な言語かフレームワークかなど，質問の傾向によって好まれた回答が異なる可能性がある．

4.3 関連研究

SO は開発者の知識共有エコシステムとして長年重要な役割を担ってきた．Anderson らはどのような質問や回答が SO コミュニティにとって将来的に有望であるかを議論し [1]，Liu らは開発者らが SO の検索をどのように扱っているかを分析して検索性の向上を目指した [7]．これらの研究はいずれも，SO をより強固なエコシステムとして構築するために重要な研究である．近年，ChatGPT などの大規模言語モデル（LLM）の台頭により，SO の役割に変容が見られる．Helic らは，AI 登場後に SO の投稿数自体は減少傾向にあるものの，投稿の長さや質問の難易度が上昇しており，SO がより複雑で困難な問題を解決するためのクラウドソース・プラットフォームとして，以前にも増して重要な役割を担っていることを明らかにした

このように高度化する SO のエコシステムにおいて，情報の信頼性と文脈を担保するリンクの存在は極めて重要である．Liu らは繰り返し引用されるリンクに注目し，リンク切れが投稿記事の信頼性を低下させる恐れを指摘した [8]．また，外部リソースの活用に関して，Gao らは API ドキュメント等への外部リンク頻度がべき乗則に従うことを示し，適切な公式ドキュメントへの自動リンク手法を提案した [5]．さらに，内部リンクによる知識の構造化について，Sun らは質問スレッドをノードとした「連結知識サブグ

ラフ」を構築し，リンクによる情報提供が投稿の活動期間を延ばすポジティブな側面を明らかにした[9]. Baltes ら[3] や Gómez ら[6] も，定性的な調査からリンク先の情報提供がユーザの新しい技術獲得やイノベーションの波及に寄与していると報告している．

一方で，生成 AI によるプログラミング支援が普及した現在，AI が生成する回答には信頼できる情報源に基づく裏付けや，開発現場特有の複雑な背景の理解が欠如しているという新たな課題が指摘されている．Zhou らは AI ペアプログラミングにおける課題を調査し，AI はコード生成は可能であるものの，開発者が抱える深い文脈の理解や背景説明が不足していることを指摘した[10]．また，Ariza らは教育現場における生成 AI の影響をレビューし，情報源の信憑性を判断する能力の重要性を説いている[2]．AI がもっともらしいが根拠の不明確な回答を生成し得る今，人間が意図を持って付与した信頼できる情報源の価値は再評価されるべきである．

5 おわりに

本研究ではユーザの回答力に着目し，ソフトウェア開発における最大規模の QA プラットフォームである Stack Ocerflow を調査した．SO の回答におけるリンクを「内部リンク」「外部リンク」「両方」に分け,「リンク無し」も加えて各リンクのベストアンサー採用率および質問文との関係性を調査した．その結果，多く回答を行う，すなわち回答の経験が多いからと行ってベストアンサー採用率が高いとは限らないと分かった．また，特に質問でソースコードを用いる際，内部リンクを用いた回答がベストアンサーとして採用されやすいと判明した．

今後は，回答や質問に含まれる他の特徴も用いて，ベストアンサー採用のメカニズムについて解明を進める．また，回答力を定式化することでユーザの回答力を定量化し，回答力の増減に応じて質問の変更やフィードバックを促す LLM 実装も目指す．

謝辞 本研究は JSPS 科研費 24K20905 の助成を受けたものです．

参 考 文 献

[1] Anderson, A., Huttenlocher, D., Kleinberg, J., and Leskovec, J.: Discovering value from community activity on focused question answering sites: a case study of stack overflow, *proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2012, pp. 850–858.

[2] Ariza, J. Á., Restrepo, M. B., and Hernández, C. H.: Generative AI in Engineering and Computing Education: A Scoping Review of Empirical Studies and Educational Practices, *IEEE Access*, Vol. 13(2025), pp. 30789–30810.

[3] Baltes, S., Treude, C., and Robillard, M. P.: Contextual Documentation Referencing on Stack Overflow, *IEEE Transactions on Software Engineering*, Vol. 48, No. 1(2022), pp. 135–149.

[4] Calefato, F., Lanubile, F., and Novielli, N.: How to ask for technical help? Evidence-based guidelines for writing questions on Stack Overflow, *Information and Software Technology*, Vol. 94(2018), pp. 186–207.

[5] Gao, S., Xing, Z., Ma, Y., Ye, D., and Lin, S.-W.: Enhancing knowledge sharing in Stack Overflow via automatic external web resources linking, *proceedings of the 2017 International Conference on Engineering of Complex Computer Systems (ICECCS)*, 2017, pp. 90–99.

[6] Gómez, C., Cleary, B., and Singer, L.: A study of innovation diffusion through link sharing on stack overflow, *proceedings of the 10th Working Conference on Mining Software Repositories*, 2013, pp. 81–84.

[7] Liu, J., Baltes, S., Treude, C., Lo, D., Zhang, Y., and Xia, X.: Characterizing search activities on stack overflow, *proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2021, pp. 919–931.

[8] Liu, J., Xia, X., Lo, D., Zhang, H., Zou, Y., Hassan, A. E., and Li, S.: Broken External Links on Stack Overflow, *IEEE Transactions on Software Engineering*, Vol. 48, No. 9(2022), pp. 3242–3267.

[9] Sun, T., Gong, L., Zhang, J., and Wei, M.: An empirical study on the characteristics of connected knowledge subgraphs on Stack Overflow, *proceedings of the 2024 IEEE 24th International Conference on Software Quality, Reliability, and Security Companion (QRS-C)*, 2024, pp. 64–71.

[10] Zhou, X., Liang, P., Zhang, B., Li, Z., Ahmad, A., Shahin, M., and Waseem, M.: Exploring the Problems, their Causes and Solutions of AI Pair Programming: A Study on GitHub and Stack Overflow, *Journal of Systems and Software*, (2024).

[11] 瀬澤勇斗, 槇原絵里奈, 近藤将成, 吉田則裕, 井上克郎: Stack Overflow における内部リンクと外部リンクの影響調査, 情報処理学会 ソフトウェアエンジニアリングシンポジウム（*SES*）, 2025, pp. 190–197.