

Stack Overflowの 回答の自動検出に向けた 投票数と更新数の分析

オープンソースソフトウェア工学研究室

学籍番号 60230010

氏名 井町 瑠貴

背景 | Stack Overflow における回答の陳腐化

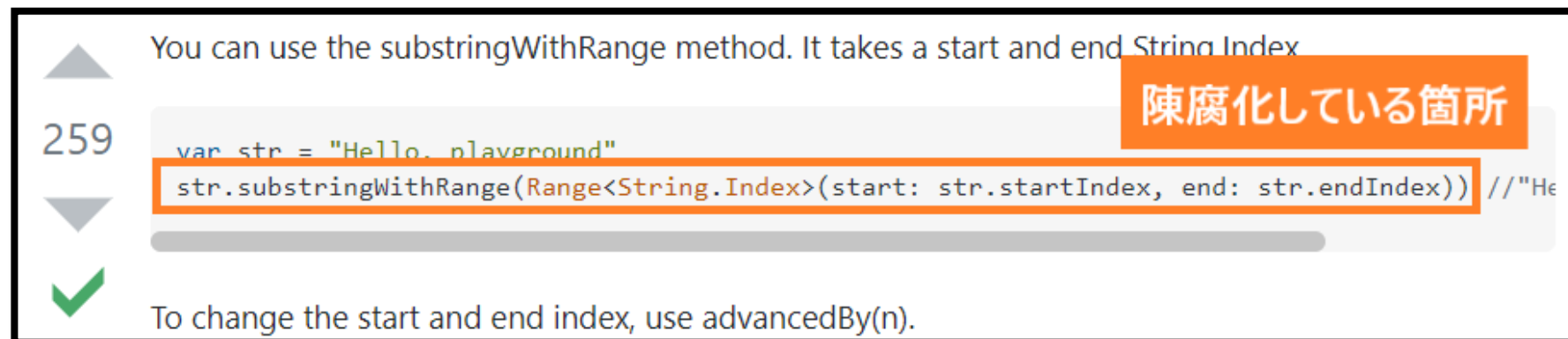


stackoverflow

人気のある技術系Q&Aサイト
膨大な量の知識を蓄積しており、
知識の共有や学習に用いられる

回答の陳腐化： ソフトウェアの進化に伴い、今ある
回答が役に立たなくなるという現象

(例) 回答された当時のSwift2系列では活用できたが、
Swift3.0以降では活用できない知識となっている^[1]



[1] <https://stackoverflow.com/questions/24044851/how-do-you-use-stringsubstringwithrange-or-how-do-ranges-work-in-swift/24045156#24045156>

背景 | Stack Overflow における回答の陳腐化

現状では、回答陳腐化の指摘と陳腐化解消は
個々のユーザーの人手に委ねられており、
陳腐化した回答は放置されたままになっている



陳腐化した回答を検出し、陳腐化を解消する
などの対応をする必要がある

本研究の着眼点

本研究では、投票数と更新数に着目する

- 投票数

- 陳腐化した回答はユーザーから支持されないため投票されることが少なくなると考えられる

- 更新数

- 回答が陳腐化するのには回答内容の更新が行われなかったためであると考えられる

目的とアプローチ

目的

陳腐化した回答の自動検出手法の構築に必要な要件を明らかにする

アプローチ

投票数と更新数の推移と回答の陳腐化のタイミングとの関係进行分析する

リサーチクエスチョン

RQ1

投票数と更新数は陳腐化を検出するうえで
重要なメトリクスになり得るか

RQ2

陳腐化指摘後も更新数が
増加している回答の特徴は何か

データセット

先行研究で利用されているデータセットを用いる^[3]

収集されている669件の回答から

以下の条件を満たす回答を抽出

- 回答が作成されてから、コメントで指摘されるまでの期間が1年以上である
- 投票数が1件以上である
- 更新数が1回以上である

これらの基準をみたす回答は**132件**であった

[3] https://github.com/SAILResearch/replication-obsolete_answers_SO

分析 | RQ1の分析方法

RQ1：投票数と更新数は陳腐化を検出するうえで
重要なメトリクスになり得るか

投票数

陳腐化が指摘されたタイミングで投票数の
増加の仕方に特徴がみられないか、
傾向を得るために移動平均を用いて調査

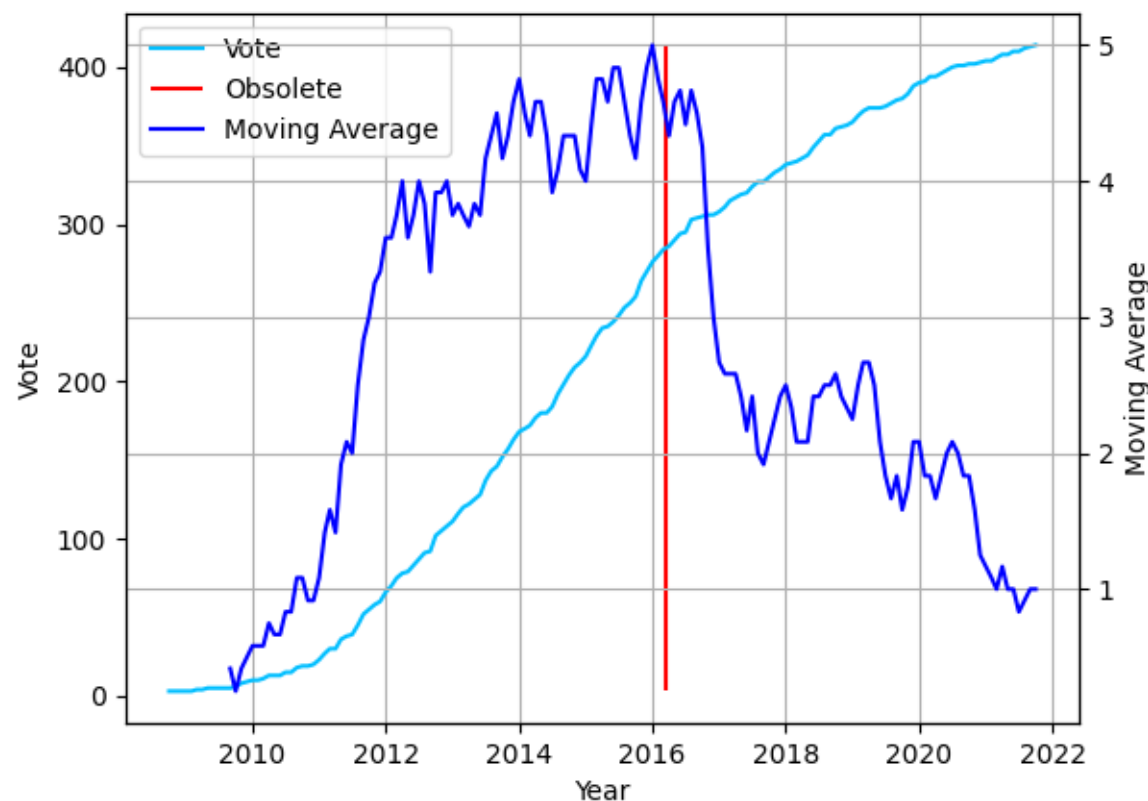
更新数

陳腐化指摘後、更新数が
増加しているかを目視で調査

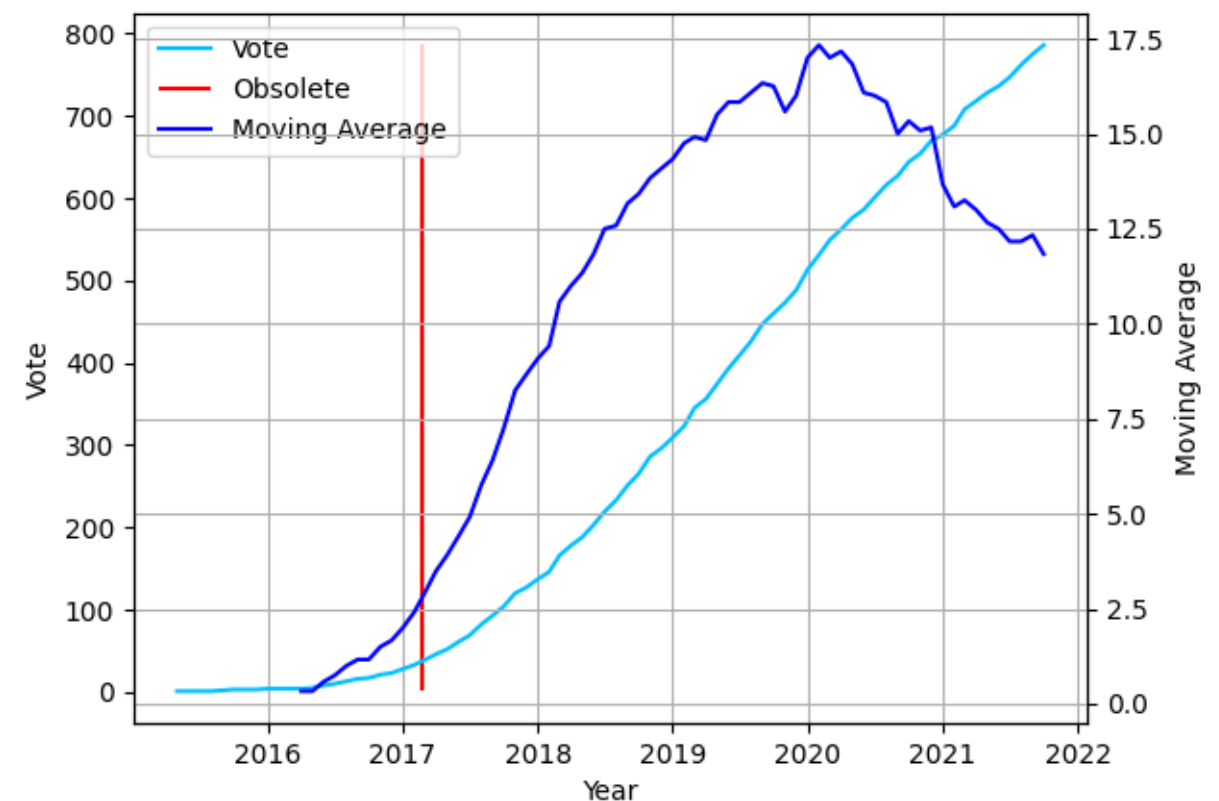
分析 | RQ1の分析結果

RQ1：投票数と更新数は陳腐化を検出するうえで重要なメトリクスになり得るか

■ 投票数



移動平均が減少

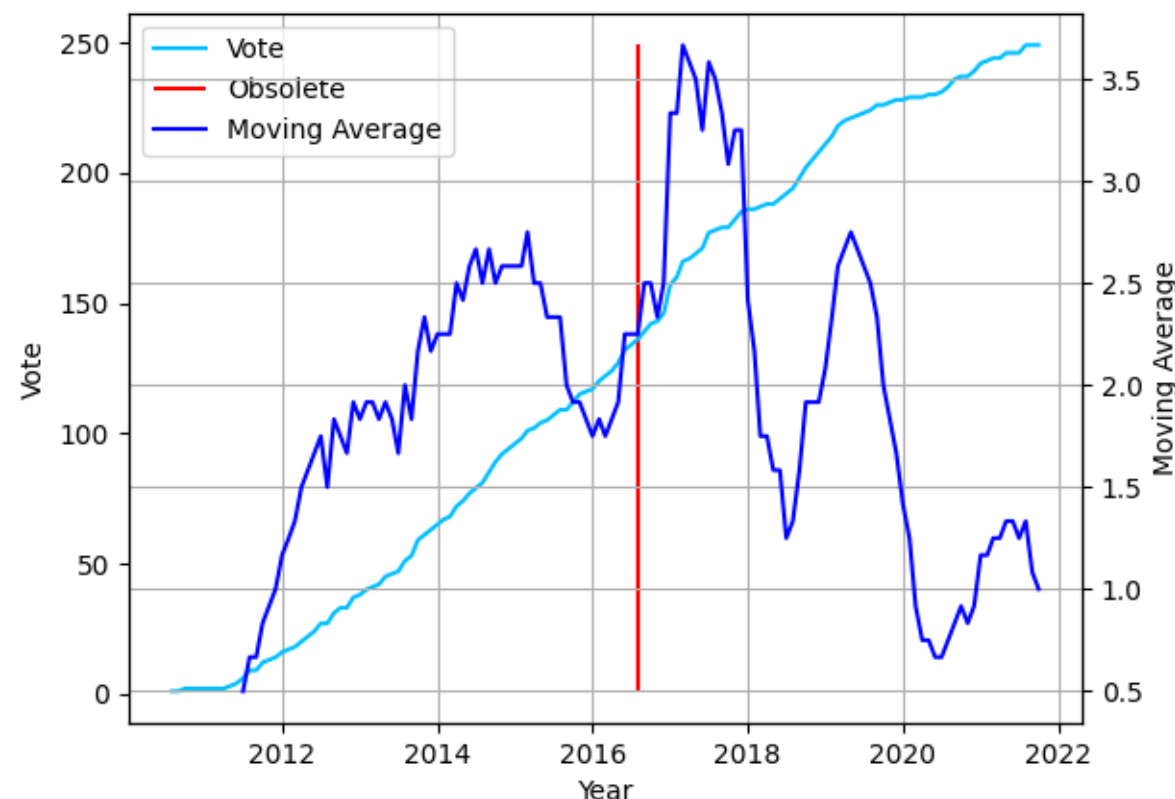


移動平均が増加

分析 | RQ1の分析結果

RQ1：投票数と更新数は陳腐化を検出するうえで重要なメトリクスになり得るか

- 投票数



陳腐化指摘に関連した変化が見られない

分析 | RQ1の分析結果

RQ1：投票数と更新数は陳腐化を検出するうえで重要なメトリクスになり得るか

- 投票数

移動平均の調査結果

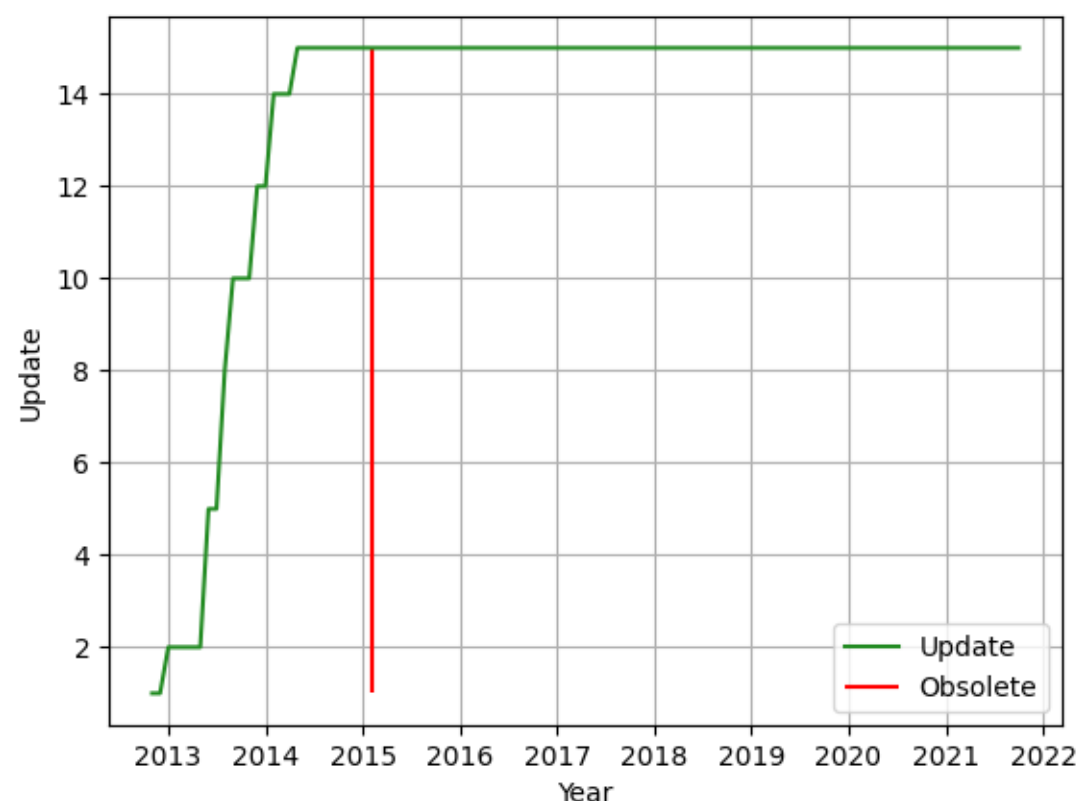
タイプ	内訳
減少	56 (42.4%)
増加	31 (23.5%)
変化なし	45 (34.1%)

陳腐化指摘後， 42.4%の回答の移動平均が減少しており， 陳腐化指摘後の移動平均の変化に着目することは重要である

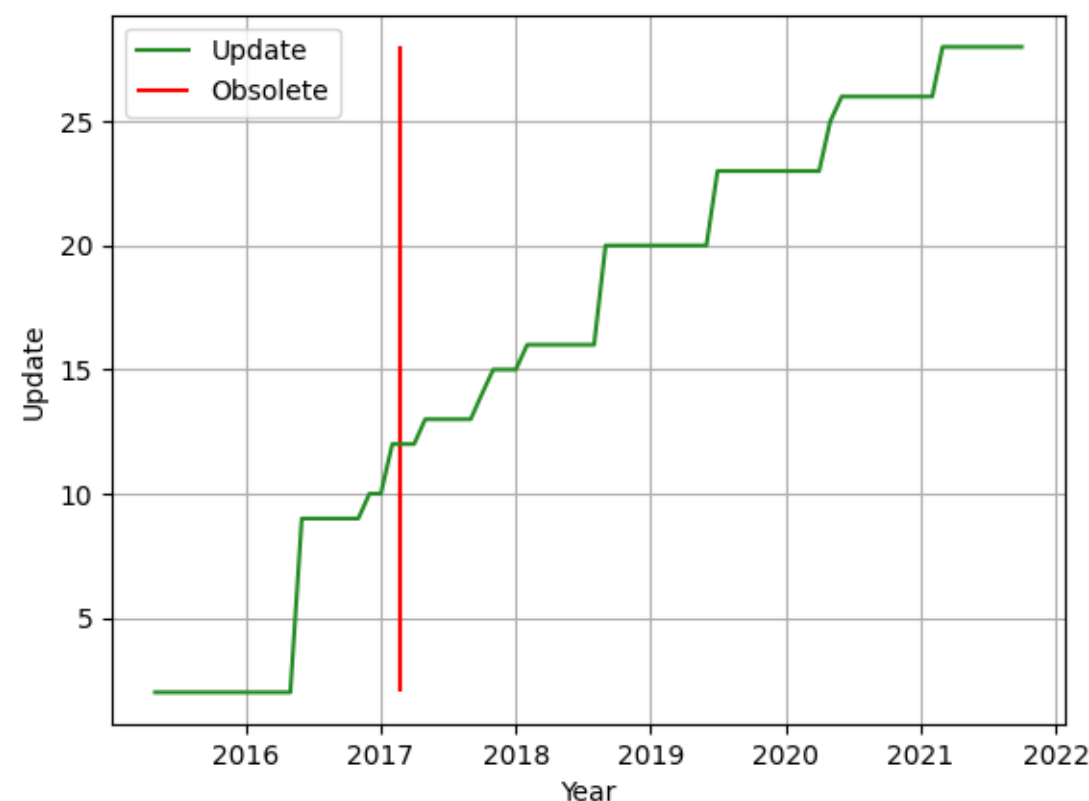
分析 | RQ1の分析結果

RQ1：投票数と更新数は陳腐化を検出するうえで
重要なメトリクスになり得るか

■ 更新数



更新数の変化なし



更新数が増加

分析 | RQ1の分析結果

RQ1：投票数と更新数は陳腐化を検出するうえで
重要なメトリクスになり得るか

- 更新数

更新数の調査結果

タイプ	内訳
変化なし	57 (43.2%)
増加	75 (56.8%)

56.8%の回答が陳腐化指摘後更新が行われており、陳腐化指摘後、更新数が増加している回答に着目することは重要である

分析 | RQ2の分析方法

RQ2：陳腐化指摘後も更新数が増加している
回答の特徴は何か

- 陳腐化が指摘された後も更新数が増加している
75件の回答を目視し、以下の3つに分類した

タイプ		内訳
更新数増加	陳腐化が解消されている	35 (26.5%)
	陳腐化が解消されていない	28 (21.2%)
	指摘が行われていない	12 (9.1%)
合計		75

- さらに、これらの3つのタイプに分類した後、
それぞれの回答の特徴について調査した

分析 | RQ2の分析結果

RQ2 : 陳腐化指摘後も投票数・更新数が増加している
回答の特徴は何か

- 陳腐化が解消している回答 (26.5%)

特徴

- 定期的にコメントの指摘があり、それをもとに更新が行われている
- 回答者による自主的な情報の更新が定期的に行われている
- スペルミスや回答をより見やすくするための修正が定期的に行われている

分析 | RQ2の分析結果

RQ2 : 陳腐化指摘後も更新数が増加している
回答の特徴は何か

- 陳腐化が解消されていない回答 (21.2%)
 - 回答の内容は変更せず, 注意を促す更新が行われているという特徴が見られた

注意を促す更新

- 回答が陳腐化していることを追記
- 動作するバージョン情報を追記
- 適切な回答へのリンクを追記

分析 | RQ2の分析結果

RQ2 : 陳腐化指摘後も投票数・更新数が増加している
回答の特徴は何か

- ・ 陳腐化の指摘が行われていない回答 (9.1%)

特徴

- ・ 回答が将来陳腐化するだろうという懸念を示している
- ・ 回答が陳腐化したかどうかを尋ねている
- ・ 議論している内容が“deprecated”, “outdated”, “obsolete”, “out of date”に関連している

データセットに誤りがある可能性があると考えられる

考察 | 陳腐化が解消していない回答

陳腐化が解消していない回答は、
陳腐化の検出を行う際に検出すべき回答である

陳腐化が解消していない回答は、
ユーザーに注意を促すような
追記が行われていることが特徴であった



更新による追記（テキスト量の増加）に着目する
ことで、陳腐化が解消していない回答を
検出することができると考えられる

まとめ

回答の陳腐化検出において、
陳腐化指摘後の投票数と
更新数の変化に着目することは重要である

陳腐化を自動検出する際、更新による
テキスト量の増加に着目することが
有効であると考えられる

今後の展望

回答の陳腐化に関連する他のメトリクスが
あるのかを明らかにし、Stack Overflow における
回答の陳腐化を自動検出するシステムの開発を行う

大学院での研究について

Stack Overflowの回答の更新提案システムの構築

背景 | Stack Overflow における回答の陳腐化

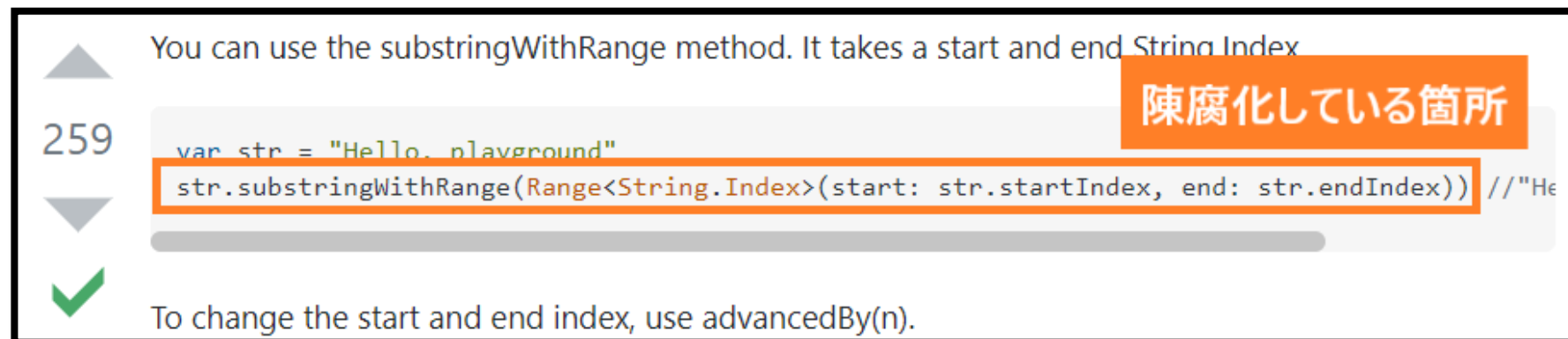


stack overflow

人気のある技術系CQAサイト
膨大な量の知識を蓄積しており、
知識の共有や学習に用いられる

回答の陳腐化： ソフトウェアの進化に伴い、今ある
回答が役に立たなくなるという現象

(例) 回答された当時のSwift2系列では活用できたが、
Swift3.0以降では活用できない知識となっている^[1]



[1] <https://stackoverflow.com/questions/24044851/how-do-you-use-stringsubstringwithrange-or-how-do-ranges-work-in-swift/24045156#24045156>

背景 | Stack Overflow における回答の更新

現状では、更新が必要な回答の指摘は、
個々のユーザの人手に委ねられており、
更新されず放置されたままの回答も多く存在する



更新が必要な回答を予測し、
更新を提案することが必要である

目的とアプローチ

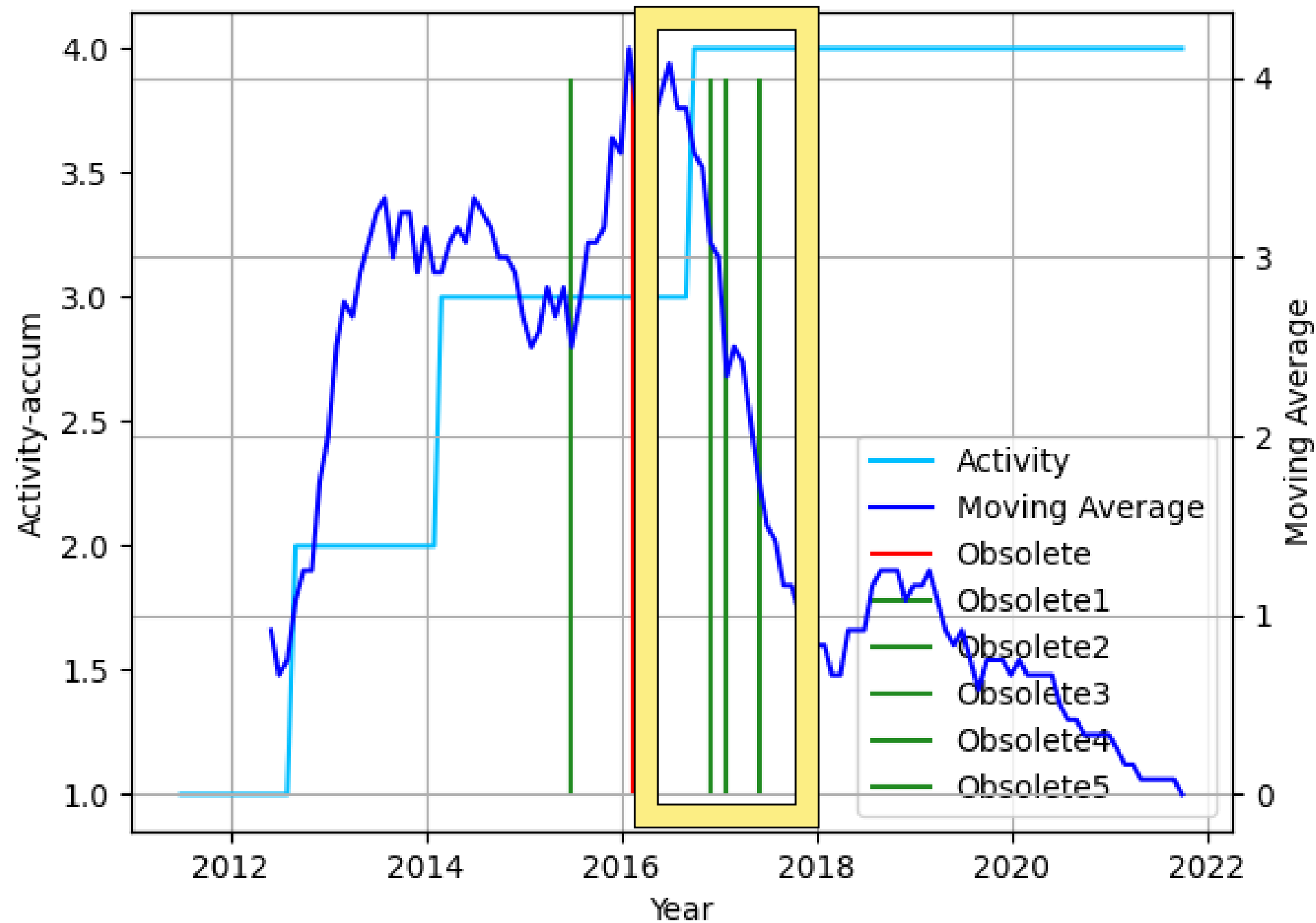
目的

更新が必要とされる回答の更新を
自動で提案する

アプローチ

回答の更新提案システムの構築

システムのイメージ



今後の予定

投票数と更新数の推移の様子を
説明できるようになる



どのようなケースがあるのかを理解し、
自分が研究を行うケースを明確にする

まとめ

- 目的
更新が必要とされる回答の更新を自動で提案する
- アプローチ
回答の更新提案システムの構築
- 今後の予定
投票数と更新数の推移の様子を説明できる
ようになる

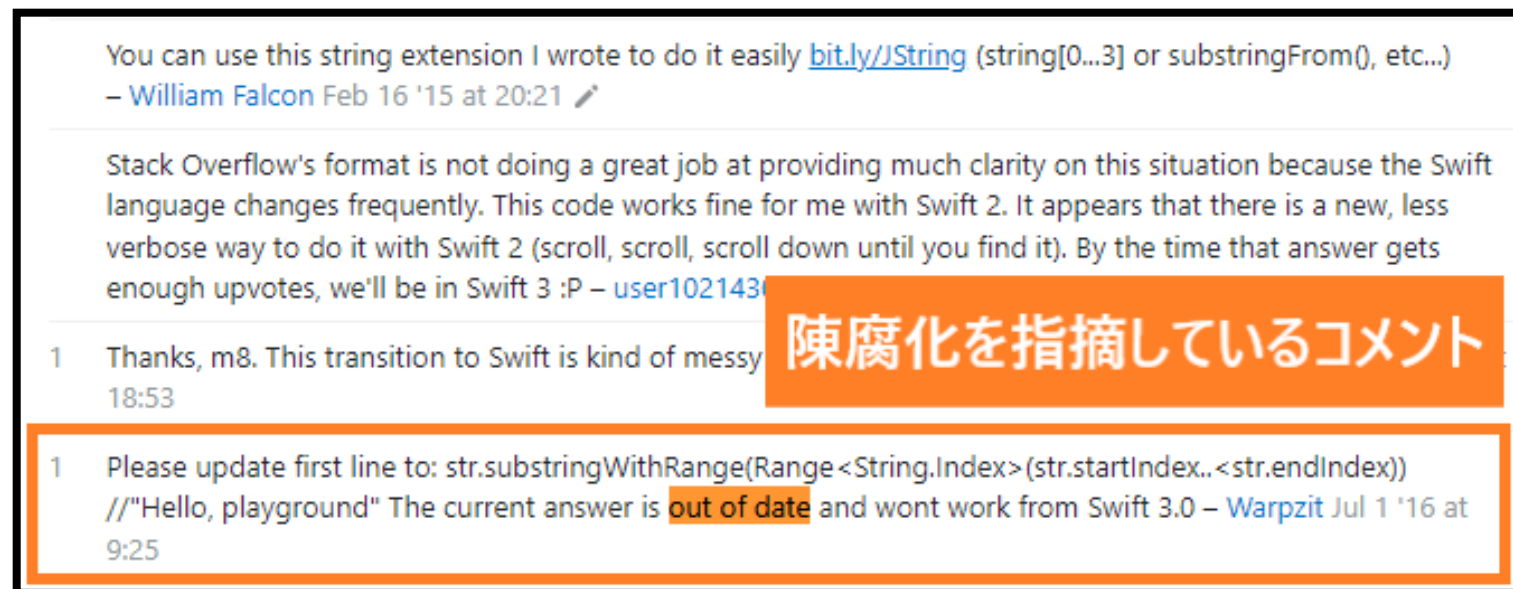
補足資料

補足. Stack Overflow | コメント機能

- ・ コメント機能とは

- 質問や回答にコメントすることができる
- ユーザーが回答の内容が陳腐化していることをコメントで指摘することがある

(例) 回答が時代遅れであり機能しないため, 回答を更新するようにコメントで提案している^[4]

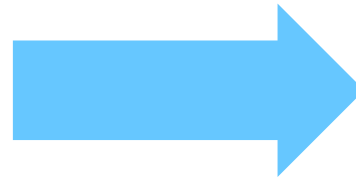


[4] https://stackoverflow.com/questions/24044851/how-do-you-use-stringsubstringwithrange-or-how-do-ranges-work-in-swift/24045156#comment63713320_24045156

補足. Stack Overflow | 投票機能

- 投票機能とは

役に立つと思われる
投稿に投票



スコアで並び変えられ
閲覧されやすくなる

スコア = プラス投票数 - マイナス投票数

```
var str = "hello, playground"
str.substringWithRange(Range<String.Index>(start: str.startIndex, end: str.startIndex.advancedBy(1)))
```

To change the start and end index, use `advancedBy(n)`.

```
var str = "Hello, playground"
```

投票機能の例^[5]

[5] <https://stackoverflow.com/questions/24044851/how-do-you-use-stringsubstringwithrange-or-how-do-ranges-work-in-swift/24045156#24045156>

補足. Stack Overflow | 投稿後の更新

- 投稿後も、ユーザーは投稿の更新を行うことができる
- 行われた更新はリビジョン履歴として記録されている



リビジョン履歴の例^[6]

[6] <https://stackoverflow.com/posts/24045156/revisions>