

Stack Overflowの 陳腐化した回答の自動検出

研究発表
2021/08/11

発表者: 井町 瑠貴

背景

StackOverflowでは、回答の陳腐化が問題となっている

回答の陳腐化の検出は、コメントや投票などの人の力で行われており、先行研究でも人力で行われている^[1]

[1] H. Zhang, S. Wang, T. -H. Chen, Y. Zou and A. E. Hassan, "An Empirical Study of Obsolete Answers on Stack Overflow," in *IEEE Transactions on Software Engineering*, vol. 47, no. 4, pp. 850-862, 1 April 2021

目的とアプローチ

- ・目的

陳腐化した回答を自動で検出する

- ・アプローチ

回答のUpVote(DownVote)数の推移に注目し,
陳腐化した回答の特徴を調査する

調査対象

- ・陳腐化した回答を自動収集することができないので、先行研究のデータセット^[2]を利用している

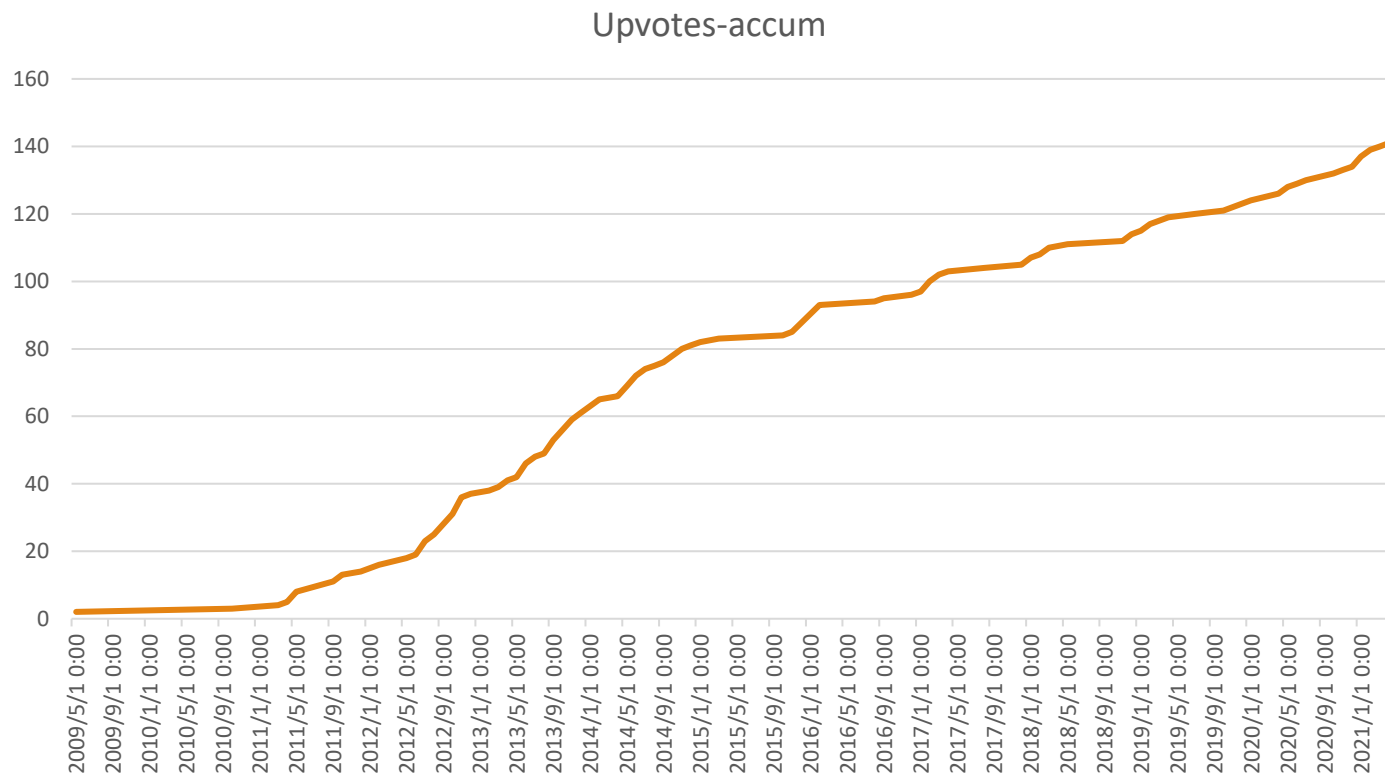
[2] https://github.com/SAILResearch/replication-obsolete_answers_SO

調査対象

- Stack Exchange Data Explorer^[3]を用いて、
陳腐化した回答のvote数の推移を取得した
 - 投稿時は正しいものであったが、次第に
古くなってしまった回答を対象に考える
- 陳腐化の指摘が回答の投稿から1日以内の
回答は調査対象から除外

[3] <https://data.stackexchange.com/>

現時点での調査結果



現時点での調査結果

vote数の増加率は、回答の陳腐化をきっかけに変動していると考えられる



vote数の増加率から回答が陳腐化しているかどうかを判断することができるのではないかと

今後の予定

- 陳腐化を判断する具体的なvote数の増加率はいくつと言えるかを調べる
- vote数の増加率の推移と陳腐化の指摘があったときを比較する

補足

Stack Exchange Data Explorer

[Home](#)[Queries](#)[Users](#)[Compose Query](#)

Viewing Query

[edit description](#)

Q&A for professional and
enthusiast programmers

```
1 DECLARE @AID int = ##PostID##
2
3 SELECT
4     DATEFROMPARTS(YEAR(v.CreationDate), MONTH(v.CreationDate), 1) AS [Year-Month],
5     v.postid AS [Post Link],
6     COUNT(CASE WHEN v.VoteTypeId = 2 THEN 1 ELSE NULL END) AS [Upvotes],
7     COUNT(CASE WHEN v.VoteTypeId = 3 THEN 1 ELSE NULL END) AS [Downvotes]
8
9 FROM Posts p
10     INNER JOIN
11         Votes v ON v.postid = p.id
12     INNER JOIN
13         Posts q ON q.id = p.parentid
14
15 WHERE p.id = @AID
16        AND v.VoteTypeId IN (2, 3) -- Limit to Up/ Downvotes
17        AND (v.CreationDate >= '##Date1?2016-07-01##') AND (v.CreationDate <= '##Date2?2021-07-
18        --AND (q.ViewCount >= '##Date1?2016-07-01##') AND (q.ViewCount <= '##Date2?2021-07-01##
19
20 GROUP BY DATEFROMPARTS(YEAR(v.CreationDate), MONTH(v.CreationDate), 1), v.postid,
21           q.ViewCount
22
23 ORDER BY DATEFROMPARTS(YEAR(v.CreationDate), MONTH(v.CreationDate), 1) ASC
24
```

Database Schema	
Posts	
Id	int
PostTypeId	tinyint
AcceptedAnswerId	int
ParentId	int
CreationDate	datetime
DeletionDate	datetime
Score	int
ViewCount	int
Body	nvarchar (max)
OwnerUserId	int
OwnerDisplayName	nvarchar (40)
LastEditorUserId	int
LastEditorDisplayName	nvarchar (40)
LastEditDate	datetime
LastActivityDate	datetime
Title	nvarchar (250)

Stack Exchange Data Explorer

Enter Parameters

PostID

Date1

Date2

Run Query

Cancel

Options: ☐ Text-only results ☐ Include execution plan

Switch to meta site



Results

Messages

 [Download CSV](#)

Year-Month ▲	Post Link ▲	Upvote...	Downvotes ▲
2011-03-01 00:00:00	Are HTTP headers case-sensitive?	2	0
2011-10-01 00:00:00	Are HTTP headers case-sensitive?	1	0
2011-11-01 00:00:00	Are HTTP headers case-sensitive?	1	0
2012-01-01 00:00:00	Are HTTP headers case-sensitive?	1	0
2012-02-01 00:00:00	Are HTTP headers case-sensitive?	2	0
2012-04-01 00:00:00	Are HTTP headers case-sensitive?	1	0
2012-05-01 00:00:00	Are HTTP headers case-sensitive?	1	0
2012-08-01 00:00:00	Are HTTP headers case-sensitive?	2	0
2012-10-01 00:00:00	Are HTTP headers case-sensitive?	2	0