

Self-Admitted Technical Debt in R Packages: An Exploratory Study

Melina Vidoni

MSR 2021

発表者：松尾 和輝

背景と目的

- 背景

Java, C++, .Netなどの伝統的なOOPで記述されたプロジェクトのSATDを研究したことが多い.

→R言語についての研究は少なかった.

- 目的

R言語におけるSATDの取り扱いについて調査を行う.

用語

- **Technical Debt(TD)**

時間がかかるよりよい方法を使わずに，今すぐ簡単な解決策を選んだことで生じる追加的な手直しの暗黙のコストを反映した比喻

- **Self-Admitted TD(SATD)**

開発者がソースコード上にコメントでTDであることを明記したTD

研究手法

- Githubから503個のRパッケージを抽出し，16万4千件以上のコメントを手作業で分析してデータセットを作成.
- Rパッケージの作者やメンテナに対し，アンケートを実施.
- コメントを「有用性」と「明確さ」の2つの観点で分類して，品質を調査.

リサーチクエスチョン

- RQ1 : RパッケージにはどれだけSATDが存在するのか?
- RQ2 : Rパッケージのコメントの質は?
- RQ3 : どのようなタイプのTDが多くあるのか?
- RQ4 : 開発者はSATDを認識しているか?

RQ1 : SATDの数

- 164262行のコメントを手作業で分類
- SATDは4962行(約3%)で、これは他の伝統的なOOPで見られた割合とほぼ同じであった.

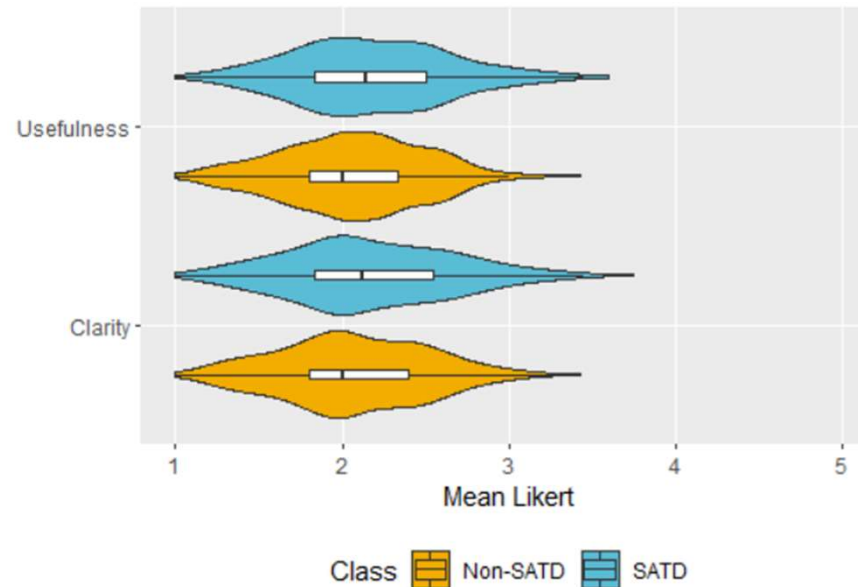
TABLE I
FILTERING OF COMMENTS, DIVIDED BY MULTI AND SINGLE LINE.

Type	Multilines		Single-lines	
	# Comments	%	# Comments	%
Explanation	17170	10.45	91820	55.89
Code	7846	4.77	9832	5.98
Title	5222	3.17	15432	9.39
SATD	2043	1.24	2919	1.77
Credits	2563	1.56	978	0.59
Exceptions	582	0.35	2252	1.37
Block Closures	237	0.14	1183	0.72
Return Notes	365	0.22	1593	0.96
Example Data	315	0.19	484	0.29
Fixes	636	0.38	790	0.48

RQ2 : コメントの品質

- アンケートにより，コメントの品質を評価.
- **SATD**コメントも非**SATD**コメントも，有用性と明確さが「悪い」という低品質なものだった.

Fig. 2. Distribution of mean Likert values for all sampled comments, in terms of Usefulness and Clarity, for each block of comments (SATD and non-SATD).



RQ3 : TDのタイプ

- 4962行のSATDコメントを手作業で分類.
- Code, Test, Defect Debtの順に多く, この三つだけで半数近くを占めている.

TABLE II
TYPES OF TD FOUND IN SATD COMMENTS.

TD Type	# Lines	% Total
Code	2015	40.6%
Test	784	15.8%
Defect	693	13.97%
Requirements	355	7.15%
Architecture	291	5.86%
Algorithm	276	5.56%
Design	221	4.45%
Build	160	3.22%
Usability	71	1.4%
Documentation	53	1.07%
People	21	0.42%
Versioning	21	0.42%

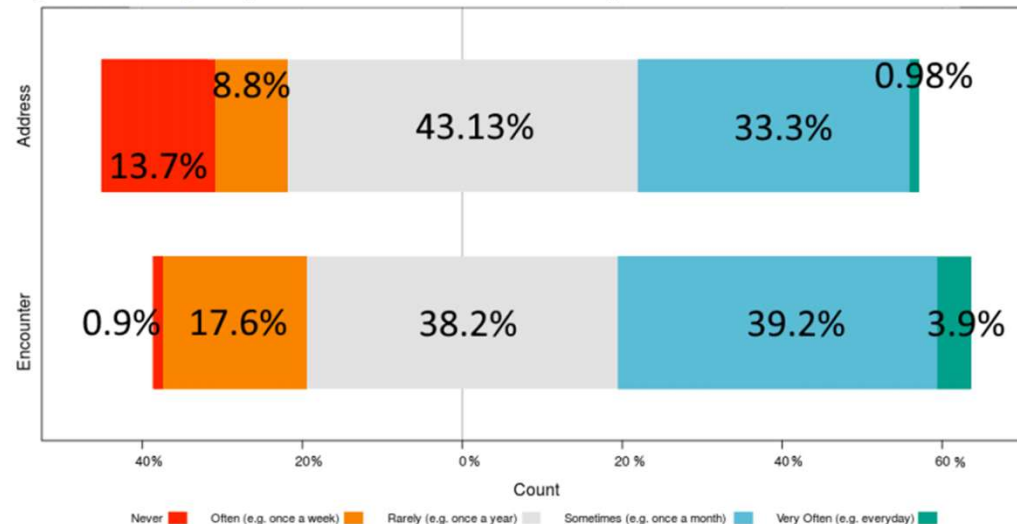
RQ3 : TDのタイプ

- SATDコメントの頻出パターンの調査.
- よく見られたのが「fixme」「hack」「ugly」「stupid」であった.
- また自動テストの実行を停止させることを目的としたSATDコメントが多数存在していた.

RQ4 : 開発者の視点

- Rパッケージの開発者及びメンテナにTDの遭遇及び対処に関する質問を行った。
- 43%が、月に一度SATDを発見するが、同じ頻度でそのSATDに対処しようと考えているのが34%以下であることが分かる。

Fig. 3. Survey responses on how often developers encounter or address SATD.



RQ4：開発者の視点

- SATDが混入する理由を質問したところ,
 - 自分への戒めとして（コメントの約28.4%）
 - 将来の計画のため（16.7%近く）
 - 手っ取り早く解決策を得るため（12.7%をわずかに上回る）
 - 知識不足や警告のため（ともに7.8%）
 - 時間がない（約16.7%）

RQ4：開発者の視点

- SATDに対応する理由を質問したところ,
 - 機能が必要になったため (17.6%)
 - コードの品質を向上させるため (15.7%)
 - 「エラーが発生して困る」「作業を完了させなければならない」
(それぞれ10.8%)
 - どのようなSATDコメントが完了したのかわからない (12.7%)

まとめ

- RパッケージのSATDについての調査.
- 16万行以上のコメントを分析，開発者及びメンテナに対してアンケートを実施.
- 約3%がSATDであった.
- Code, Test, Defect debtが多かった.
- 「自分への戒め」としてSATDが混入され，遭遇したTDに対処することはほとんどない

所感

- どんな言語であろうとも混入するSATDの割合は変わらないのでは？
 - 結局開発者がどれだけTDを意識しているかに依存すると思う。
- 半数程度が、月に一回SATDを発見するとあるが、意外と少なく感じた。
 - 開発者がTDであることを分かっていないだけなのでは？