

Program Comprehension and Code Complexity Metrics : An fMRI Study

Norman Peitek, Sven Apel, Chris Parnin,
Anddre Brechmann, Janet Siegmund

ICSE 2021 Technical Track “Program Comprehension and Code Complexity Metrics: An fMRI Study”

発表者: 田中 智達

背景：コードの複雑さ

- 以下のコードスニペットについて考える.

```
public List<Integer> compute (int a, int b, int c) {  
    if (a > b) { int temp = b; b = a; a = temp; }  
    if (a > c) { int temp = c; c = a; a = temp; }  
    if (b > c) { int temp = c; c = b; b = temp; }  
    return Arrays.asList(a, b, c);  
}
```

- どう複雑さを評価するか
 - ソフトウェア工学の研究ではさまざまな構造特性に基づいて計算された、複雑さを測定する基準がいくつかある[2].

背景：コードの複雑さ

- 複雑さを測定する基準がいくつかある[2].
 - コード行数(**LOC**が代表)：7
 - **Halstead**の複雑度^[3](語彙数)：21
 - **McCabe**のサイクロマティック複雑度^[4](制御フロー)：4
 - Beyerの**DepDegree**^[5](データフロー)：31
 - 値はかなり異なっており，共通の尺度でないので一概には判断できない.
- コードメトリクスは人間の認知や行動を捉えるには限界がある.

[2] A. S. N. Varela, H. G. Perez-Gonzalez, F. E. Martinez-Perez, and C. Soubervielle-Montalvo, “Source Code Metrics: A Systematic Mapping Study,” J. Systems and Software, vol. 128, pp. 164–197, 2017.

先行研究

- Scalabrinoらの研究^[16]
 - 複雑さの測定基準とコードの理解度の間は僅かな相関しかない.
- Ajamiらの研究^[17]
 - 複雑さメトリクスは人間がどうコードを処理するかを考慮していない.
- JbaraとFeitelsonらの研究^[18]
 - 複雑さの測定基準が、繰り返されるコード構造の理解しやすさの増加を見逃している.
- しかしコードの複雑さの認識, コードを理解する上で必要な認知負荷の予測などに, 複雑さメトリクスを使用することは魅力的.
- 多くの研究が行われているが, その全体像は不透明.

本研究の目的

- 著者らの過去の研究の副産物として、ニューロン活動といくつかのコード複雑性メトリクスとの間に、測定可能な相関の最初の兆候を発見^[26].
- 複雑性指標と認知過程の関係は不透明なまま.



- コード複雑度指標のばらつきは、プログラム理解時のプログラマの認知の違いを説明できるか.

[26] N. Peitek, J. Siegmund, S. Apel, C. Kastner, C. Parnin, A. Bethmann, T. Leich, G. Saake, and A. Brechmann, "A Look into Programmers' Heads," IEEE Trans. Softw. Eng., vol. 46, no. 4, pp. 442–462, 2020.

研究目標

- 以下の関係を明らかにする.
 - 複雑度メトリクスは、振る舞いと相関があるか.
 - 複雑度メトリクスは、プログラマ認知的負荷と相関があるか.
 - 複雑度メトリクスは、主観的な認識と相関があるか.

実験準備

- プログラム理解に関する先行研究 [27,28,29] から実験用にJavaコードスニペットを50件入手.
- その中からfMRI研究に最も適したものを選択.
 - 7名のプログラム経験者にスニペットの理解と意見共有を行った.
 - fMRI実験では4種類の複雑指標にバランスよく関連するスニペットを最終的に選択.
- 最後に先行研究[27,28]同様にスニペットから文脈情報をすべて排除.
 - 認知負荷に対する過去の経験の影響を低減するため.

実験準備：資料

- 研究に使用されたコードスニペットと4つの選択された複雑さのメトリックスコアと実験結果

Snippet		Complexity Metrics				Experiment Results		
		Code Size (LOC) ¹	Vocabulary (Halstead) ¹	Control Flow (McCabe) ¹	Data Flow (DepDegree) ²	Correctness (in %)	Time (in sec.)	Subjective Complexity Low → Medium → High
Loop	Average of array	17	12.64	3	17	47%	49.0	
	Contains substring	26	25.50	7	29	32%	50.2	
	Count vowels in string	19	13.00	5	22	89%	30.4	
	Greatest common divisor	24	26.63	5	33	47%	50.0	
	h index	21	16.25	4	20	53%	46.0	
	Length of last word	22	18.58	8	17	58%	43.3	
	Palindrome check	17	16.72	4	17	79%	38.4	
	Square root of array	23	39.83	5	27	68%	40.1	
Recursion	Binary to decimal	17	16.75	4	10	68%	42.3	
	Cross sum	12	14.66	3	4	84%	27.6	
	Factorial	12	16.50	3	4	95%	22.3	
	Fibonacci variation	12	10.88	3	4	84%	35.6	
	Power	17	15.38	4	8	79%	33.3	
If/Else	Contains yes or no	22	6.13	6	7	95%	23.4	
	Hurricane check	17	9.00	7	13	100%	21.5	
	Sort four elements	17	30.30	7	61	79%	41.4	

¹ <https://github.com/BasLeijdekkers/MetricsReloaded/>

² <https://www.sosy-lab.org/~dbeyer/DepDigger/>

fMRI実験(1/2)

- 脳が活性化・非活性化した際に変化する酸素濃度の変化を測定するもの.
- 脳のDMNに関する研究^[34,35]により, タスク中のDMNの不活性化のレベルは認知的負荷の強度に関連すると示唆^[36].
 - これによりタスク中の認知負荷を判断できる.
- 参加者はプログラミング経験は中程度であった.

[34] M. Raichle, A. MacLeod, A. Snyder, W. Powers, D. Gusnard, and G. Shulman, “A Default Mode of Brain Function,” *Proc. Nat’l Academy of Sciences*, vol. 98, no. 2, pp. 676–682, 2001.

[35] L. Mak, L. Minuzzi, G. MacQueen, G. Hall, S. Kennedy, and R. Milev, “The Default Mode Network in Healthy Individuals: A Systematic Review and Meta-Analysis,” *Brain Connectivity*, vol. 7, no. 1, pp. 25–33, 2017.

fMRI実験(2/2)

- まず以下のタスクを3スニペット分行う。
 1. スニペットを実行したらどうなるかを判断(最大1分)
 2. 反射的思考を減らすためd2課題^[40]を実行(10秒)
 3. 休息(30秒)
- その後別のスニペットを見て、開始のカッコ部をマーク。
- 以上を参加者が16のスニペットを完了するまで繰り返す。



データ収集

- 先程のタスク中にfMRIスキャンを行い，脳の活動データを収集.
- 参加者の行動(反応時間・反応の正しさ)と参加者の行動(回答時間・回答正答率)を観察.
- 一連のタスク終了後にスニペットを再度見せ，主観的な複雑さを事後インタビューで記録.

結果

- 従属変数のKendallの τ と説明される分散 r^2 . セルが濃いほど相関が強い.
 - なし($\tau < 0.1$)
 - 中($0.3 < \tau < 0.5$)
 - 弱($0.1 < \tau < 0.3$)
 - 強($0.5 < \tau$)

		LOC	Complexity Metrics			BA 6	Activation			Deactivation		Subjective Complexity
			Halstead	McCabe	DepDegree		BA 21	BA 39	BA 44/45	BA 31	BA 32	
Complexity Metrics	LOC											.16 (.04)
	Halstead	.32										.20 (.04)
	McCabe	.57	.25									-.07 (.01)
	DepDegree	.59	.50	.49								.16 (.01)
Activation	BA 6	.26 (.05)	.38 (.20)	.04 (.00)	.32 (.11)							.20 (.09)
	BA 21	.43 (.39)	.32 (.27)	.09 (.04)	.41 (.24)							.18 (.09)
	BA 39	.17 (.10)	.40 (.18)	.07 (.01)	.36 (.21)							.17 (.10)
	BA 44/45	.15 (.04)	.17 (.06)	-.04 (.00)	.22 (.09)							.22 (.06)
Deactivation	BA 31	-.30 (.21)	-.30 (.11)	.05 (.00)	-.24 (.04)							-.44 (.44)
	BA 32	-.39 (.24)	-.42 (.22)	.04 (.00)	-.29 (.04)							-.69 (.69)
Behavioral Data	Correctness	-.46 (.29)	-.45 (.26)	-.09 (.02)	-.41 (.22)	-.63 (.53)	-.63 (.67)	-.58 (.47)	-.34 (.13)	.59 (.58)	.71 (.62)	-.77 (.71)
	Time	.22 (.10)	.24 (.09)	.06 (.00)	.26 (.07)	.29 (.19)	.22 (.10)	.31 (.22)	.24 (.13)	-.22 (.11)	-.21 (.10)	.34 (.18)

結果(メトリクスと振る舞い)

- 正答率とMcCabe以外のメトリクスは中程度の相関。
- 回答時間とMcCabe以外のメトリクスは弱い相関。

		LOC	Complexity Metrics			BA 6	Activation			Deactivation		Subjective Complexity
			Halstead	McCabe	DepDegree		BA 21	BA 39	BA 44/45	BA 31	BA 32	
Complexity Metrics	LOC											.16 (.04)
	Halstead	.32										.20 (.04)
	McCabe	.57	.25									-.07 (.01)
	DepDegree	.59	.50	.49								.16 (.01)
Activation	BA 6	.26 (.05)	.38 (.20)	.04 (.00)	.32 (.11)							.20 (.09)
	BA 21	.43 (.39)	.32 (.27)	.09 (.04)	.41 (.24)							.18 (.09)
	BA 39	.17 (.10)	.40 (.18)	.07 (.01)	.36 (.21)							.17 (.10)
	BA 44/45	.15 (.04)	.17 (.06)	-.04 (.00)	.22 (.09)							.22 (.06)
Deactivation	BA 31	-.30 (.21)	-.30 (.11)	.05 (.00)	-.24 (.04)							-.44 (.44)
	BA 32	-.39 (.24)	-.42 (.22)	.04 (.00)	-.29 (.04)							-.69 (.69)
Behavioral Data	Correctness	-.46 (.29)	-.45 (.26)	-.09 (.02)	-.41 (.22)	-.63 (.53)	-.63 (.67)	-.58 (.47)	-.34 (.13)	.59 (.58)	.71 (.62)	-.77 (.71)
	Time	.22 (.10)	.24 (.09)	.06 (.00)	.26 (.07)	.29 (.19)	.22 (.10)	.31 (.22)	.24 (.13)	-.22 (.11)	-.21 (.10)	.34 (.18)

結果(メトリクスと認知負荷)

- BA31, BA32はデフォルトモードネットワーク(DMN)に属している.
- McCabe以外は小～中程度の相関.

		LOC	Complexity Metrics			Activation				Deactivation		Subjective Complexity
			Halstead	M McCabe	DepDegree	BA 6	BA 21	BA 39	BA 44/45	BA 31	BA 32	
Complexity Metrics	LOC											.16 (.04)
	Halstead	.32										.20 (.04)
	M McCabe	.57	.25									-.07 (.01)
	DepDegree	.59	.50	.49								.16 (.01)
Activation	BA 6	.26 (.05)	.38 (.20)	.04 (.00)	.32 (.11)							.20 (.09)
	BA 21	.43 (.39)	.32 (.27)	.09 (.04)	.41 (.24)							.18 (.09)
	BA 39	.17 (.10)	.40 (.18)	.07 (.01)	.36 (.21)							.17 (.10)
	BA 44/45	.15 (.04)	.17 (.06)	.04 (.00)	.22 (.00)							.22 (.06)
Deactivation	BA 31	-.30 (.21)	-.30 (.11)	.05 (.00)	-.24 (.04)							-.44 (.44)
	BA 32	-.39 (.24)	-.42 (.22)	.04 (.00)	-.29 (.04)							-.69 (.69)
Behavioral Data	Correctness	-.46 (.29)	-.45 (.26)	-.09 (.02)	-.41 (.22)	-.63 (.53)	-.63 (.67)	-.58 (.47)	-.34 (.13)	.59 (.58)	.71 (.62)	-.77 (.71)
	Time	.22 (.10)	.24 (.09)	.06 (.00)	.26 (.07)	.29 (.19)	.22 (.10)	.31 (.22)	.24 (.13)	-.22 (.11)	-.21 (.10)	.34 (.18)

結果(メトリクスと主観的な認識)

- メトリクスとの相関は小さいが、正答率、認知負荷とは中程度以上の相関がある。

		LOC	Complexity Metrics			BA 6	Activation			Deactivation		Subjective Complexity
			Halstead	McCabe	DepDegree		BA 21	BA 39	BA 44/45	BA 31	BA 32	
Complexity Metrics	LOC											.16 (.04)
	Halstead	.32										.20 (.04)
	McCabe	.57	.25									-.07 (.01)
	DepDegree	.59	.50	.49								.16 (.01)
Activation	BA 6	.26 (.05)	.38 (.20)	.04 (.00)	.32 (.11)							.20 (.09)
	BA 21	.43 (.39)	.32 (.27)	.09 (.04)	.41 (.24)							.18 (.09)
	BA 39	.17 (.10)	.40 (.18)	.07 (.01)	.36 (.21)							.17 (.10)
	BA 44/45	.15 (.04)	.17 (.06)	-.04 (.00)	.22 (.09)							.22 (.06)
Deactivation	BA 31	-.30 (.21)	-.30 (.11)	.05 (.00)	-.24 (.04)							-.44 (.44)
	BA 32	-.39 (.24)	-.42 (.22)	.04 (.00)	-.29 (.04)							-.69 (.69)
Behavioral Data	Correctness	-.46 (.29)	-.45 (.26)	-.09 (.02)	-.41 (.22)	-.63 (.53)	-.63 (.67)	-.58 (.47)	-.34 (.13)	.59 (.58)	.71 (.62)	-.77 (.71)
	Time	.22 (.10)	.24 (.09)	.06 (.00)	.26 (.07)	.29 (.19)	.22 (.10)	.31 (.22)	.24 (.13)	-.22 (.11)	-.21 (.10)	.34 (.18)

結論

- McCabeは一貫して認知労力との関係は薄い.
- 一般的なコードの外観の評価を使用する複雑さの推定は、複雑さを誤って伝える可能性がある.
 - サイズに基づく測定低基準(LOC, Halstead)は認知の負荷と関係がある.
 - プログラム理解時にはDepDegreeとの関係が強くなる.
- コード複雑度メトリクスを使用し人間の認知を記述する場合、1つの指標に頼ってはいけない.

所感

- 複雑さという大まかな概念を評価することの難しさを改めて認識した.
- 脳のスキャンデータを取り扱っているのが興味深かった.