

Program Comprehension and Code Complexity Metrics: An fMRI Study

Norman Peitek <i>Leibniz Institute for Neurobiology</i> Magdeburg, Germany	Sven Apel <i>Saarland University, Saarland Informatics Campus</i> Saarbrücken, Germany	Chris Parnin <i>NC State University</i> Raleigh, North Carolina, USA
André Brechmann <i>Leibniz Institute for Neurobiology</i> Magdeburg, Germany	Janet Siegmund <i>Chemnitz University of Technology</i> Chemnitz, Germany	

*2021 IEEE/ACM 43rd International Conference on
Software Engineering (ICSE), 2021, pp. 524-536*

背景

- プログラムの複雑さを測定するためのメトリクスが数多く存在している
- 複雑度メトリクスは様々な場面で利用されている
- 単純すぎるメトリクスはプログラムの状態を正確に反映できない危険性がある[15]

既存研究

- 記号の数が多いほど認知処理の要求が高くなる
[23]
- 制御パスが増えるとルールに基づく条件付き推論に関連する脳領域が活性化する[24,25]
- データフローの依存関係の数が多いほど、ブローカ野（構文上の作業メモリ）が活性化する[22]

研究の動機

- プログラム理解と脳の活動の関係について多くの研究がなされてきた
- メトリクスと脳の活動の関係は既存研究の副産物として明らかになった
- 様々な複雑度メトリクスとプログラム理解中の脳の活動の間関係は不明なままである

目的・アプローチ

- 目的

- 既存の複雑度メトリクスはプログラム理解の難易度を反映できているか明らかにする

- アプローチ

- 複雑度メトリクスとプログラム理解中の脳の働きの間
の相関をとる

実験目標

- 以下に示す関係を相関分析で明らかにする
 - 複雑度メトリクスと振る舞い
 - 複雑度メトリクスと認知的負荷
 - 複雑度メトリクスと主観的複雑度

実験準備

- 既存研究で用いられた**50件**の実験用コード片を用意
- **7人**の開発経験者にコード片の内容を理解してもらい、意見を集める
- **fMRI**を用いた実験に適していると思われるコード片を選択
- 余計な認知負荷を避けるためにコード片の周辺情報を除去

実験に用いるコード片

TABLE I: Code snippets used in the study with four selected complexity metric scores and experimental results: behavioral data (correctness, time) and subjective complexity. A higher intensity of a cell's background color indicates a higher complexity. The histogram plots show the distribution of subjective complexity; skewness to the right represents higher subjective complexity.

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を用いた実験

- fMRI(functional Magnetic Resonance Imaging)
 - 脳内の酸素レベルの変化を検出し，活性化（非活性化）した脳領域を特定する装置
- 被験者のプログラム理解中の脳の活動を記録
- タスク中のDMNの非活性化のレベルは認知的負荷の強度に関連している[36]
 - DMN(Default Mode Network)：意識的な活動をしていない脳の状態

fMRI スキャン中のタスク

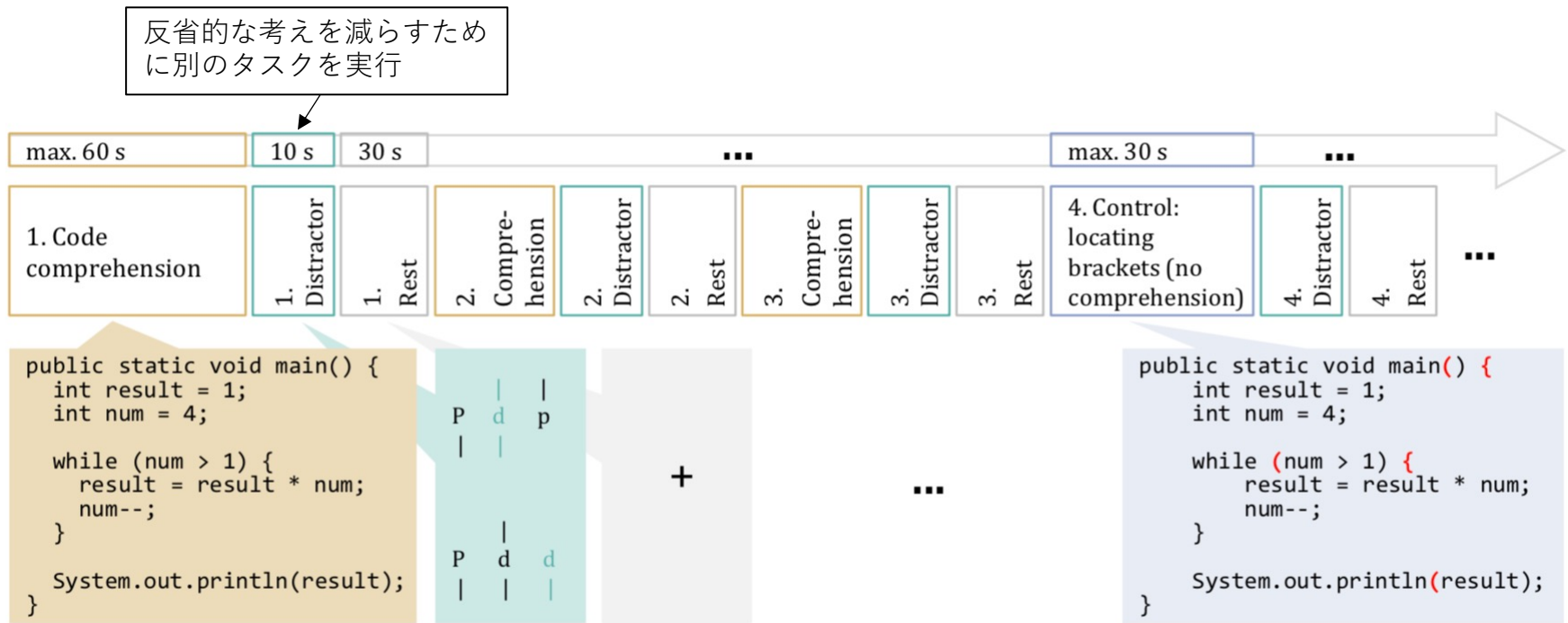


Fig. 1: Illustration of one (out of five) experiment trials.

データ収集

- タスク中のfMRIスキャン結果から脳の活動データを収集
- タスクにかかった時間と答案の正誤（振る舞い）を収集
- タスク後にもう一度同じコード片を見てもらい、主観的な複雑度をインタビューで収集

プログラム理解中の脳の(非)活性

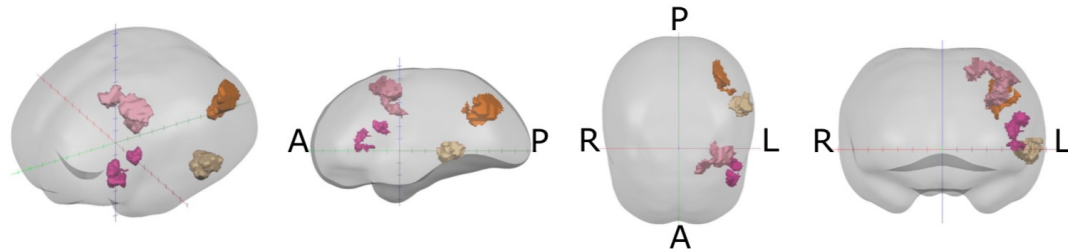


Fig. 2: Visualization of the activated brain areas (all in the left hemisphere): **BA 6** (4315 voxel, TAL: -35, 10, 47), **BA 21** (3306 voxel, TAL: -57, -39, 1), **BA 39** (3527 voxel, TAL: -37, -65, 35), and **BA 44/45 (Broca's area)** (1618 voxel, TAL: -49, -22, 14). A: anterior, P: posterior, L: left, R: right.

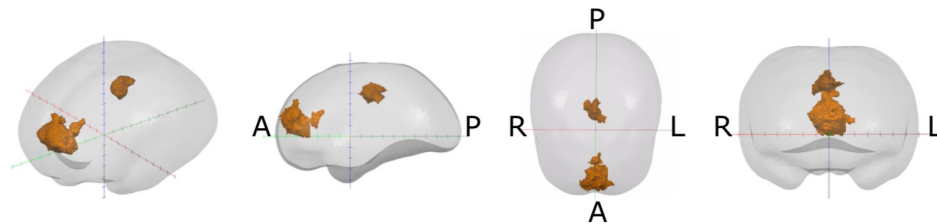


Fig. 3: Visualization of the deactivated brain areas: **BA 32** (10'651 voxels, TAL: -1, 51, 14) and **BA 31** (2672 voxels, TAL: 3, -20, 40), which both are part of the default mode network. A: anterior, P: posterior, L: left, R: right.

結果：メトリクス，振る舞い

TABLE II: Kendall's τ and the explained variance (r^2 , in brackets) of the dependent variables. A darker cell shading indicates a stronger correlation: none ($\tau < 0.1$), small ($0.1 < \tau < 0.3$), medium ($0.3 < \tau < 0.5$), and strong ($0.5 < \tau$) [49].

		Complexity Metrics				Activation				Deactivation		Subjective Complexity
		LOC	Halstead	McCabe	DepDegree	BA 6	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)

ケンドールの順位相関係数 τ で各データ間の関係を分析

- 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

結果：メトリクス, 脳活動

TABLE II: Kendall's τ and the explained variance (r^2 , in brackets) of the dependent variables. A darker cell shading indicates a stronger correlation: none ($\tau < 0.1$), small ($0.1 < \tau < 0.3$), medium ($0.3 < \tau < 0.5$), and strong ($0.5 < \tau$) [49].

		Complexity Metrics				Activation				Deactivation		Subjective Complexity
		LOC	Halstead	McCabe	DepDegree	BA 6	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)

ケンドールの順位相関係数 τ で各データ間の関係进行分析

- 活性化領域とMcCabe以外のメトリクスは中程度または弱い相関
- 非活性化領域も同様

結果：振る舞い， 脳活動

TABLE II: Kendall's τ and the explained variance (r^2 , in brackets) of the dependent variables. A darker cell shading indicates a stronger correlation: none ($\tau < 0.1$), small ($0.1 < \tau < 0.3$), medium ($0.3 < \tau < 0.5$), and strong ($0.5 < \tau$) [49].

		Complexity Metrics				Activation				Deactivation		Subjective Complexity
		LOC	Halstead	McCabe	DepDegree	BA 6	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)

ケンドールの順位相関係数 τ で各データ間の関係进行分析

- 正答率と脳活動は強い相関
- 回答時間と脳活動は中または弱い相関

結果：主観的複雑度，各データ

TABLE II: Kendall's τ and the explained variance (r^2 , in brackets) of the dependent variables. A darker cell shading indicates a stronger correlation: none ($\tau < 0.1$), small ($0.1 < \tau < 0.3$), medium ($0.3 < \tau < 0.5$), and strong ($0.5 < \tau$) [49].

		Complexity Metrics				Activation				Deactivation		Subjective Complexity
		LOC	Halstead	McCabe	DepDegree	BA 6	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)

ケンドールの順位相関係数 τ で各データ間の関係进行分析

- McCabe以外のデータで相関がある
- 特に非活性脳状態と振る舞いに強い相関がある

議論：非活性領域と認知的負荷

- 非活性化レベルは認知的負荷の尺度[36]
- LOCとHalstedで中程度の相関
 - 文字数，語彙の多さは認知的負荷を大きくする

議論：活性領域と認知プロセス

- LOCとB21の中程度の相関
 - B21は言語理解，プログラム理解中の意味処理に関連する
 - LOCを増やすだけで識別子と記号の認知処理が増える
- Halsted, DepDegreeとB6, B21, B39の中程度の相関
 - B6は注意力と作業記憶が必要なときに活性化
 - B39は複雑な認知作業時に活性化され，情報を統合するハブとして機能
 - 語彙数とデータ依存関係の認知作業時に幅広い処理ネットワークを要する

議論：さらなる複雑度メトリクスの調査

TABLE III: Kendall's τ correlation between brain activation and the unique, differentiating top 26 of the 37 explored complexity metrics. The metric's text color indicates its class (size, vocabulary, control flow, data flow, other). The cell shading highlights strong correlations (cf. Table II). All metrics, raw data, and results are available on our supplementary Web site.

Metric	BRANCH	NP	NBD	D	LOOP	EXP	N	V	CRCT	MAINT	iv(G)	RLOC	n	CALL	RLBTY	E	IF_NEST	CogCompl	CONTR	STAT	ev(G)	v(G)	EXEC	RETURN	CALLED	CDENS
BA 6	.50	.40	.36	.51	.36	.45	.41	.41	.41	.40	.34	.37	.37	.27	.34	−0.06	.31	.30	.27	.29	.23	.21	.16	−0.37	−0.18	.09
BA 21	.56	.27	.54	.41	.48	.34	.36	.36	.35	.33	.38	.37	.35	.37	.27	−0.07	.27	.24	.21	.23	.27	.10	.10	.10	.14	−0.33
BA 39	.29	.56	.29	.45	.22	.32	.34	.26	.31	.33	.10	.32	.17	.21	.27	−0.01	.31	.14	.13	.25	.19	.12	.19	−0.18	.00	−0.07
BA 44/45	.16	.50	.20	.41	.12	.16	.18	.14	.15	.17	.06	.16	.12	.16	.11	.03	.27	.04	.01	.07	.08	.00	.02	−0.08	.13	−0.05

- MetricsReloadedとSONARQUBEで用いられている39種のメトリクスの値と脳の活性度とで相関を計測
 - シンプルな制御フローメトリクスと語彙メトリクスで強い相関が得られた
- シンプルなメトリクスを組み合わせた方がプログラム理解における認知的負荷を効果的に測定できる

まとめ

- 複雑度メトリクスとプログラム理解の関係をfMRIを用いて調査
- McCabe以外のメトリクスで弱から中程度の相関があることがわかった
- シンプルなメトリクスがプログラム理解の難易度を測定する上で効果的であることがわかった

所管

- 脳の状態をデータとして扱った点が興味深かった
- 周辺の情報を含んだ状態のプログラム理解では脳の状態はどう変わってくるのか気になった