

Game Software Engineering: A Controlled Experiment Comparing Automated Content Generation Techniques

Mar Zamorano López, África Domingo, Carlos Cetina, Federica
Sarro, Thu 24 Oct 2024.

ESEIW 2024 ESEM Technical Papers

背景

論文の日本語訳

ゲームソフトウェアエンジニアリング：自動コンテンツ生成技術を比較する制御実験

論文概要

ゲーム作成にソフトウェアは不可欠だがデザイン，楽しさ，難易度の評価のプラクティスが確立されていない．そこでコンテンツを自動生成する最新のアプローチに対し，ソフトウェア工学で確率されている手法で実証的評価を行った．

ゲームとソフトウェア

ゲーム開発の課題

課題の1つとして新しいコンテンツの高い需要が挙げられるが、これは他のソフトウェア作成よりも特に時間、労力、コストがかかり、エラーも発生しやすい。

PCGとRCG

課題に対応するために、コンテンツ生成を自動化する手段として様々な研究が進んでいるソフトウェアの自動生成手法。

今回はこの2つのどちらがより優れたコンテンツを生成するのかを比較する。

PCG と RCG

PCG（プロシージャル・コンテンツ生成）

ランダムに組み合わせたアルゴリズムに基づいて、データを生成する手法。利点としてファイルサイズを小さくできる、大量のコンテンツを用意できることなどが挙げられる。

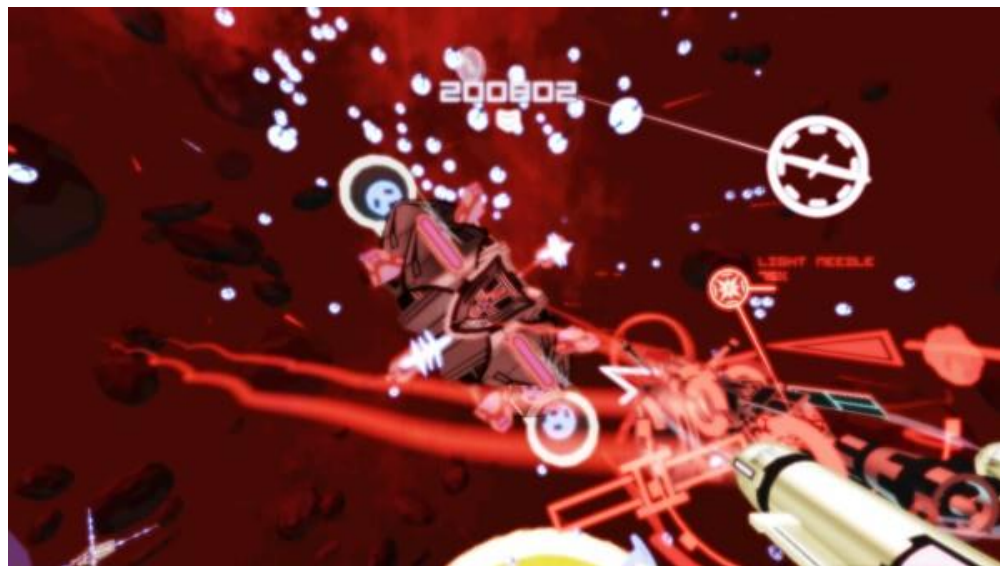
RCG（再利用ベースのコンテンツ生成）

ゲーム内の機能を再利用してデータを生成する手法。具体的には提供側から機能を選び、受け手側となる他のコンテンツに移植するという手法。

評価

評価の対象

PlayStation 4の「Kromaia」というゲームを対象とし，参加者にはPCG及びRCGで生成されたゲーム内コンテンツ（ラスボス戦）をプレイしてもらい，その体験からゲームを評価してもらった．



評価

評価の指標

ビデオゲーム特有の品質指標である**難易度**，**デザイン**，**楽しさ**，**没入感**に基づいていて，プレイヤー向けと開発者向けの計3回の実験を行った．

評価方法

3回のセッションに分けてプレイヤ目線，製作者目線の両方の立場からPCG版とRCG版をプレイし，2つを比較してもらった．

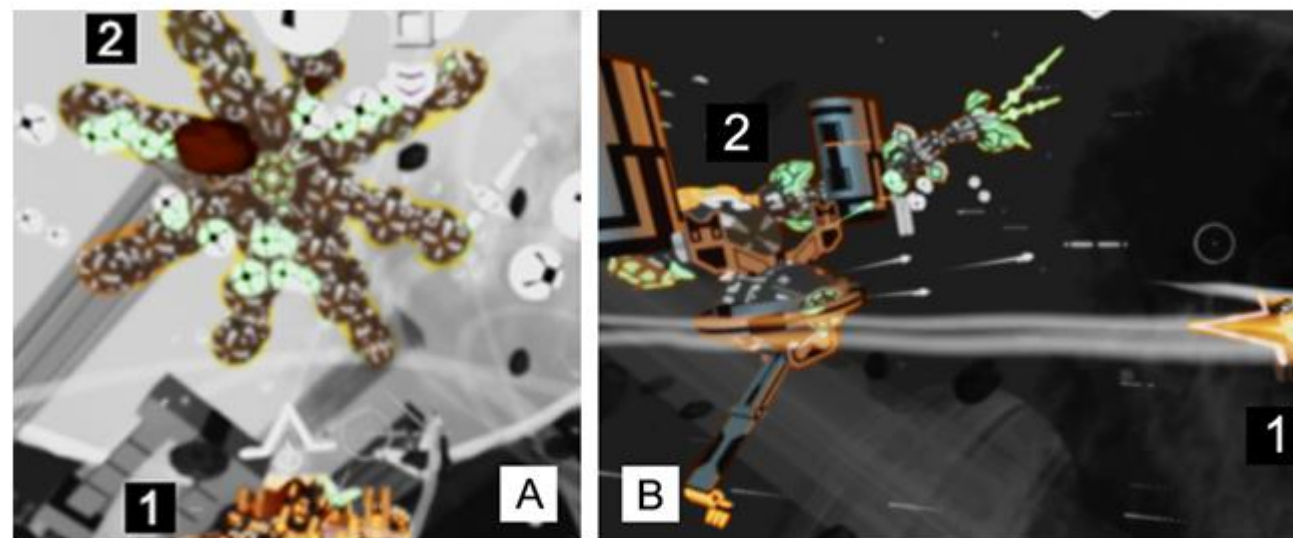


Figure 1: (A) PCG boss. (B) RCG boss.

評価分析

分析方法

繰り返し測定に起因する相関データを扱え、ブロック変数の影響を調査できる**線形混合モデル（LMM）**を統計解析に用いた。

また応用変数（RV）とブロック関数を定義するために、**主成分分析（PCA）を2度実施**して評価結果を分析した。

- ・ ブロック変数の例

参加者のプロフィール， ゲーム開発経験， etc

- ・ 応用変数の例

勝率， 難易度， 楽しさ， デザイン， 没入感， etc

評価結果

分析，抽出結果

PCAの結果， 4つのコンポーネントが抽出された．

- 1: デザイン，楽しさに関する質問への回答のまとめ
- 2: 被験者によるコメントに関する応用変数のまとめ
- 3: 勝率と難易度のまとめ
- 4: ゲームのプレイ時間のまとめ

※没入感についての回答は全てに含まれており，明確に要因を定義できなかった．

→まとめられた成分は同様の観点から結果が得られることが分かる．

評価結果

各応用変数の平均値と標準偏差

	Technique	Profile		Developing Games		Gamer Profile			Group	
		Players	Developers	More than 10 h/Week	Less than 10 h/Week	Target Audience	Neutral	Non Target Audience	G1 (RCG-PCG)	G2 (PCG-RCG)
Game Duration	RCG	4.24±2.85	4.18±3.23	4.38±1.52	4.05±3.27	4.57±1.95	4.57±4.36	3.22±2.22	5.33±2.77	4.16±2.93
	PCG	2.01±1.76	2.19±2.02	1.54±0.55	2.39±2.06	1.34±0.68	1.58±0.54	2.01±1.38	2.13±2.34	2.21±2.28
	All	3.12±2.61	3.18±2.85	2.96±1.83	3.22±2.83	2.95±2.18	3.07±3.33	2.62±1.92	3.73±3	3.19±2.77
Won rate	RCG	0.32±0.37	0.33±0.39	0.29±0.33	0.3±0.39	0.36±0.35	0±0	0.25±0.32	0.5±0.39	0.41±0.38
	PCG	0.71±0.39	0.7±0.4	0.73±0.4	0.6±0.42	0.9±0.26	0±0	0.68±0.36	0.95±0.16	0.76±0.4
	All	0.52±0.43	0.52±0.43	0.51±0.42	0.45±0.43	0.63±0.41	0±0	0.46±0.4	0.72±0.37	0.59±0.42
Boss Difficulty	RCG	5.41±1.68	5.28±1.59	5.75±1.91	5.39±1.73	5.44±1.63	2.8±1.48	5.86±1.42	5.61±1.38	5.48±1.31
	PCG	3.05±2.09	2.84±2	3.58±2.31	3.61±2.25	2.06±1.34	6.2±1.79	3.43±1.96	1.72±0.9	2.96±2.16
	All	4.23±2.23	4.06±2.17	4.67±2.35	4.5±2.18	3.75±2.26	4.5±2.37	4.64±2.09	3.67±2.28	4.22±2.18
Design	RCG	4.72±1.66	4.53±1.64	5.22±1.66	4.63±1.79	4.88±1.42	4.6±2.23	4.73±1.7	4.74±1.54	4.17±1.61
	PCG	3.53±1.47	3.54±1.48	3.5±1.5	3.67±1.45	3.29±1.51	3.27±1.46	3.57±1.4	3.56±1.62	3.3±1.47
	All	4.13±1.67	4.04±1.63	4.36±1.78	4.15±1.69	4.08±1.65	3.93±1.91	4.15±1.64	4.15±1.67	3.74±1.59
Fun	RCG	4.35±1.99	4.13±2.05	4.96±1.76	4.18±1.98	4.66±2.03	4.2±2.17	4.29±1.96	4.47±2.09	4.09±1.92
	PCG	3.4±1.81	3.38±1.89	3.46±1.67	3.39±1.73	3.41±2.01	2.1±1.34	3.57±1.65	3.56±2.04	3.04±1.8
	All	3.88±1.95	3.75±1.99	4.21±1.85	3.79±1.89	4.03±2.09	3.15±2.03	3.93±1.82	4.01±2.09	3.57±1.91
Immersiveness	RCG	4.35±1.98	4.09±2.16	5.04±1.23	4.11±1.96	4.78±2.01	3.6±1.98	4.43±1.75	4.47±2.28	4.17±1.84
	PCG	4.16±1.81	4.06±1.78	4.42±1.94	4.16±1.66	4.16±2.1	3.4±2.27	4.38±1.58	4.11±1.97	4.07±1.71
	All	4.26±1.89	4.08±1.96	4.73±1.62	4.13±1.8	4.47±2.04	3.5±2.01	4.41±1.65	4.29±2.11	4.12±1.76
Comment Length	RCG	200.5±275	120±136	415±417	205±321	193±177	121±164	202±357	221±193	236±346
	PCG	177±223	86±807	336±156	160±156	144±155	123±170	177±170	136±133	148±172
	All	201±275	103±112	375±311	182±251	168±165	122±157	189±273	179±169	192±274
Comment Type	RCG	2.68±1.55	2.41±1.6	3.42±1.17	2.64±1.59	2.75±1.53	1.6±1.82	2.38±1.6	3.33±1.19	2.61±1.62
	PCG	2.55±1.62	1.94±1.63	3.67±1.16	2.32±1.7	2.56±1.71	1.6±2.19	2.38±1.75	2.67±1.5	2.09±1.62
	All	2.68±1.55	2.17±1.62	3.54±1.14	2.48±1.64	2.66±1.6	1.6±1.9	2.38±1.65	3±1.37	2.35±1.62

表の見方

色が濃い方がPCGとRCGに大きな違いが見られる

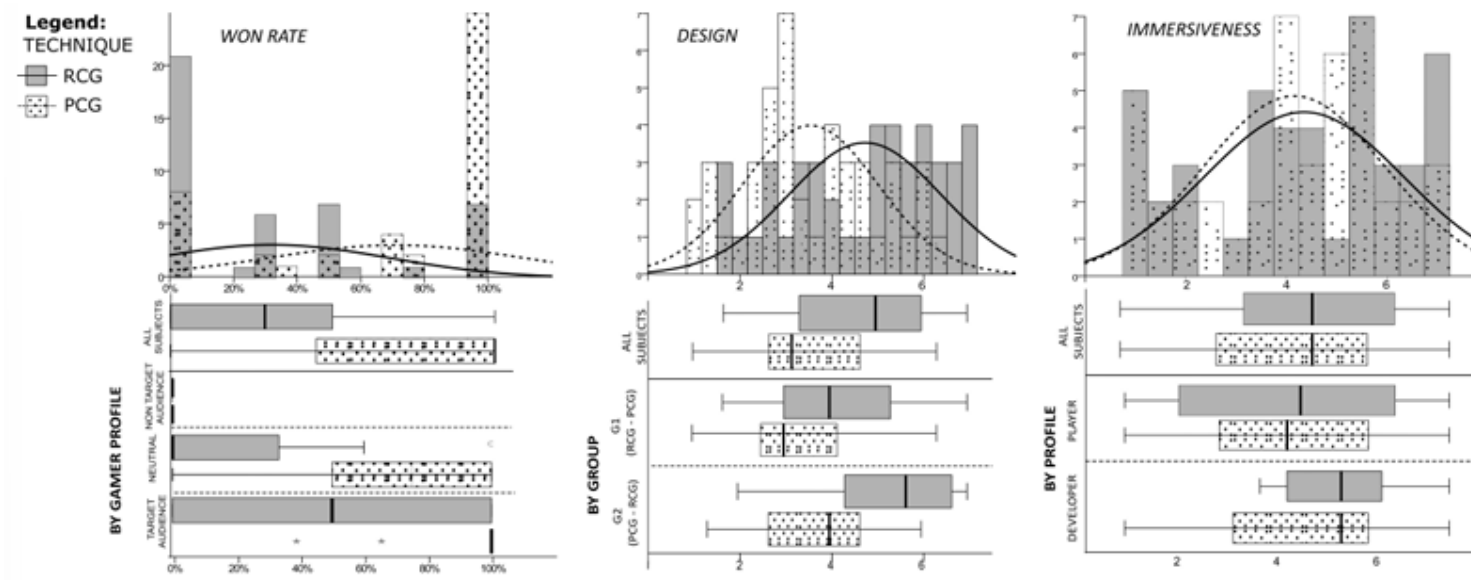
=プレイ時間，勝率，難易度に大きな違いがあることが分かる

Cohen's dを用いた分析

	Technique (RCG vs PCG)	Profile (Players vs Developers)	Developing Games (< 10h/week vs ≥ 10h/week)	Gamer Profile (1vs2, 1vs3, 2vs3)	Group (G1vsG2)
Game duration	0.941	0.086	0.103	(0.203,-0.213,-0.448)	0.051
Won rate	-1.024	0.010	-0.434	(-1.265,-2.166,-0.667)	0.353
Boss difficulty	1.248	-0.272	0.339	(-0.067,0.363,0.448)	-0.009
Design	0.760	-0.194	0.039	(-0.128,-0.125,0.002)	-0.497
Fun	0.501	-0.235	-0.125	(-0.418,-0.417,-0.044)	-0.335
Immersiveness	0.102	-0.347	-0.177	(-0.527,-0.379,0.060)	-0.151
Comment Length	0.209	-1.456	0.061	(-0.261,0.338,0.046)	0.141
Comment Type	0.168	-0.910	-0.541	(-0.460,-0.936,-0.405)	-0.257

評価結果

勝率，デザイン，没入感に関する技術の差



勝率はPCG，デザインと没入感はRCGの方が上回る結果となった。

評価のまとめ

勝率はPCG，それ以外の要素はRCGの方が効果があるということが図や表より分かった。

RQおよびその回答

リサーチクエスチョン

RQ1 : ゲームにおいて自動的にソフトウェアを生成する技術は、ゲームの知覚される品質に影響を与えるのか？

回答 : 被験者によって要素の捉え方が異なったため、生成技術が知覚される質に影響を与えていると言える。

RQ2 : 評価する者のプロフィールが異なると、ゲームの品質評価も異なるのか？

回答 : 勝率とゲーム時間以外の変数からプロフィール要因による統計的有意な影響が確認できなかったため、評価者のプロフィールが評価に影響を与えとは言いきれない。

考察と結論

考察

- 先行研究などでの考察

ゲームにおいて“再利用”はプレイヤーに“使いまわし”と認識されるという懸念点がある。その点PCGのランダム性は新しいコンテンツを広げるという点で優位であると考えられる。

- 今回の考察

RCGの手法はむしろ潜在的なコンテンツを活かし、PCGよりも優れたものを生み出すということが評価から分かった。

なぜこの認識の差が？

これまでの研究は被験者をプレイヤーに絞っていたため。

今回は開発者目線でも実験を行ったため、先行研究の考察が一様にそうではないことが明らかになった。

考察と結論

結論

これまでのゲームソフトウェア工学におけるコンテンツ実験の多くは、ソフトウェア工学の実証実験で求められるベストプラクティスに準拠していなかった。（仮説の提示，妥当性，分析，etc）

本研究ではゲームの品質指標と確率された実証方法を統合して実験を行った結果、**コンテンツの再利用が実際には有利に働くという従来の通説に反する結果**が得られた。

今後、開発者の実験参加を促す新たな可能性を示すことができた。

所感

被験者, ゲームの種類によって結果の差がありそう.

- 被験者を開発者のみに絞るとどうなるのか
- シューティングゲーム以外, ボス戦以外ではどうなのか

GUI関連の研究に使えるかどうか

- RCGを利用してGUI作成? = LLM利用と同じ?
- ゲーム“作成”関連の研究はLLMを使わなさそう.