

# 意味的類似度による自動マージを備えた UML クラス図共同編集環境

橋浦研究室

122D028 岡部 優弥

122D037 柄澤 淳吉

## 1 はじめに

ソフトウェアの設計において、UML のクラス図はシステムの静的構造を表現する重要な成果物である。しかし、初学者が個人でクラス図を作成する場合、客観的な視点が入らないことによるレビュー不足が設計品質の低下を招く一因となる [1]。この問題を解決するアプローチとして、複数人で議論しながら作業を進める共同編集が有効であると考えられる。共同編集は、認識のずれを防ぎ、設計ミスを早期に発見できる可能性が高まる。

## 2 関連研究

個人編集の問題点として、Olivia ら [1] によるとクラス図作成では設計者の視点が限定されるため、仕様の抜け漏れや誤りが生じやすく、本来必要であるべきクラスや属性、関連が欠落する可能性が高いと指摘している。共同編集の有効性として、Dittmar ら [2] は、共同モデリングが多様なアイデアや代替案の創出を促し、質の高い設計成果につながることを示唆しているが、その品質に関する定量的評価は十分に行われていない。モデル統合のガイドラインでは、Lutz ら [3] は、UML モデルのマージにおいて"比較"が中心的な活動であるとし、アルゴリズムによるサポートや個々のワークフロー支援を含むツール設計のガイドラインを提唱している。

## 3 研究目的

本研究では以下の 4 つの Research Question (RQ) を設定する。

**RQ1:** PC 環境は、ホワイトボード環境に比べ、UML クラス図の"関連"に関する設計ミスを減少させるか？

**RQ2:** PC 環境は、ホワイトボード環境に比べ、UML クラス図の"クラス"に関する設計ミスを減少させるか？

**RQ3:** PC 環境は、ホワイトボード環境に比べ、UML クラス図全体に関する設計ミスを減少させるか？

**RQ4:** クラス図の比較プロセスにおいて、ツールの自動化支援は、参加者の協調的な問題解決プロセスにどのような影響を与えるか？

RQ を検証するために、本研究では PC 環境"実験群"とホワイトボード環境"統制群"の 2 つのグループによる比較実験を行う。

## 4 実装

本研究では、UML の複雑さを解消するため、クラス図作成に特化したモデリングツールである KIFU [4] を拡張し、図 1 に示すように、WebSocket および、OT [5] を搭載したリアルタイム共同編集環境を構築した。また、軽量言語モデル EmbeddingGemma-300 [6] を活

用し、クラス名や属性名の意味的類似度を考慮した自動マージ機能の実装を行った。

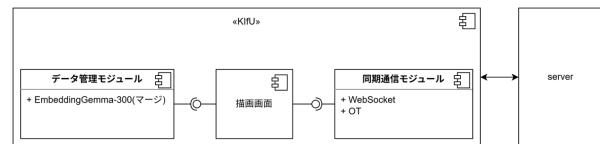


図 1: 提案システムの全体像

以下の図 2 は作成したツール KIFU5.0 の描画画面であり、リアルタイム共同編集機能に加え、自動マージ機能を備えている。

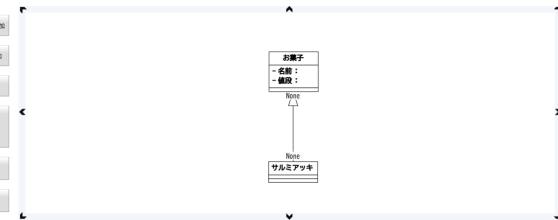


図 2: クラス図共同編集環境の画面

PC 環境による支援がホワイトボードを用いた従来の手動共同編集と比較して、成果物とプロセスにどのような影響を与えるかを検証する。

## 5 実験計画

被験者を"PC 環境グループ"実験群"と"ホワイトボード環境グループ"統制群"の 2 群にランダムに割り当て、2 人 1 組のペアで UML クラス図作成課題に取り組ませる。実験手順として、個人編集(30 分)、比較(10 分)、マージ・統合(10 分)"実験群は自動マージ、統制群は手動"を行う。共同編集(20 分)、確認(15 分)の順に進める。

## 6 評価

### 6.1 評価指標と分析方法

**類似度評価:** 正解データに対する"関連類似度 ( $RS_{all}$ )" [7] および"クラス類似度 ( $CS_{all}$ )" [7], "クラス図全体類似度 ( $CDS$ )" [7] を算出する。これらを用いて、PC 環境が"関連"や"クラス", "クラス図全体"の設計ミス減少に寄与するかをウェルチの t 検定により検証する。

**プロトコル分析:** Lutz ら [3] らのフレームワークを参考に、プロトコル分析を行う。録音データを文字起こし、参加者の発話内容を分析する。分類基準の定義として、ツール(アルゴリズム)で代替可能かを境界線として定義した。具体的には、以下の 2 つのカテゴリに分類する。

- 1. 低次な比較 :**互いのクラス図の要素を提示し、その類似点・相違点を比較する発話

## 2. 高次な設計議論：クラス図の設計意図や要件適合性に関する議論を行う発話

### 7 実験結果

RQ1-3 では有意差は確認できなかったが、以下の図 3 に示す通り、RQ2 に対するクラス単体の細分化を行った結果、社員クラスの類似度において、実験群が統制群を有意に上回った“スチューデントの t 検定(片側), p 値=0.0038, ( $\alpha=0.05$ )”。

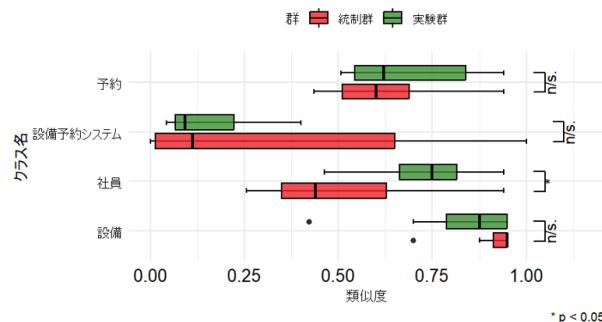


図 3: クラスごとの 2 群の類似度の比較

RQ4 のプロトコル分析の結果、実験群は統制群に比べ、低次な比較に費やすコストが $27.9\% \rightarrow 12.0\%$ に減少し、以下の図 4 に示す通り、高次な設計議論に費やすコストが $72.1\% \rightarrow 88.0\%$ に増加した。

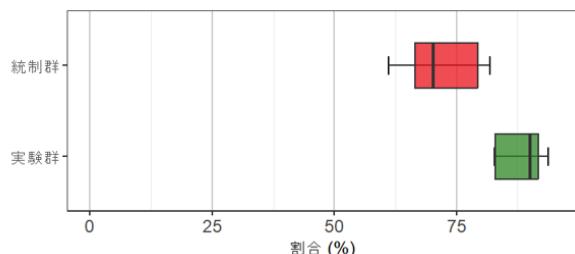


図 4: 両群による高次な設計議論の発話比率の比較

### 8 考察

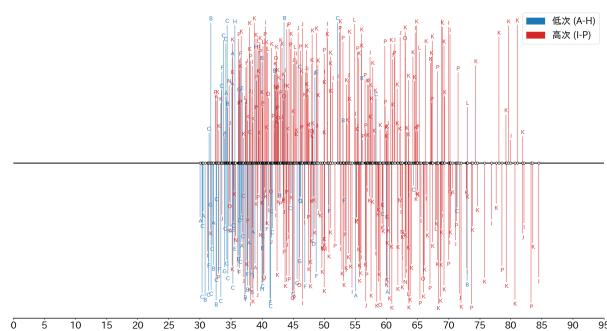


図 5: 2 群の発話比率カテゴリの比較

上記の図 5 は、各群の発話カテゴリの割合を示している。議論の質的な違いとして、低次な比較では、議論の初期段階に差分確認やクラス名の一致に関するマッチングに関する発話が多く見られた。一方、高次な設計議論では、設計意図の共有や多重度、関連に関する議論が多く観察された。高次な議論が多いペア

は、単なるマージではなく、新たな構造の創出や不要な要素の削除といった高度な編集を行っている可能性がある。これにより、正解データとの単純な類似度は下がったとしても、モデルとしての論理的整合性は高まっている可能性があると考えられる。

### 9 結論

PC 環境は、ホワイトボード環境に比べ、UML クラス図の“関連”や“クラス”，“クラス図全体”に関する設計ミスを有意に減少させることは確認できなかったが、社員クラスに関する設計ミスを有意に減少させることが示された。また、PC 環境は、ホワイトボード環境に比べ、低次な比較に費やすコストを有意に減少させ、高次な設計議論に費やすコストを有意に増加させることが確認された。共同編集はプロセスの質を高める上で有効だが、単にツールで自動化して高度な議論に集中するだけでは、必ずしも認識のずれ“構造的な不一致”が解消されるわけではない。設計品質を担保し認識を揃えるためには、自動化支援を受けつつも、人間があえて意識的に事実確認(低次な比較)を行うプロセスを組み込むことが重要である。

### 10 まとめ

本研究では、UML クラス図の共同編集における成果物の質向上を目的とし、LLM による自動マージ支援とリアルタイム同期機能を備えた環境を構築した。本研究で構築したシステムは、被験者を煩雑な手動マージ作業から解放し、より本質的な設計議論へリソースを向けさせることで、共同編集プロセスを効果的に支援できることが示された。

#### 参考文献

- [1] O. G. Fragoso-Díaz, J. A. Sandoval-Acosta, Francisco Javier Álvarez-Rodríguez, J. C. Rojas-Pérez, and R. Santaolaya-Salgado, “Systematic review of quality in class diagrams for software engineering competencies,” *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje (online)*, Vol.17, No.4, pp.351–357, 2022. (DOI:10.1109/RITA.2022.3217169)
- [2] Anke Dittmar, Gregor Buchholz, and Mathias Kühn, “Effects of facilitation on collaborative modeling sessions with a multi-touch uml editor,” *Proceedings of the IEEE/ACM 39th International Conference on Software Engineering: Software Engineering Education and Training Track (ICSE-SEET) (online)*, pp.97–106, 2017. (DOI:10.1109/ICSE-SEET.2017.14)
- [3] Rainer Lutz, David Wurfel, and Stephan Diehl, “How humans merge uml-models,” *2011 International Symposium on Empirical Software Engineering and Measurement (online)*, pp.177–186, 2011. (DOI:10.1109/ESEM.2011.26)
- [4] Takafumi Tanaka, Hiroaki Hashiura, Kousuke Mouri, Atsuo Hazeyama, and Keiichi Kaneko, “An automated evaluation method of conceptual data models considering similarities of attribute sets,” *2018 Seventh ICT International Student Project Conference (ICT-ISPC) (online)*, pp.1–5, 2018. (DOI:10.1109/ICT-ISPC.2018.8523988)
- [5] Clarence A Ellis and Simon J Gibbs, “Concurrency control in groupware systems,” *Proceedings of the 1989 ACM SIGMOD international conference on Management of dataACM*, pp.399–407 1989.
- [6] Henrique Schechter Vera, et al., “Embeddinggemma: Powerful and lightweight text representations,” 2025. <<https://arxiv.org/abs/2509.20354>>
- [7] Takafumi Tanaka, Atsuo Hazeyama, Hiroaki Hashiura, and Seiichi Komiya, “Do learners to create an artifact with good quality make a number of trials and errors during the editing process?” *Proceedings of the 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence (online)*, pp.28–33, 2015. (DOI:10.1109/ACIT-CSI.2015.14)