

Minecraft での人間同士の共同作業における対話の分析

市川拓茉¹ 東中竜一郎²

¹ 名古屋大学情報学部 ² 名古屋大学大学院情報学研究科
{ichikawa.takuma.w0@s.mail,higashinaka@i}.nagoya-u.ac.jp

概要

近年、共同作業が可能な対話システムの構築に向けた研究が盛んに行われているが、創造的な課題の遂行に関する研究はほとんどなされていない。本研究では、創造的な共同作業が可能な対話システムを目指し、その基礎データとして、Minecraftでの人間同士の共同作業データ 500 対話を収集した。加えて、質の高い共同作業に重要な対話的要因を調べるために、質の高い共同作業と対話との関係性を分析した。その結果、質の高い共同作業のためには、作業者の間で同意を求め、それに対して同意していくプロセスをより多く行うことや、序盤にある程度のイメージを合意したうえで作業を進め、その後、実物を見ながら修正箇所や細かい部分を話し合うことが重要であることが分かった。

1 はじめに

対話システムと会話をする機会の増加に伴い [1, 2], 近年では、より高度な対話システムを目指して、人間と協調的にタスクを遂行する共同作業型の対話システムの研究が始まっている [3, 4].

対話システムにおける共同作業の研究では、Minecraft¹⁾を用いた研究が多い [5, 6]. Minecraft はサンドボックス型の 3D ゲームである。プレイヤーはアバターを操作することで Minecraft の世界で自由に活動することができる。また、マルチプレイに対応しており、複数のプレイヤーが同一の環境で共同作業をすることが容易である。Minecraft を用いた研究の多くは所定の問題を解決するタイプの共同作業に関するものであり [7, 8, 9], 創造的な共同作業についてはほとんど検討されていない。人間とシステムが協働する社会の実現のためには、システムは単に命令されて行動するのではなく、人間の仲間として共同で価値を創造していくことが重要である。

本研究では、人間とシステムが共同して新しい

1) <https://www.minecraft.net/>

ものを作り出す創造的な共同作業を遂行できる対話システムを目指し、その基礎的なデータとして、Minecraft での人間同士の共同作業データを収集し分析した。具体的には、創造的な共同作業として共同建築タスクを設定し、Minecraft で作業を行う過程のテキストチャット、行動ログ、作業後のアンケートを収集した。また、共同作業の質を評価するために、収集された作品に対して個性と美しさの観点から第三者評価を収集した。主観評価および第三者評価の高い共同作業に重要な対話的特徴について分析した結果、質の高い共同作業のためには、作業者の間で同意を求め、それに対して同意していくプロセスをより多く行うことや、序盤にある程度のイメージを合意したうえで作業を進め、その後、実物を見ながら修正箇所や細かい部分を話し合うことが重要であることが分かった。

2 Minecraft 共同作業コーパス

創造的な共同作業が可能な対話システムの実現に向けて、共同建築タスクの設定および基礎的なデータの構築を行った。なお、データ収集の実施にあたっては、倫理的関連において所属組織の承認を得ている。

2.1 共同建築タスク

Minecraft での創造的な共同作業を実現するために、2名の作業者が協力して個人的かつ美しい庭を作成する共同建築タスクを設定した。作成対象を庭とすることで、作業の難易度と作品のバリエーションを両立できると考えた。また、個性や美しさといった個人の価値観に左右される基準を設けることにより、作業者同士の話し合いによって創造的に作業を行えるようにした。

共同建築タスクでは、テキストチャットを用いて対話をしながらブロックの操作を行い、個人的かつ美しい庭を作成する。建築可能範囲は縦横 10 マス、高さ 4 マスとした。使用できるブロックは 17

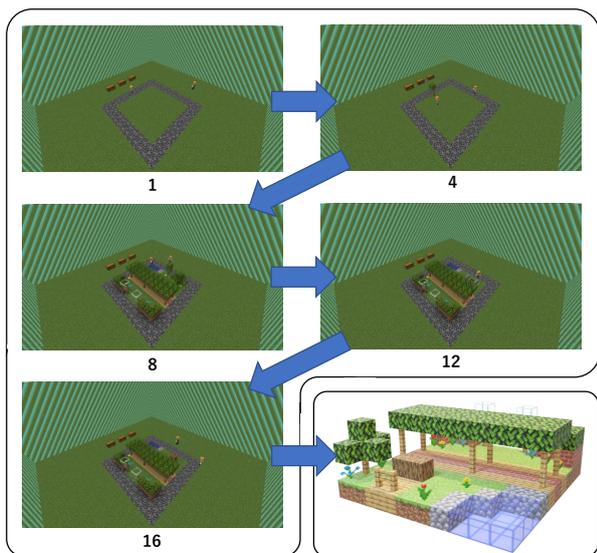


図1 共同建築タスクにおける共同作業の様子と作品(実際に収集したデータから抜粋). 番号は表1における発話IDであり, その発話が行われた時点に対応している.

種類に限った. 作業の制限時間は20分である. 以降, 共同作業1回分のことを1対話と呼ぶ.

2.2 データ収集の手順

データ収集には Minecraft Java Edition 1.16.5 を使用し, データ収集基盤には SpigotMC²⁾ を使用する.

収集するデータはゲーム内テキストチャットのほか, ブロックの設置, 破壊動作およびプレイヤーの移動動作であり, いずれもタイムスタンプが付与されている. 各動作において収集する情報は, 作業者の位置情報(座標データ(x,y,z), 視点情報(yaw,pitch)), 動作固有の情報(チャット内容, ブロックの種類等), 建築範囲の情報(10×10×4の範囲におけるブロックの種類)である.

また, 作業者に対し, 共同作業の満足度を問うアンケートを実施する. アンケートでは, 対話における主張や合意, 作った作品などについて7つの質問を用意し, 5段階のリッカート尺度を用いて集める.

2.3 収集されたデータ

表2に今回収集した500対話の統計量を示す. データ収集実験にはクラウドソーシング³⁾で募った79名が参加し, 1名当たりの作業回数は平均12.7回, 最大24回であった. それぞれの共同作業は異なるペアによって実施された. 1回の作業時間は約

2) <https://www.spigotmc.org/>

3) <https://www.lancers.jp/>

表1 図1の作業時の対話発話

ID	話者	発話
1	A	なにか作りたいものありますか?
2	B	藤?みたいな屋根みたいなつくってみたいです
3	A	いいですね!
4	A	真ん中にどーんと作ってみてください!
5	B	道を真ん中に作ってみます
6	B	両脇にもなにかつくりたいです
7	A	横, こんな感じで大丈夫ですか?
8	B	いいとおもいます!
9	B	2本の木のところに, 試しに椅子置いてみてもいいですか?
10	A	どうぞ! 邪魔だったら壊してください!
11	B	ありがとうございます
12	B	いいかんじの椅子と机の作り方わかりますか?
13	A	思いつかないですね... ハーフブロックとかあればいいのに...
14	B	はい...
15	A	でっかいですね!
16	B	これ一個でテーブルは無理がありますかね...
17	A	いや... いけます!
18	B	いいです!

表2 収集した共同作業データの統計情報

総作業者数	79
総作業ペア数	500
総対話数	500
総発話数	16,221
対話の平均発話数	32.4
発話の平均文字数	13.9
総ブロック設置回数	182,355
総ブロック破壊回数	131,552
作業者の平均ブロック設置回数	182.4
作業者の平均ブロック破壊回数	131.6
作業者の平均移動距離	1,899.5
作品の平均ブロック数	201.8
作品の平均ブロック種類	11.6

25分であり, 作業報酬は1名当たり585円であった.

図1に収集された対話における作業の様子, 表1にその対話内容を示す. この対話では, はじめにメインとなる要素を話し合ってから, それに付け足す形で細かい部分を作り上げている.

アンケート7項目のそれぞれのスコアは, 平均が4.34から4.65, 標準偏差が0.67から1.02であり, 質の高い共同作業のデータが多く得られた.

2.4 作品に対する第三者評価

作業者が主観的によい作品ができたと思っけていても, 客観的によい作品とは限らない. そこで, クラウドソーシング⁴⁾を用いて, 収集された500個の作品に対してより客観的と思われる第三者評価を実施

4) <https://crowdworks.jp/>

した。254名の評価者が参加した。1個の作品に対して10名の評価者が評価を行った。平均作業時間は約42分であり、作業報酬は1名あたり700円であった。

評価者は、はじめに、別途実施した予備実験において収集した30個の作品を見て、自身の中で評価基準を定め、その後、20個の作品に対して個性と美しさの観点から評価を行った。個性と美しさの評価には7段階のリッカート尺度を用いた。

10名による評価値の平均は、個性が平均4.84、標準偏差0.73、美しさが平均4.32、標準偏差0.75であった。第三者評価についても比較的高いスコアが得られており、質の高い共同作業が実施できたことが確認できた。

3 共同作業における対話の分析

将来的に共同建築タスクを実施可能な対話システムを構築するために、主観評価および第三者評価の高い共同作業を可能にする対話的要因を分析した。

3.1 共同作業に有用な頻出表現

質の高い共同作業に重要な表現を見つけるために、単語頻度の分析を行った。共同作業の優劣は、アンケートによる作業者の主観評価7項目の総和(満足度)および作品の第三者評価(個性、美しさ)の計3つの分類基準によって付け、収集された500対話すべてを上位20%と下位80%に分類した。

単体である程度の意味を持つ表現が抽出できるように、単語の4-gramに対して分析を行った。分析の対象となる表現は収集した16,221発話に含まれる頻出上位300個の4-gramとした。収集した500対話に対して、それがある4-gramを含むか否か、上位の対話に属するか否かをもとに2×2のクロス表を作成し、それに対してフィッシャーの正確確率検定を行った。単語の抽出には形態素解析器としてJanome⁵⁾を使用した。また、4-gramとして助詞および助動詞のみで構成される表現については分析対象から除外した。

表3に各分類基準において上位の対話に有意に多く見られたものを示す。なお、主観評価による分類では多数の表現が得られたが、第三者評価による分類では得られた表現は少数であった。

満足度の高い対話に頻出する表現は主に2つに分類できる。1つ目は「がいいですか」、「(して)みま

表3 分類基準(満足度、個性、美しさ)ごとの頻出表現(p値はフィッシャーの正確確率検定による。*はp<0.05, **はp<0.01)

分類基準	単語の4-gram
満足度	ありがとうございます!*, がいいですか**, いいですかね*, にしてみ*, になりました*, みましようか**, なんですよ**, どうでしょう**, そうですよね*, いいですよ!*, 感じてどうでしょ*, してみてください*, もよさそう*, よかったです!*, あたりします**, 感ありますね*, こんな感じでどう**, ありがとうございました**, てもよさ*,
個性	のはどうです*, っぽくなりました*
美しさ	しますか?*, てみても*, みてもいい*, してみたいです*, いいかんじですね*, もいいと思い**

しょうか」、「どうでしょう」等の相手に同意を求める表現である。これらは、相手にアイデアを提案する際や作業の方針を決める際の発話に確認された。2つ目は「なんですよ」、「そうですよね」、「いいですよ!」等の相手に好意的な評価を示す表現である。これらは、相手の意見や作ったものに対して好意的な評価を示し、同意する際の発話に確認された。

第三者評価の高い対話については、頻出表現があまり得られなかったが、個性による分類では、「のはどうです(か)」と「っぽくなりました」が確認された。これらのうち、「のはどうです(か)」は相手にアイデアを主張し、それに対して同意を求める表現である。美しさによる分類では、6つの表現が得られた。これらのうち、「しますか?」、「てみてもいい(ですか)」は相手に同意を求める表現であり、「いいかんじですね」、「もいいと思い(ます)」は相手に好意的な評価を示す表現である。

これらの分析をまとめると、いずれの分類においても同意に関する表現が多く見られた。このことから、質の高い共同作業には作業仲間間で同意を求め、それに対して同意していくプロセスをより多く含むことが重要であることが定量的に確認できた。

3.2 共同作業の流れの分析

どのような流れの対話が質の高い共同作業につながるかを調査した。まず、共同作業の流れを分析するため、発話時における作品の完成度を時系列データとしてプロットし、それに対してクラスター分析を

5) <https://github.com/mocobeta/janome>

行った。

共同建築タスクでは2名の作業者が協力して作品を完成させるが、最後まで残っていた部分についてはすでに両者が納得していると考えられる。これを作品の完成度と捉え、以下の様に定義した。

$$\text{作品完成度} = \frac{|Blocks_t \cap Blocks_{last}|}{|Blocks_{last}|} \quad (1)$$

$Blocks_t$ は時刻 t に設置されているブロックの集合、 $Blocks_{last}$ は最終的な作品に使用されているブロックの集合を表す。発話の時点とその時の作品完成度を時系列データとし、クラスタリングを行った。クラスタリングには k-Shape[10] を使用した。また、光田ら [11] に従い、作品完成度は平均 0、分散 1 の正規分布に従うように正規化した。クラスタリングでは、クラスタ内誤差平方和が下げ止まった箇所が複数存在したため、各クラスタ数による結果を著者の 1 人が確認し、類似するクラスタが出現しないクラスタ数 6 を選択した。

図 2 に得られたクラスタを示す。各クラスタが全体に占める割合は、クラスタ 1 が 4.4%、クラスタ 2 が 19.6%、クラスタ 3 が 16.4%、クラスタ 4 が 3.0%、クラスタ 5 が 32.2%、クラスタ 6 が 24.4% である。クラスタ 1 とクラスタ 4 では、終盤に作品完成度が急増している。序盤に何をやるか話し合いによって決めた後、黙々と作業を続けた対話がこのに含まれる。クラスタ 3、クラスタ 5 では、作品完成度が中盤にやや大きく増加している。序盤に話し合った大まかな内容で作業を進め、中盤以降は細かい部分について話し合った対話がこのに含まれる。クラスタ 2 とクラスタ 6 については、序盤から終盤にかけて作品完成度がほぼ一定に増加している。常に話し合いながら作業を行い、少しずつ合意を重ねて作品を作った対話がこのに含まれる。

3.3 共同作業の流れと評価値の関連

よい共同作業がいずれのクラスタのものなのかを把握するために、得られたクラスタに対して質の高い対話が頻出するかを調べた。対話を質によって分類するために、主観評価 7 項目それぞれとその総和、第三者評価 2 項目それぞれの計 10 項目を対話の評価基準として使用し、そのスコアが高い上位 20% を質のよい対話、下位 80% を質の悪い対話として分類した。収集した 500 対話に対して、それがああるクラスタに属するか否か、よい対話に属するか否かをもとに 2×2 のクロス表を作成し、それに対し

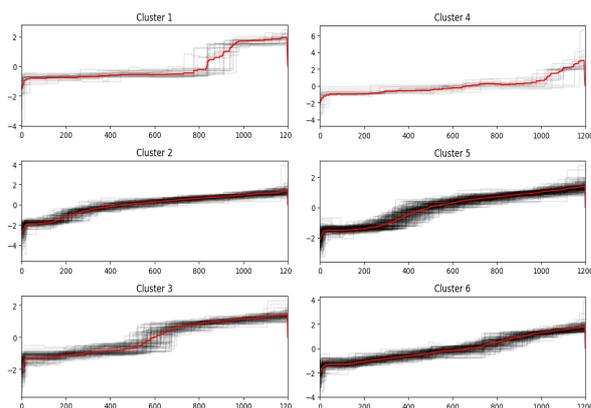


図 2 作品完成度の時系列クラスタリング結果

てフィッシャーの正確確率検定を行った。

クラスタ 3 では、第三者評価の「個性」および「美しさ」によって対話を分類した際に、クラスタ 5 では、主観評価 7 項目の総和によって分類した際に、それぞれ 5% の有意水準で質の高い対話が有意に多く出現することが認められた。一方、クラスタ 2 やクラスタ 6 では質の高い対話を有意に多く含むことは認められなかった。常に話し合い、合意をしながら作業をすることがよりよい共同作業につながると思われがちだが、そうっていない点は興味深い。

この分析から、序盤にある程度のイメージを合意したうえで作業を行い、その後、細部について話し合って作品を完成させるような共同作業が、いずれの評価においても重要であることが分かった。

4 おわりに

本研究では、創造的な共同作業が可能な対話システムを目指し、その基礎データとして、共同建築タスクにおける人間同士の作業の様子を Minecraft を用いて収集した。また、対話に対する評価を作業者による主観評価と第三者評価によって行い、それらの評価値が高い対話における特徴を分析した。その結果、質の高い共同作業のためには、作業者の間で同意を求め、それに対して同意していくプロセスをより多く行うことや、序盤にある程度のイメージを合意したうえで作業を進め、その後、実物を見ながら修正箇所や細かい部分を話し合うことが重要であることが分かった。

今後は、イメージを合意するための対話手法、実物について評価したり修正を提案する手法などを検討し、実際にユーザと創造的な共同作業を対話を通して行うシステムを構築したい。

謝辞

本研究は科研費「モジュール連動に基づく対話システム基盤技術の構築」(課題番号 19H05692)の支援を受けた。

参考文献

- [1] 中野幹生, 駒谷和範, 船越孝太郎, 中野有紀子, 奥村学. 対話システム(自然言語処理シリーズ). コロナ社, 2015.
- [2] 東中竜一郎, 稲葉通将, 水上雅博. Python でつくる対話システム. オーム社, 2020.
- [3] Charles Rich, Candace L. Sidner, and Neal Lesh. Colla-gen: Applying collaborative discourse theory to human-computer interaction. **AI Magazine**, Vol. 22, No. 4, p. 15, 2001.
- [4] James Allen, Nathanael Chambers, George Ferguson, Lucian Galescu, Hyuckchul Jung, Mary Swift, and William Taysom. Plow: A collaborative task learning agent. In **Proceedings of the 22nd National Conference on Artificial Intelligence - Volume 2**, p. 1514–1519, 2007.
- [5] Arthur Szlam, Jonathan Gray, Kavya Srinet, Yacine Jernite, Armand Joulin, Gabriel Synnaeve, Douwe Kiela, Haonan Yu, Zhuoyuan Chen, Siddharth Goyal, et al. Why build an assistant in minecraft? **arXiv preprint arXiv:1907.09273**, 2019.
- [6] Haruna Ogawa, Hitoshi Nishikawa, Takenobu Tokunaga, and Hikaru Yokono. Gamification platform for collecting task-oriented dialogue data. In **Proceedings of the 12th Language Resources and Evaluation Conference**, pp. 7084–7093, 2020.
- [7] Anjali Narayan-Chen, Prashant Jayannavar, and Julia Hockenmaier. Collaborative dialogue in Minecraft. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 5405–5415, 2019.
- [8] Arne Köhn, Julia Wichlacz, Christine Schäfer, Álvaro Torralba, Joerg Hoffmann, and Alexander Koller. MC-saar-instruct: a platform for Minecraft instruction giving agents. In **Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue**, pp. 53–56, 2020.
- [9] Cristian-Paul Bara, Sky CH-Wang, and Joyce Chai. Mind-Craft: Theory of mind modeling for situated dialogue in collaborative tasks. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 1112–1125, 2021.
- [10] John Paparrizos and Luis Gravano. k-shape: Efficient and accurate clustering of time series. In **Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data**, pp. 1855–1870, 2015.
- [11] 光田航, 東中竜一郎, 大賀悠平, 杵渕哲也. 共同図形配置課題における対話の共通基盤構築過程の分析. 言語処理学会第 27 回年次大会発表論文集, 2021.