2025 年度

修　士　論　文

マルチエージェント強化学習を用いた連合ニューラル翻訳の自己組織化

指導教員: 村上 陽平

立命館大学大学院 情報理工学研究科
情報理工学専攻 博士課程前期課程
計算機科学コース

学生証番号: 6611240011-8

氏名: 北川 勘太朗

マルチエージェント強化学習を用いた
連合ニューラル翻訳の自己組織化

北川 勘太朗

内容梗概

ニューラル機械翻訳の誕生により近年，機械翻訳の精度がより一層高くなり，世間から大きな注目を集めている．高精度なニューラル機械翻訳には大量で高品質な対訳データが必要になるが，その構築には膨大なコストがかかる．また，著作権や機密情報の保護のために，異なる組織との対訳データを共有することも難しい．この問題の解決策の一つが連合学習である．連合学習を用いることで複数のデータ所有者が各自の持つデータを秘匿にしたまま，翻訳モデルのみを共有でき，協力してニューラル機械翻訳を構築できる．しかしながら，各データ所有者の持つデータの分布がそれぞれ異なっている場合，連合学習では非独立同一分布 (Non-independent and identically distributed (Non-iid)) と呼ばれ，モデルの精度低下に繋がるため，全ての翻訳モデルを統合することが必ずしもモデルの翻訳精度の向上に繋がるとは限らない．

そこで，本研究では連合学習において，集約プロセスのたびに動的に連携相手を選択し，翻訳モデルを統合する手法を提案する．各データ所有者をエージェントとし，エージェントは深層強化学習を通じて連携相手を評価し，最適なモデルの統合先を選択する方策を学習する．本手法の実現にあたり，取り込むべき課題は以下の 2 点である．

マルチエージェント強化学習に基づく自己組織化の設計

連合学習に参加する各クライアントは，自身のデータ特性に適した翻訳モデルの獲得を目指すが，従来の手法では，クライアント間のデータ分布の違いや個別の協調要求を反映できない．この課題に対し，本研究では，クライアントが自身の性能向上に寄与する相手のみを選択して連携し，学習ごとに協力関係を柔軟に更新できる仕組みの重要性を指摘する．

エージェントの内部モデル

本研究におけるエージェントは，他クライアントとの連携が自身の翻訳性能に与える影響を観測し，その結果に基づいて次の連携相手を選択する．その内部モデルは，過去の連携による翻訳精度の変化から，特定の相手と協調した場合の性能向上・低下を推定する役割を担う．一方で，エージェントの行動を決

定する方策は，内部モデルではなく報酬に基づいて強化学習により学習される．そのため，翻訳精度の向上を適切に反映した報酬を設計することが重要である．

前者の課題に対しては，エージェントの状態を検証データに基づく翻訳精度行列，報酬を検証データから得られる翻訳精度の向上量，行動を連携するクライアントの選択として定義する．このように自己組織化のための状態・報酬・行動を設計することで，エージェントは自身の得られる状態と報酬を基準に，性能向上に寄与する連携相手を選択可能とする．さらに，強化学習を通じてエージェントは個別の方策を獲得し，新たなデータセットに対しても動的に最適な連携関係を構築できる．

後者の課題に対しては，本研究ではエージェントの内部モデルは特定の相手と連携した場合に自分の翻訳性能がどの程度変化するかを推定するための仕組みであり，その評価指標として翻訳精度を用いる．内部モデルの構成は，性能を評価する際に用いる検証データの作り方に依存する．具体的には，自ドメインの検証データのみで評価する利己的エージェントと，全ドメインを均等に含む検証データで評価する協調的エージェントの 2 種類を設定し，これらの行動と性能を比較した．

提案手法によって構築された翻訳モデルを用いて, 評価データを翻訳し, BLEU スコアを用いてその精度の評価を行った．本研究の貢献は以下の通りである．

マルチエージェント強化学習に基づく自己組織化の設計

マルチエージェント深層強化学習の導入によって，連合学習のクライアントが将来的のモデルの精度上昇を考慮しながら，最適な連携相手を動的に選択することで， 従来手法のモデルと比較して提案手法によって構築されたモデルの精度は平均で 21.5%高くなり，その有効性を検証した．

エージェントの内部モデル

内部モデルの設計を変化させることで，エージェントが目指す最適行動が変化し，その結果，利己的エージェントは，自身のドメインにおける翻訳精度を最大化することを目的として，データセットの類似度が高いクライアント同士が頻繁に連携する傾向を示した．一方，協調的エージェントは，全ドメインにわたる平均的な翻訳性能の向上を目標として行動を選択するため，特定ドメインへの最適化よりも，全体のバランスを重視した協調関係を構築する傾向を示した．

Self-Organization of Federated Neural Machine Translation via Multi-Agent Reinforcement Learning

Kantaro KITAGAWA

Abstract

Neural machine translation has made machine translation even more accurate. Highly accurate neural machine translation requires large amounts of high-quality bilingual data, which is very expensive to construct. It is also difficult to share bilingual data with different organizations for copyright and privacy reasons. However, using federated learning, multiple data owners can integrate only the translation model while keeping their own data confidential, allowing them to cooperate in the construction of neural machine translation. However, if the data distributions of individual data owners differ, the accuracy of the models will be reduced, so integrating all translation models does not necessarily improve the translation accuracy of all models.

Therefore, we propose a cooperative agent that integrates translation models by dynamically selecting a partner for each aggregation process using deep reinforcement learning in coalition learning. The following two issues needed to be addressed to realize this method.

Self-organizing design based on multi-agent reinforcement learning

Conventional averaging-based aggregation cannot reflect differences in data distributions or individual collaboration preferences. We therefore design a self-organizing mechanism that allows each agent to selectively collaborate with partners that contribute to its performance improvement.

Internal model of agents

Agents must evaluate the impact of collaboration on translation quality. In the proposed approach, agents estimate performance changes using an internal model, while action policies are learned based on rewards defined as improvements in translation accuracy.

To address the former challenge, we define the agent' s state as a translation accuracy matrix on validation data, the reward as the corresponding accuracy improvement, and the action as the selection of collaboration partners. This design enables agents to learn, via reinforcement learning, individualized policies

that dynamically select beneficial partners, even under changing data distributions.

For the latter challenge, the internal model estimates the impact of collaboration on translation performance using validation accuracy. We consider two agent types: a *selfish agent* that evaluates performance using only in-domain validation data, and a *cooperative agent* that uses validation data that equally covers all domains. Their behaviors and performances are compared.

Using the neural machine translation models constructed by the proposed method, we translate evaluation datasets and assess translation quality using the BLEU score. The main contributions of this study are summarized as follows.

Self-organizing design based on multi-agent reinforcement learning

By introducing multi-agent deep reinforcement learning, participants in federated learning are able to dynamically select optimal collaboration partners while taking future improvements in model accuracy into account. As a result, the models constructed using the proposed method achieved, on average, a 21.5% higher translation accuracy compared with models built using conventional methods, thereby experimentally demonstrating the effectiveness of the proposed approach.

Internal model of agents

By varying the design of the agents' internal models, the optimal behaviors pursued by the agents change accordingly. As a result, selfish agents tend to prioritize collaboration partners that are expected to yield the largest short-term performance gains, with the objective of maximizing translation accuracy within their own domains. In contrast, cooperative agents select their actions with the goal of improving average translation performance across all domains, and thus tend to establish collaborative relationships that emphasize overall balance rather than optimization for a specific domain.

マルチエージェント強化学習を用いた
連合ニューラル翻訳の自己組織化

目次

第1章　はじめに

ニューラル機械翻訳 [1] の誕生により近年，機械翻訳の精度がより一層高くなり，世間から大きな注目を集めている [2]．2016 年に発表された Google の機械翻訳システムについての論文 [3] では，従来のフレーズベースと比較して，誤り率を 60%も削減し，平均的なバイリンガルの翻訳精度に迫るスコアを達成している．ニューラル機械翻訳とは，人間の脳神経回路が情報伝達を行う仕組みを模倣したニューラルネットワークを用いて，入力の単語列を符号化し，対訳の確率の高い訳語を選択して，出力の単語列を生成する翻訳アルゴリズムである．高精度のニューラル機械翻訳を構築するためには，高品質の対訳データが大量に必要となる．確かに，web サイトなどで大量の対訳データは集めることは可能であるが，このような対訳データは汎用的で，品質についても高品質である保証がないため容易に使用しづらい．したがって，特定のドメインに特化した高品質な対訳データを一組織で構築するには膨大なコストがかかる．また，複数の組織が対訳データを共有するにも，ドメインが違っていたり，著作権やプライバシー，セキュリティ等の問題で対訳データを共有しがたい．したがって，複数の組織が持つ対訳データのドメインが異なるときに，各自の持つデータを秘匿にしたまま，各組織において高い翻訳精度を達成するモデルを獲得したい．解決策は，対訳データを一カ所に集約して学習するのではなく，各データ所有者の下で学習したモデルを連携させる連合学習 [4] を用いる．連合学習を用いることで，ユーザ間での対訳データの共有を必要としなくなり，組織内に蓄積された非公開の大量の対訳データの利用を促進することが期待され，ニューラル機械翻訳の構築において大規模かつ高品質な対訳コーパスが不十分という問題を解決できる．しかしながら，各データ所有者の持つデータの分布がそれぞれ異なっている場合，連合学習において非独立同一分布 (Non-independent and identically distributed (Non-iid)) と呼ばれ，モデルの精度低下に繋がるため，全ての翻訳モデルを統合することが必ずしも全てのモデルの翻訳精度の向上に繋がるとは限らない．

そこで，本研究では連合学習において，集約プロセスのたびに動的に連携相手を選択し，翻訳モデルを統合する協調的なエージェントを提案する．各データ所有者をエージェントとし，エージェントは深層強化学習を通じて連携相手を評価し，最適なモデルの統合先を選択する方策を学習する．

本手法の実現にあたり，取り込むべき課題は以下の 2 点である．

マルチエージェント強化学習に基づく自己組織化の設計

連合学習に参加する各クライアントは，自身が保有するデータの特性に適した翻訳モデルを獲得することを目指す．しかしながら，従来の連合学習手法では，全てのクライアントのモデルを一度の集約で機械的に平均化する方式が一般的であり，個々のクライアントが持つデータ分布の違いや，どのクライアントと協調したいかといった個別の要求を反映できない．特に Non-iid 環境では，データ内容が大きく異なるクライアント同士のモデルを無差別に統合すると，特定のクライアントにとっては不要な特徴が取り込まれることで翻訳性能が低下する場合がある．この問題を避けるためには，クライアントが自身のモデルの性能向上に寄与する相手だけと連携できる仕組みを導入し，学習のたびに適切な協力関係を選び直せるようにする必要がある．

エージェントの内部モデル

本研究におけるエージェントは，他クライアントとの連携が自身の翻訳性能に与える影響を観測し，その結果に基づいて次の連携相手を選択する．その内部モデルは，過去の連携による翻訳精度の変化から，特定の相手と協調した場合の性能向上・低下を推定する役割を担う．一方で，エージェントの行動を決定する方策は，内部モデルではなく報酬に基づいて強化学習により学習される．そのため，翻訳精度の向上を適切に反映した報酬を設計することが重要である．本研究では，連携前後の翻訳精度の変化を報酬として与えることで，エージェントが性能向上に寄与する連携戦略を学習できるようにする．

本論文は以下の構成で進める．第 2 章では，ニューラル機械翻訳およびそのための連合学習の基礎と関連研究を概説する．続く第 3 章では，連合学習をニューラル機械翻訳に適用する際に生じる集約の課題を整理し，自己組織化手法として FedAvg，greedy 法，およびマルチエージェント強化学習に基づく手法を形式的に定式化する．

第 4 章では，ニューラル機械翻訳向けマルチエージェント強化学習のエージェントの状態・報酬・行動設計について述べ，第 5 章では利己的エージェントと協調的エージェントという 2 種類の内部モデルの構成およびその意義を明らかにする．第 6 章では，これらの自己組織化手法を LLM エージェントに拡張する際の設計上の要点を示す．

第 7 章では，実験環境，データセット，評価指標，および強化学習エージェ

ントの実装方法について説明する．第 8 章では，以上の手法に基づく検証結果を示し，9 章で得られた知見について考察する．最後に，本論文の結論と今後の展望をまとめる．

第2章　関連研究

2.1　ニューラル機械翻訳

ニューラル機械翻訳（Neural Machine Translation: NMT）は，深層学習を用いてある言語のテキストを別の言語へ自動的に翻訳する，現時点で最も先進的な翻訳技術である．従来のルールベース機械翻訳（Rule-Based Machine Translation: RMT）や統計的機械翻訳（Statistical Machine Translation: SMT）[5] のように手作業で設計された特徴量や翻訳ルールに依存する手法とは異なり，NMT はニューラルネットワークによってデータから直接翻訳パターンを学習する．こうしたエンド・ツー・エンドの枠組みにより，NMT は高い精度と自然な流暢性を実現し，現在では機械翻訳分野における主流のアプローチとなっている．

NMT システムは，ソース言語とターゲット言語のペア文から構成される対訳コーパスを用いて学習される．典型的な構造はエンコーダ・デコーダモデルであり，エンコーダは入力文を連続ベクトルに符号化し，デコーダはその表現を基に翻訳文を生成する．さらに，Transformer モデル [6] に代表されるアテンション機構の導入により，デコーダが入力文中の関連箇所に動的に注意を向けることが可能となり，翻訳性能が大幅に向上した．

2.2　ニューラル機械翻訳のための連合学習

NMT は，学習時に対訳コーパスを用いるため，翻訳性能が学習データの量および質に強く依存する．一般に，大規模かつ高品質な対訳コーパスを用いることで，翻訳精度は向上する．近年では，OpenAI の GPT [7] や Google の BERT [8] に代表される大規模事前学習言語モデルの登場により，文脈理解能力が飛躍的に向上し，より自然で文脈に即した翻訳が可能となっている．一方で，十分な量の対訳データを単一の組織で収集・管理することは，著作権，プライバシー，およびセキュリティ上の制約から依然として困難である．

この問題に対して有望な解決策として注目されているのが，連合学習（Federated Learning: FL）を活用した連合型 NMT である．FL は，2016 年に提唱された複数の組織やデバイスが生データを共有せずに協調的に機械学習のモデルを構築するための分散学習の枠組みである．図 1 のように各クライアントはローカルデータを用いて学習を行い，モデルの更新パラメータのみを中央サーバに送信する．サーバはそれらを集約してグローバルモデルを更新し，再度ク

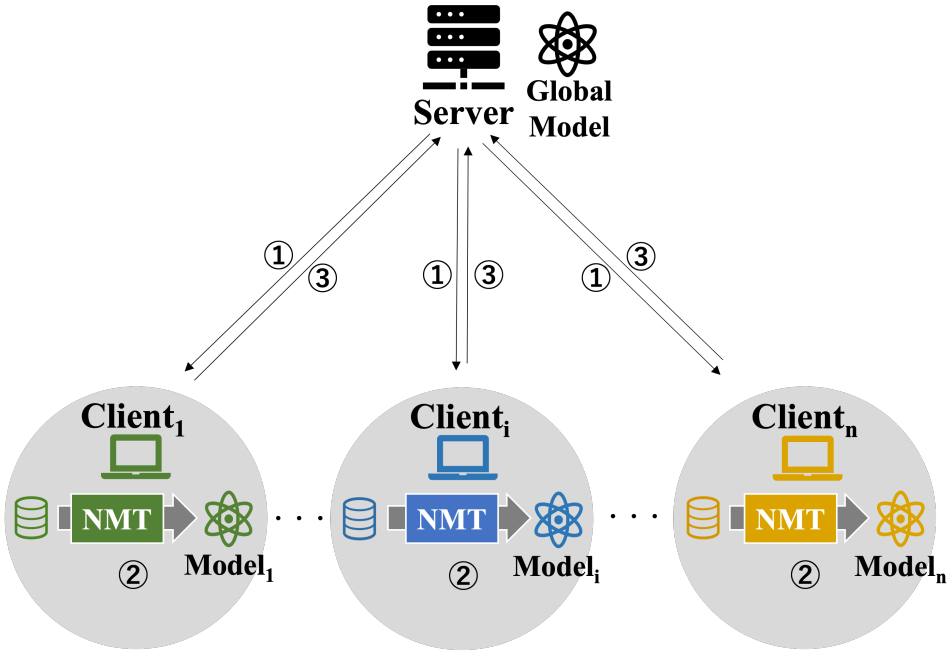

図 1: 連合学習

ライアントに配布する．この反復的プロセスにより，データを一箇所に集めることなく，プライバシーを保ちながら高性能な翻訳モデルを構築できる．つまり，各組織は自身のデータを保持しながら，モデルの学習に寄与し，複数の組織間での協力的な学習が実現される．プライバシーやデータの所有権の制約，また組織間でのデータ共有に関する法的な制約やセキュリティ上の懸念がある場合でも，FL を使用することでこれらの問題を解決する可能性があり，その有望性から研究や実践で広く注目されている．

しかしながら，FL には Non-IID という根本的な課題が存在する．各クライアントのデータは個別に収集されるため，分布の違いが大きく，これがモデルの偏りや収束遅延を引き起こし，翻訳精度を低下させる要因となる [9]．この課題を解決するために，パーソナライズド層の導入 [10]，ローカルファインチューニング [11]，グローバル共有データの併用 [12]，および Deep Q-Learning によるクライアント選択最適化 [13] などが提案されてきた．

2.3 マルチエージェント強化学習

強化学習（Reinforcement Learning: RL）[14] は，エージェントが環境との相互作用を通じて得られる報酬を手がかりに，環境に適応する方策を獲得する機

械学習の一種である．その概念は心理学における強化理論に由来しており，刺激と応答の間に与えられる強化子によって特定の行動パターンが学習・増強されるという考え方に基づいている．

強化学習は教師データを必要とせず，行動の結果として観測可能な情報が変化する環境において，累積報酬を最大化する行動系列を探索することを目的とする．一般的な設定では，自律的に行動する主体をエージェント，その行動の影響を受ける対象を環境と呼ぶ．エージェントが環境に対して行う操作を行動，環境の状況を表す構成を状態と定義し，状態遷移の結果として得られる数値的な評価指標が報酬である．エージェントが選択する行動に応じて，次に遷移する状態および得られる報酬が変化するため，強化学習では，未知の環境における行動の良否を評価する指標としてスカラー値の報酬が用いられる．

強化学習の近代的なアプローチは，1950 年代ごろから，Q 学習 [15][16] や SARSA など価値関数を用いた手法が取り組まれてきた．Q 学習では，Q 関数と呼ばれる行動価値関数を学習し，制御を実現する．Q 関数は，ある時刻ある状態のときにある行動を行った場合に，その先でどれくらいの報酬が期待されるかを出力する関数である．さらに近年では，多層ニューラルネットワークの高い表現力を用いて，強化学習における Q 関数 の近似に多層ニューラルネットワークを応用した深層強化学習（Deep Reinforcement Learning: DRL）が広く研究されている．深層強化学習は，Mnih ら [17][18] によって提案された Q 学習の拡張である Deep Q-Network (DQN) などが代表的な手法であり，画像入力を対象とした高い識別能力を示している [19]．

実世界の多くの場面では，複数のエージェントが同じ環境内で相互作用しながら学習する．複数のエージェントが同一環境下で相互に影響を与えながら学習するの状況を扱う枠組みがマルチエージェント強化学習（Multi-Agent Reinforcement Learning: MARL）であり，エージェント間の協力や競争を通じて，単一エージェントでは得られない戦略や適応行動が獲得できる．

また，MARL は，エージェント間の利害関係に基づき以下のように分類される．一つ目は全てのエージェントが協力し，システム全体の報酬を最大化する完全協力型 (Fully Cooperative) である．二つ目はあるエージェントが勝利すると他のエージェントが負けになる完全競争型（Fully Competitive）である．三つ目はエージェントたちに競争と協力の関係が同時に存在する混合：協力＆競争型（Mixed Cooperative Competitive）である．四つ目はエージェントが自分

の利益だけを最大化にする利己型（Self-interested）である．

そして, 学習の実装形態によっても, 次の三つに分類される．シングルエージェントの場合と同様に，中央集権的なエージェントが他のエージェントの学習，行動をコントロールする完全中央集権型（Fully Centralized）がある．各エージェントが独立して学習，行動を決定する完全非中央集権型（Fully Decentralized）もある．中央集権的なエージェントに学習して，他エージェントが行動を決定する混合型（Mixed：Centralized Decentralized）も存在する．これらの学習方法の中で，完全中央集権型の学習が最も安定し，収束しやすい一方で，エージェント間の相互作用を考える必要があり，学習空間が大きくなる傾向がある．一方，完全非中央集権型ではエージェント間の相互作用を考える必要がないため，学習空間が小さくて済むが，その分エージェントたちが互いの状態を把握しきれず，学習が収束しにくく，不安定になりやすいという課題がある．

2.4　連合学習とマルチエージェント強化学習

連合学習とマルチエージェント強化学習はいずれも分散型学習に基づいており，相互に高い親和性を持つ技術である．マルチエージェント強化学習（Multi-Agent Reinforcement Learning，MARL）は，複数のエージェントが協力して目標を達成するために学習するフレームワークであり，連合学習は，各クライアントが独自のモデルを保持しつつ，分散されたデータを共有せずに，他のクライアントとの協調を通じて自身のモデルを改善する手法である．これらの技術の組み合わせには双方向の可能性があり，連合学習を改善するためにマルチエージェント強化学習を利用したり，逆にマルチエージェント強化学習を強化するために連合学習を適用したりするシナリオが考えられる．

Zhang らは，連合学習においてマルチエージェント強化学習を用いてクライアント選択を最適化することで，モデル精度と通信効率の両立を実現した [20]．また，無線通信における連合学習システムでのデバイスの参加意思決定を最適化するため，マルチエージェント強化学習とメタ学習を活用したメカニズムを提案し，効率的なクライアント選択を達成した [21]．Klein らは，マルチエージェント強化学習を利用して，デバイスの学習行動を評価し，効率的なクライアント選択を実現にした [22]．一方で，Yuan らは，通信分野におけるマルチエージェント強化学習のプライバシー保護のために連合学習を活用し，さらに分割学習を導入することで効率を向上させ，オーバーヘッドを削減する効率的なプ

ライバシー保護方式を実現した [23]．Liam らは，連合強化学習に Transformer を導入し，エージェント間の文脈的関係を学習することで，非同質環境下でも精度とスケーラビリティに優れたモデル統合を実現した [24].

本研究は，特に連合学習の改善に向けてマルチエージェント強化学習を活用する方針を採用している．特に，クライアントに対して連携するパートナーを選択する自主性を与えることで，Non-iid 環境における翻訳精度の向上を目指す．また，これらの既存の研究は，クライアントごとのモデルを作成するのではなく，主に単一のグローバルモデルを作成することを目的としている．これに対して，本研究では各クライアントのローカルモデルの集約ごとに誰と連携するかを決定する方策をクライアントが取得できるようにするため，連合学習にマルチエージェント強化学習を導入している．既存のアプローチとは異なり，この方法は単一のグローバルモデルではなく複数のグローバルモデルを生成し，各クライアントが自らのニーズに最適なグローバルモデルを追求できるようにする．その結果，各クライアントは従来の手法である FedAvg を超える翻訳精度をそれぞれのドメインで達成できることが期待される．

第3章　自己組織化連合学習の定式化

本章では，既存手法である FedAvg，本研究で提案する greedy 法に基づく自己組織化手法，およびマルチエージェント強化学習に基づく自己組織化手法について，それぞれの概要と定式化を述べる．具体的には，各手法におけるモデル集約，クライアント選択，およびモデル更新の方法について，サーバおよびクライアントの両視点から比較し，その相違点を明らかにする．表 1 に，本定式化で用いる記号の一覧を示す．これらの記号には，グローバルモデルおよびローカルモデルの表現，クライアント集合，最適化パラメータに加え，状態・行動・Q 値といった強化学習に関連する要素が含まれる．この記号体系を基盤として，クライアントの異質性，データプライバシー，およびシステムの適応性といった連合学習における重要課題に取り組む．

本研究では，連合学習の開始時に同一の NMT モデルを初期モデルとして各組織に配布し，各組織は受信した初期モデルを，自身が保有するドメイン固有のデータセットを用いてローカルで学習する．その後，学習済みモデルを連合学習の枠組みにおいて相互に連携・集約することで，翻訳モデルの性能向上を図る．

3.1　Federated Average

Federated Average（FedAvg）[4] は，データを共有することなく協調学習を実現する，連合学習の代表的かつ基礎的なアルゴリズムである．FedAvg では，サーバとクライアントが反復的に相互作用することで，各クライアントが保持するローカルデータを外部に公開することなく，グローバルモデルを協調的に学習する．

図 2 に示す動作プロセスにおいて，サーバは各ラウンドごとにクライアントをランダムに選択し，現在のグローバルモデルを配布する．選択されたクライアントは，受信したモデルを自身のローカルデータで学習し，更新後のモデルパラメータをサーバへ送信する．サーバは，各クライアントが保有するサンプル数に基づく重み付き平均を用いてこれらの更新を統合し，次ラウンドのグローバルモデルを生成する [4].

定式化すると，まず各学習ラウンドにおいて，サーバは利用可能なすべてのクライアントを選択し，現在のグローバルモデル $w_g^{(t)}$ をそれぞれに送信する．

表 1: 記号一覧表

記号	説明
τ	各クライアントのローカルトレーニング反復回数
$\tau_{\max}$	各ラウンドにおけるローカル学習の最大反復回数
$w_g^{(t)}$	連合学習における反復回数 t のグローバルモデル
$w_i^{(t,\tau)}$	連合学習における反復回数 t において，τ 回のローカル学習後のクライアント c_i のローカルモデル
$w_i^{(t)}$	ラウンド t において，$\tau_{\max}$ 回のローカル学習を完了した後のクライアント c_i の最終ローカルモデル（$w_i^{(t)} = w_i^{(t,\tau_{\max})}$）
n	クライアントの総数
D	全クライアントの集合 $D = \{c_1, c_2, \ldots, c_n\}$
D_k	クライアント集合 D の部分集合 $D_k \subseteq D$
$\mathcal{P}(D)$	クライアント集合 D の冪集合（すべての部分集合の集合）
$W_g^{(t+1)}$	すべての部分集合 $D_k \in \mathcal{P}(D)$ から生成されるグローバルモデル $w_{g,k}^{(t+1)}$ の集合
η	ローカルモデル更新における学習率
$F_i(w)$	モデル w に対するクライアント c_i のローカル損失関数
$\nabla F_i(w)$	クライアント i のローカル損失関数の勾配
$s_i^{(t)}$	連合学習の反復回数 t におけるクライアント i の状態
A	行動集合（例：選択可能なクライアント部分集合）
$a \in A$	行動集合から選択された行動
$Q(s_i^{(t)}, a)$	状態 $s_i^{(t)}$ において行動 a を取るときの，クライアント i の Q 値
α	Q 学習における学習率
γ	Q 学習における割引率

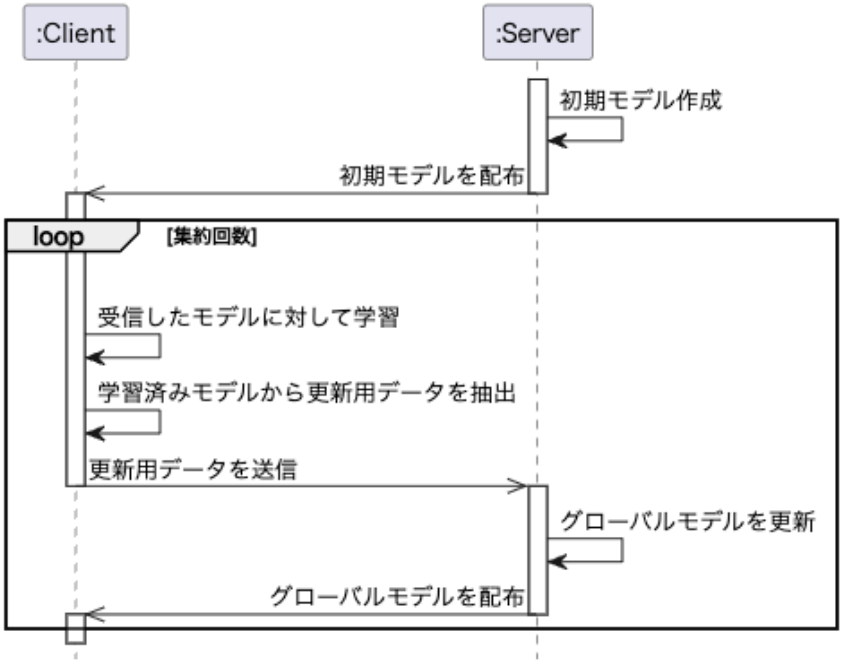

図 2: FedAvg のシーケンス図

各クライアント c_i は，ローカルモデルを $w_i^{(t,0)} = w_g^{(t)}$ として初期化し，自身のローカルデータセット上でモデルを更新する．具体的には以下のように配降下法によるローカル更新を繰り返す：

$$w_i^{(t,0)} = w_g^{(t)}$$
$$w_i^{(t,\tau+1)} = w_i^{(t,\tau)} - \eta \nabla F_i(w_i^{(t,\tau)})$$

$\tau_{\max}$ 回のローカル学習を完了した後，各クライアントは更新後のローカルモデル $w_i^{(t)} = w_i^{(t,\tau_{\max})}$ をサーバに送信する．サーバは，受信したすべてのローカルモデルを集約し，次ラウンドのグローバルモデルを次式で算出する：

$$w_g^{(t+1)} = \frac{1}{n} \sum_{i=1}^{N} w_i^{(t)}$$

この集約方法では，データ分布の違いに関わらず，すべてのクライアントが等しく貢献することが前提となっている．そのため，FedAvg は多くの状況で有効である一方，クライアント間でデータ分布が大きく異なる Non-IID 環境では，更新が偏り，グローバルモデルの精度が低下するという問題が生じる．

3.2 greedy 法に基づく自己組織化手法

FedAvg が抱える課題を緩和することを目的として，本研究では greedy 法に基づく自己組織化手法を導入する．この手法では，モデル集約に参加するクライアントを動的に最適化することで，自身のドメインに適合した NMT モデルを構築することが可能である．

greedy 法とは，大域的な最適解を一度に求めるのではなく，問題を複数の部分問題に分割し，各段階において局所的に最適な選択を繰り返すことで，全体として良好な解を得る手法である．

本研究で用いる greedy 法は，連合学習に参加する各クライアントが，自身の管理する対訳コーパスのドメインに特化した NMT モデルを構築することを目的とした自己組織化手法である．図 3 に示すように，まずサーバは初期グローバルモデルを生成し，各クライアントへ配布する．各クライアントは受け取ったモデルを用いてローカル学習を行い，所定の学習ステップに到達した後，更新されたモデルをサーバへ送信する．

次に，サーバは各クライアントから収集したローカルモデルを基に，クライアント集合の全ての組み合わせに対する集約モデルを生成する．その後，各クライアントはサーバから全ての集約モデルを取得し，評価用データを用いてローカル環境で翻訳精度を測定する．各クライアントは，最も高い翻訳精度を示した集約モデルを次ラウンドにおけるグローバルモデルとして選択し，次の学習フェーズへ進む．

最終ラウンドの集約後，各クライアントは再び全ての集約モデルを評価し，最良のモデルに対して自身が保有する学習データを用いて追加学習を行うことで，自身のドメインにおける翻訳精度のさらなる向上を図る．このように greedy 法による自己組織化手法を用いることで，各組織は連合学習の枠組みを維持しつつ，自身のドメインに適合した NMT モデルを構築することが可能となる．

以上より，本研究で提案する greedy 法に基づく自己組織化手法は，複数の組織が連合学習により NMT モデルを構築するという大域的な問題を，ローカルモデルの学習とグローバルモデルの集約という一連の部分問題に分解し，各集約段階において翻訳精度を最大化するモデル選択を行うことで，全体性能の向上を実現する手法である．

定式化すると，本手法では，まずサーバが参加クライアントの異なる部分集

合から複数の候補グローバルモデルを生成する．全クライアント集合を $D = \{c_1, c_2, \ldots, c_n\}$ とし，その冪集合を $\mathcal{P}(D)$ と定義する．各部分集合 $D_k \in \mathcal{P}(D)$ は，モデル更新に参加するクライアントの組み合わせを表す．部分集合 D_k に対して生成されるグローバルモデル $w_{g,k}^{(t+1)}$ は，次式により計算される：

$$w_{g,k}^{(t+1)} = \frac{1}{|D_k|} \sum_{j \in D_k} w_j^{(t)}$$

これにより生成されるすべての候補グローバルモデルの集合は，

$$W_g^{(t+1)} = \{w_{g,k}^{(t+1)} \mid D_k \in \mathcal{P}(D)\}$$

と定義され，これはサーバが生成可能なすべての集約結果を表している．各クライアントは，これらの候補モデルをサーバから受信し，自身のローカルデータセット上で評価を行う．その上で，最も高い精度を示すグローバルモデル $w_{g,i}^{(t+1)}$ を選択する：

$$w_{g,i}^{(t+1)} = \underset{w_{g,k}^{(t+1)} \in W_g^{(t+1)}}{\arg\max} \; Accuracy_i\left(w_{g,k}^{(t+1)}\right)$$

ここで，$Accuracy_i(w)$ はモデル w をクライアント c_i のローカルデータで評価した際の精度である．

このように，各クライアントがローカル性能に基づいて最適なモデルを選択することで，データ分布の異質性に対応した柔軟な集約が実現され，全体的な翻訳精度の向上が期待できる．

3.3 マルチエージェント強化学習に基づく自己組織化手法

マルチエージェント強化学習は，複数のエージェントが同時に学習・行動し，互いに影響を与え合う自律分散システムが構築される．MARL に基づく自己組織化手法は，図 4 のようにマルチエージェント強化学習を連合学習システムに統合し，連合学習の枠組みに動的な意思決定機構を導入するものである．FedAvg や greedy 型の自己組織化手法とは異なり，本手法では，クライアントが他クライアントのモデルや振る舞いに応じて適応的な協調方策を学習することが可能となる．各クライアントは自律的なエージェントとして振る舞い，環境を観測し，行動を選択し，サーバおよび他クライアントとの相互作用を通じて方策を

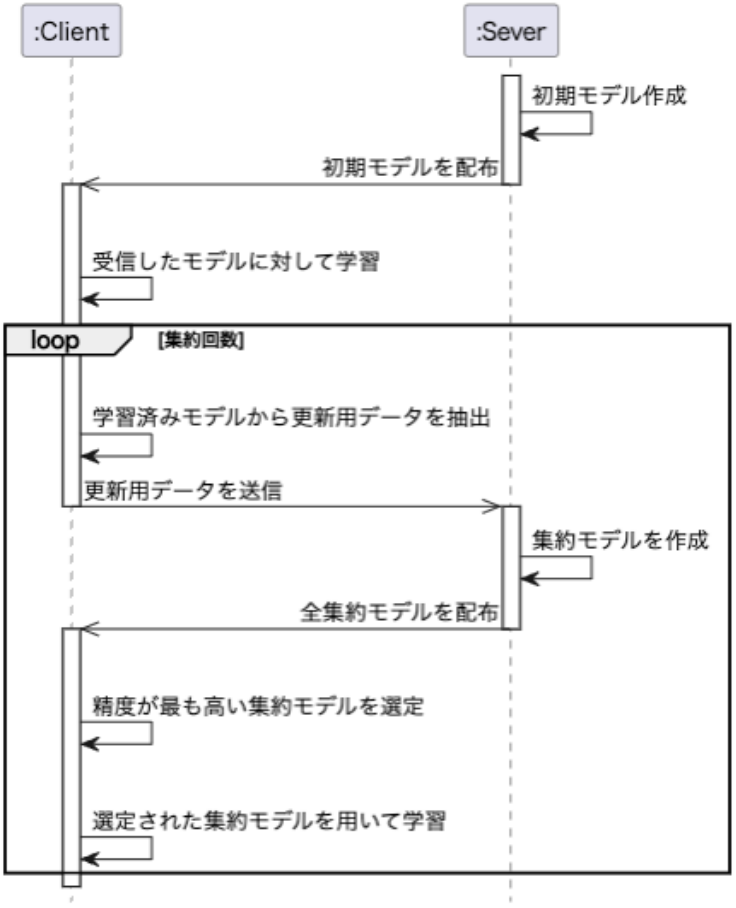

図 3: greedy 法に基づく自己組織化手法のシーケンス図

更新する．

図 5 に示すように，まずサーバは初期グローバルモデルを生成し，各クライアントへ配布する．各クライアントは受け取ったモデルを用いてローカル学習を行い，所定の学習ステップに到達した後，更新されたモデルをサーバへ送信する．

各連合学習のラウンド t において，クライアント c_i は，前回の集約で得られたグローバルモデルを用いてローカル学習を行う．その後，対応するエージェントが学習結果を評価し，現在の状態 $s_i^{(t)}$ を観測する．この状態に基づき，深層強化学習エージェントは次ラウンドの集約に向けた行動 $a_i^{(t)}$ を選択する．行動は，単一クライアントのみで連携する場合や，複数クライアントを同一組織化として構成する場合などを含む．行動集合 A はクライアント集合の冪集合として定義され，$A = \mathcal{P}(D)$ と表される．クライアント数を n とすると，可能な

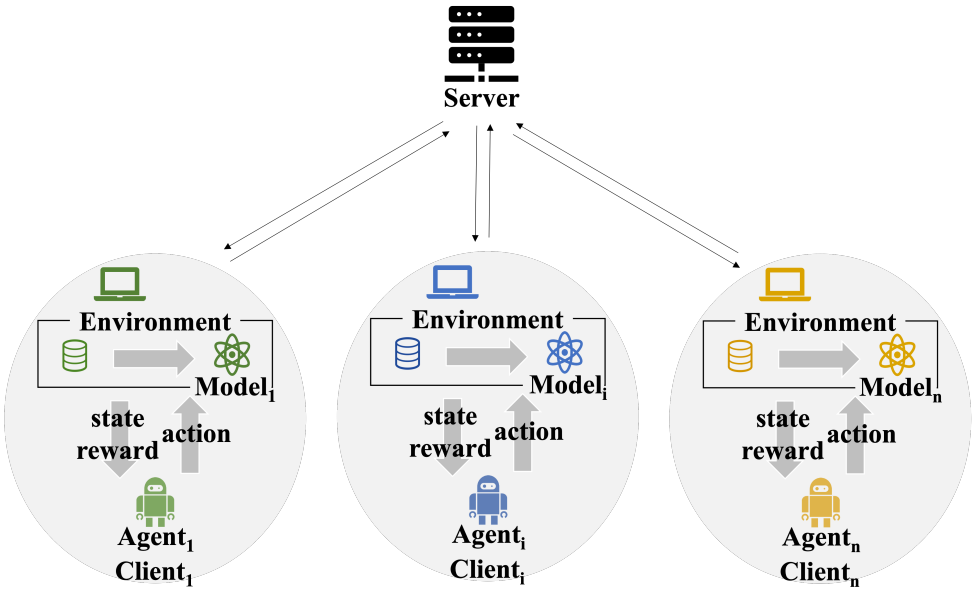

図 4: MARL を統合した FL

組織化構成（行動）の総数は次式で与えられる：

$$\sum_{k=1}^{n} \binom{n}{k} = 2^n - 1 \tag{1}$$

行動 $a_i^{(t)}$ を選択した後，エージェントは対応するクライアントに対して，該当する集約要求とローカルモデルの送信を指示する．サーバは，各組織化要求に従って受信したローカルモデルを集約し，新たなグローバルモデル $w_{g,i}^{(t+1)}$ を生成してクライアントへ配布する．各クライアントはこの新しいグローバルモデルを用いて再びローカル学習を行い，エージェントはその結果を評価して報酬 $r_i^{(t+1)}$ を算出し，次状態 $s_i^{(t+1)}$ を観測する．これらを用いて，エージェントは以下の Q 学習更新式により行動価値関数を更新する [25].

$$\begin{aligned} Q(s_t, a_t) \leftarrow\ & Q(s_t, a_t) \\ & + \alpha \left(r_{t+1} + \gamma \max_{a_{t+1} \in A} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \right) \end{aligned} \tag{2}$$

学習された Q 値は，次回以降の組織化構成の意思決定に利用される．ラウンド t において，各クライアント c_i のエージェントは，期待効用を最大化するク

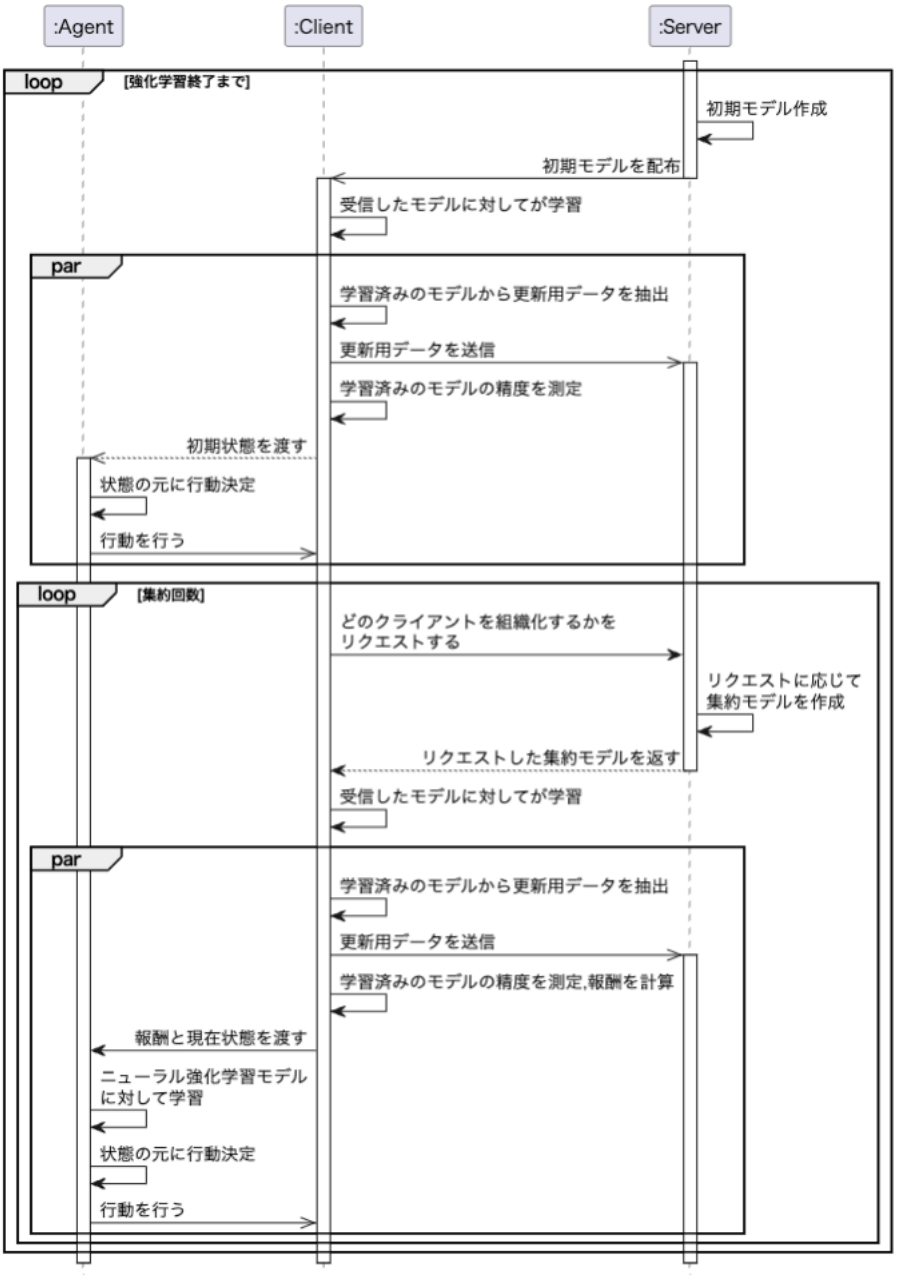

図 5: MARL 統合手法のシーケンス図

ライアント部分集合 $D_k \subseteq D$ を行動として自律的に選択する．

$$a_i^{(t)} = D_i = \underset{D_k \in \mathcal{P}(D)=A}{\arg\max}\ Q(s_i^{(t)}, D_k)$$

ここで，$Q(s_i^{(t)}, D_k)$ は，部分集合 D_k と協調することによって得られる期待利益を定量化したものである．このように，本手法は「状態観測 → 行動選択

→ 報酬獲得 → 方策更新」というフィードバックループを形成する．この方策駆動型の自己組織化機構により，本システムは Non-IID なデータ分布に対して動的に適応し，ドメイン特性を考慮した効果的な協調を実現する．各クライアントが自身のローカルモデルに最も適した協調相手を自律的に選択できるため，連合型 NMT における翻訳精度の向上が期待される．

3.4 手法の比較

本節では，本章で紹介した三つの手法——FedAvg，greedy 法に基づく自己組織化手法，および MARL 統合手法——の相違点を明確化するため，各手法がクライアント選択およびモデル集約をいかに扱うかに着目して論じる．

まず FedAvg は，全クライアントが各ラウンドに一様に参加し，それぞれのモデルパラメータを単純平均することでグローバルモデルを更新する．しかしながら，この一様な集約はクライアント間のデータ分布の不均一性を考慮しておらず，特に Non-IID 環境においては最適性を欠く更新を招く可能性が高い．

これに対し，greedy 法に基づく自己組織化手法は，グローバル性能の即時的な向上が期待されるクライアント集合を選択的に集約することで，FedAvg の問題点を部分的に緩和する．ただし，この戦略はヒューリスティックあるいは静的な最適化に依存しており，クライアントのデータ特性や協調関係が動的に変化する状況では適応性に限界が生じる．

一方で，提案する MARL ベースのフレームワークは，各クライアントをエージェントとみなし，強化学習の報酬構造を通じて協調ポリシーを学習させる．特定の静的ルールに依拠するのではなく，エージェントは過去の相互作用から得られる環境情報に基づいて自律的に最適な協調相手を決定する．その結果，データ分布の類似したクライアント間に自然な自己組織的協調が形成され，Non-IID 環境下においても安定的かつ効果的な知識共有が実現される．

第4章　NMT向けMARLエージェントの設計

本章では，NMT タスクを対象として設計した MARL エージェントの構成要素および設計方針について詳述する．

本研究では，連合学習に参加する各クライアントを自律的に意思決定を行うエージェントとして捉え，エージェント同士が状況に応じて協調関係を形成しながら翻訳モデルを学習・改善していく枠組みを採用する．MARL に基づく連合学習を NMT に適用するにあたり，本研究ではエージェントを以下のようにモデル化する．まずエージェントが観測する状態の設計について述べ，次に連携関係を制御する行動空間を定義する．最後に，翻訳性能の向上を学習目標とするための報酬設計について詳しく説明する．

4.1　状態

図 6 のように，各クライアント c_i にはエージェントが対応付けられており，ローカルモデルを評価することで現在の状態 $s_i^{(t)}$ を取得する．連合学習ラウンド t におけるクライアント c_i のローカル NMT モデルを $w_i^{(t)}$ とし，これは入力文 src を翻訳文へ写像する関数として扱う．評価には，固定されたテストセット

$$\mathcal{T}^{\text{test}} = \{(src_j, ref_j)\}_{j=1}^{k}$$

を用いる．翻訳品質の評価指標としては，システム出力と参照訳の n-gram 一致度を測定する標準的な指標である BLEU [26] を用いる．

状態 $s_i^{(t)}$ は，テストセット $\mathcal{T}^{test}$ に対する文単位 BLEU スコアのベクトルとして定義される：

$$\begin{aligned} s_i^{(t)} = \big[&\text{BLEU}\big(w_i^{(t)}(src_1),\, ref_1\big),\\ &\text{BLEU}\big(w_i^{(t)}(src_2),\, ref_2\big),\\ &\cdots,\\ &\text{BLEU}\big(w_i^{(t)}(src_k),\, ref_k\big)\big]. \end{aligned}$$

特に断りのない限り，初期状態 $s_i^{(0)}$ は，クライアント c_i が初期グローバルモデル $w_g^{(0)}$ を用いて学習を行った後に算出される．

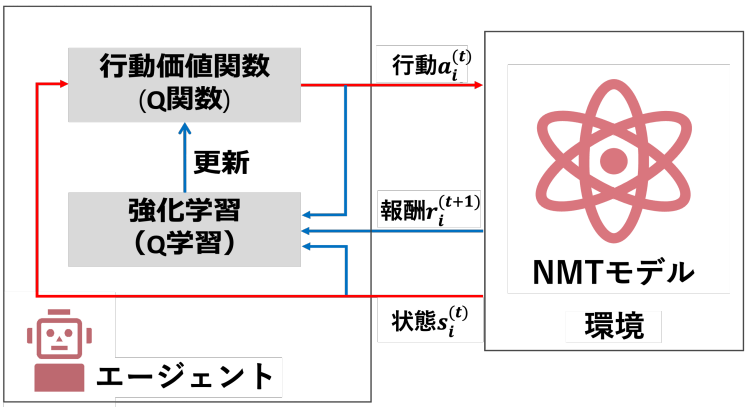

図 6: MARL エージェント

4.2 報酬

状態 $s_i^{(t+1)}$ を取得した後，各エージェントは，自身の行動 $a_i^{(t)}$ によって生じた翻訳性能の変化を反映する報酬 $r_i^{(t+1)}$ を計算する．報酬 $r_i^{(t+1)}$ は，連続する 2 ラウンド間におけるローカル NMT モデルの翻訳精度の向上量を表す：

$$r_i^{(t+1)} = score_i^{(t+1)} - score_i^{(t)}$$

ここで $score_i^{(t)}$ は，連合学習ラウンド t におけるテストデータセット $\mathcal{T}^{test}$ 上での文単位 BLEU スコアの平均値を示す．

報酬設計は，翻訳性能を向上させる意思決定へとエージェントを導く上で極めて重要である．エージェントは環境の観測と内部モデルの両方に基づいて意思決定を行い，将来の性能変化を予測しながら協調相手を選択する．したがって，分散環境で動作するエージェントにとって，翻訳性能を正確に推定する内部モデルは不可欠である．本研究では，エージェントの報酬は，特定のテストデータセット上で評価されたローカル NMT モデルの BLEU 性能に基づいて定義される．

4.3 行動

エージェントは，ペアになっているクライアントの環境をセンシングし，現在の状態 $s_i^{(t+1)}$ を取得する．次に，深層強化学習エージェントは状態 $s_i^{(t+1)}$ のもとで，自分が保持している強化学習モデルを使用して次の行動 $a_i^{(t+1)}$ を決定する．ここでの行動 $a_i^{(t+1)}$ は，どのクライアントと組織化するかを指す．組織内

には複数のクライアントが存在することもあるが，単一のクライアントが組織になることも考えられる．すなわち，クライアント数が n の時，行動の種類は 3.3 節の式 (1) のように表せる．これは，クライアント集合 D の冪集合 $\mathcal{P}(D)$ から空集合を除いた集合に対応しており，エージェントが選択できる協調形態の多様性を表している．

さらに，エージェントは選択した行動 $a_i^{(t+1)}$ に基づき，3.3 節の式 (2) に示した行動価値関数を参照しながら，クライアントがサーバに送信する集約リクエストの内容を制御する．これにより，各エージェントは自身の目的に応じた協調関係を動的に構築することが可能となる．

第5章　エージェントの内部モデル

本章では，本研究で設計した MARL エージェントが，連合型 NMT 環境においてどのような内部モデルを形成し，それに基づいて行動選択を行うかについて述べる．特に，エージェントが翻訳性能をどのように認識・評価し，その情報を用いて利己的あるいは協調的な戦略を選択する仕組みに焦点を当てる．

5.1　報酬設計

一般に，エージェントは外部環境を直接観測することはできず，限られた観測情報から環境の状態を推定し，意思決定を行う．このとき用いられる内部モデルは，環境の構造や自身の行動がもたらす影響を表現する役割を担う．連合型 NMT においては，各クライアントが異なるデータ分布を持つため，翻訳性能の評価基準そのものがエージェントごとに異なり得る．本研究では，この評価基準を内部モデルの中心的要素として位置づけ，翻訳精度を通じて環境を理解する枠組みを採用する．

具体的には，本研究における内部モデルは，「現在の連携関係や学習結果が，自身の翻訳性能にどのような影響を与えているか」を表現するものと定義する．翻訳性能の測定には，事前に用意した検証データセットを用い，その構成の違いによってエージェントの行動傾向が変化するよう設計する．この検証データセットの違いが，結果として利己的戦略と協調的戦略という二つの内部モデル設計の差異を生み出す．

本研究では，報酬設計に用いる検証データセットとして，以下のように二種類を定義する．

- **ドメイン特化型データ**：法律, 医療など, 特定の専門領域に限定されたデータ
- **一様分布型データ**：複数の領域を均等に含む，多様性を重視したデータ

5.2　利己的エージェント

ドメイン特化型データを評価基準として内部モデルを構築するエージェントは，自身のドメインにおける翻訳性能の最大化を重視する利己的戦略を採用する．この戦略では，特定ドメインにおける翻訳精度の向上が直接的な目的となるため，専門用語や分野固有の表現への適応が強く促される．法務，医療，技術文書といった専門領域では，文脈依存性が高く，翻訳の正確性が厳しく要求

される．利己的エージェントは，こうした要求に応える形で局所的な最適化を進めるため，担当ドメインにおいて高い翻訳品質を達成しやすい．一方で，他ドメインに対する汎化性能は内部モデル上で明示的に考慮されないため，適用範囲が限定される可能性がある．

5.3　協調的エージェント

一様分布型データを評価基準とするエージェントは，全ドメインに対してバランスの取れた性能向上を目指す協調的戦略を採用する．この内部モデルでは，単一ドメインの性能ではなく，複数ドメインにまたがる平均的な翻訳性能が重視される．複数領域を含む検証データを用いることで，エージェントは自身の行動や他エージェントとの連携が，全体的な翻訳品質に与える影響を学習する．その結果，特定分野に過度に偏らない汎用的な翻訳モデルの構築が促進される．協調的エージェントは，連合学習における知識共有や相互補完の効果を最大化する役割を担う．

5.4　二つの戦略の意義

本研究は，利己的戦略と協調的戦略という異なる内部モデル設計を併存させることで，連合学習における根本的課題である「専門性」と「汎用性」の両立を図る．すなわち，個々のクライアントが自身のデータ特性に基づいて性能を最大化しようとする要求と，全体としての協調的な性能向上とのトレードオフを明示的に扱う点に，本研究の特徴がある．

提案する報酬モデルにより，各エージェントは連携によって生じた翻訳性能の変化を内部モデルに反映し，状況に応じて利己的行動と協調的行動を選択できる．このような戦略の違いを内部モデルとして明確に区別する枠組みは，連合学習における「クライアント個別最適化」と「全体最適化」の関係を分析する上で有効な視点を提供する．

利己的戦略は特定ドメインで高い精度を実現しやすい一方で，汎用性が低下する可能性がある．これに対し，協調的戦略は広範なドメインへの適応性を持つが，専門領域での性能が相対的に抑えられる場合がある．本研究は，これら二つの戦略を対立概念としてではなく，相補的に扱うことで，連合学習環境下における NMT システムの性能向上に新たな可能性を示す．

第6章　LLMエージェントによる協調翻訳

6.1　LLMにおける翻訳タスクへの応用

近年，ChatGPT や Llama3 に代表される大規模言語モデル（Large Language Model; LLM）の研究は急速に発展しており，多様な分野への応用が進展している．モデル規模の拡大に伴って内包される知識量が増大し，汎用的な推論能力や高度な対話能力が向上した結果，問い合わせ対応，教育支援，情報検索，ソフトウェア開発補助，エンターテインメントなど，幅広いタスクにおいて高品質な応答生成が可能となった．これらのモデルは，単に正確な情報を提示するだけでなく，文脈や背景知識，ユーザー意図を考慮した柔軟な生成を実現しており，ユーザー体験の向上に大きく寄与している．

一方で，LLM の実用性が高まるにつれ，特定領域に特化したモデルに対する需要も依然として高い．金融，法務，医療といった専門性の高い分野では，汎用モデルのみでは十分な性能が得られない場合が多く，Google 社による医療特化モデル Med-PaLM [27] など，ドメイン特化型 LLM の研究開発が進められている．また，Cheng ら [28] は，特定領域データを用いた継続学習が，当該ドメインにおける性能向上に有効であることを示している．しかし，ドメイン特化型モデルの構築には大量の学習データが必要であり，多言語対応や領域偏り，データ収集コストといった課題が依然として残されている．特に専門領域のデータは英語に偏在する傾向が強く，日本語を含む多言語で高性能なモデルを構築することは容易ではない．

さらに，日本語に関しては，高性能かつオープンな LLM が依然として不足している．ChatGPT をはじめとするクローズドモデルは日本語において高い性能を示すものの，内部構造や学習過程が非公開であるため研究応用の柔軟性に制約があり，再学習や制御が困難である．また，機密情報や個人情報を扱う環境では，プライバシー保護の観点から利用が制限される場合も多い．特に GPT-5 系モデルは最高水準の性能を有する一方でブラックボックス性が強く，学術研究や産業応用におけるモデルの制御や検証が難しいという課題がある．このため，日本語に特化しつつ，安全性と透明性を確保した高性能 LLM の構築手法を確立することは，国内における生成 AI 技術の普及と利活用を推進する上で重要な課題である．

また，LLM の利便性向上には，ユーザーごとに最適化された応答を生成する

パーソナライゼーション技術の発展も不可欠である．ユーザーの嗜好や作業履歴，専門知識レベルに応じて応答を適応させる手法は数多く提案されているが，個人情報を安全に扱いながら柔軟にモデルをカスタマイズすることは依然として困難である．特に，大規模モデルを再学習することは計算資源およびコストの面で非現実的であるため，近年ではプロンプト設計，LoRA，知識編集，報酬モデル制御など，既存モデルの振る舞いを局所的に調整する軽量な手法に注目が集まっている．

以上のように，LLM は自然言語処理分野の中核技術として急速な発展を遂げている一方で，ドメイン特化，多言語対応，日本語性能の向上，パーソナライゼーション，および学習コストの削減といった課題が依然として残されている．これらの課題を解決するための新たな技術的枠組みの確立は，次世代 LLM の実用化および社会実装に向けた重要な研究テーマである．

6.2 LLM の事後学習

LLM を特定タスクや特定領域に適応させるためには，モデル全体を再学習するのではなく，既存モデルの能力を最大限に活用しつつ，必要な部分のみを局所的に補正する事後学習（post-training）の重要性が高まっている．事後学習は，計算資源やデータ量に制約のある実環境において，効率的なモデル適応を実現する手法として注目されている．代表的な事後学習のアプローチは，大きく以下の三つに分類できる．

(1) プロンプトベース適応　プロンプトにタスク指示や制約条件，スタイル情報などを付与することで，モデルの出力挙動を制御する手法である．プロンプトエンジニアリング，In-Context Learning（ICL），Chain-of-Thought（CoT）などがこれに含まれる．これらの手法は，モデル内部のパラメータを更新することなく性能向上を図れる点が特徴であり，学習データが限られた状況や迅速な適応が求められる場合に有効である．

(2) パラメータ効率の高い微調整　LoRA，QLoRA，Adapter などの手法により，モデル全体ではなく一部の低ランクパラメータや付加的モジュールのみを更新することで，少量のデータでも効率的な知識適応を実現する．これらの手法は，専門領域知識の付与，日本語特化，出力スタイルの制御などの用途で広く利用されており，大規模モデルの計算コストを抑えつつ高い適応性能を達成できる点が利点である．

(3) 事後的な知識編集　既存モデルに内在する特定の知識のみを直接更新・修正する手法であり，ROME，MEMIT，SERAC などが代表例として挙げられる．これらの手法は，特定の事実更新や誤り訂正を低コストで実現できるため，頻繁な知識更新や即時性が求められるアプリケーションにおいて有効である．

これらの事後学習手法は，モデル全体を再学習する従来手法と比較して，計算資源，必要データ量，および学習時間を大幅に削減できる点が大きな利点である．本研究で対象とする連合学習環境においても，事後学習は通信量の削減やクライアント側の計算負荷低減に寄与し，ドメイン特化型知識を安全かつ柔軟に付与する手法として極めて有効である．

6.3　LLM マルチエージェントの連携

LLM は高い言語生成能力を有する一方で，単一モデルでは特定ドメインへの適応や，分散環境における柔軟な運用に限界がある．この課題に対し，本研究では複数の LLM をエージェントとして配置し，各エージェントを特定ドメインに特化させた上で協調的に翻訳を行う枠組みを導入する．本節で扱う手法は，強化学習による自動的な組織化を用いず，あらかじめ定義された協調方式に基づいて翻訳を行う手法であり，本研究で提案する強化学習を用いた連合翻訳の自動組織化手法と比較するためのベースラインとして位置づけられる．

本枠組みにおいて，各 LLM エージェントは異なる専門領域に基づく知識を活用して翻訳生成を行い，エージェント間の連携を通じて知識の相互補完を実現する．これにより，専門性を維持しつつ汎用性を向上させることが可能となり，単一モデルでは達成が困難であった高品質な翻訳生成が期待される．この考え方は，分散データ環境において複数クライアントが協調する連合型 NMT の枠組みと概念的に対応しているが，本手法では意思決定や組織化を学習によって最適化することは行わない点が大きな違いである．

本研究では，LLM 翻訳エージェントの協調方式として，以下の二つのメカニズムを検討する．これらはいずれも，固定的なルールに基づく協調翻訳方式であり，強化学習を用いた自動組織化手法との比較対象として用いる．

(1) 並列型アプローチ　複数の LLM エージェントが同一の入力文に対して独立に翻訳を生成し，それらの翻訳結果に対して LLM エージェント間の多数決により最終的な翻訳を決定する方式である．各エージェントは同条件下で翻訳を行い，得られた複数の翻訳候補の中で最も多く支持されたものを採用する．本

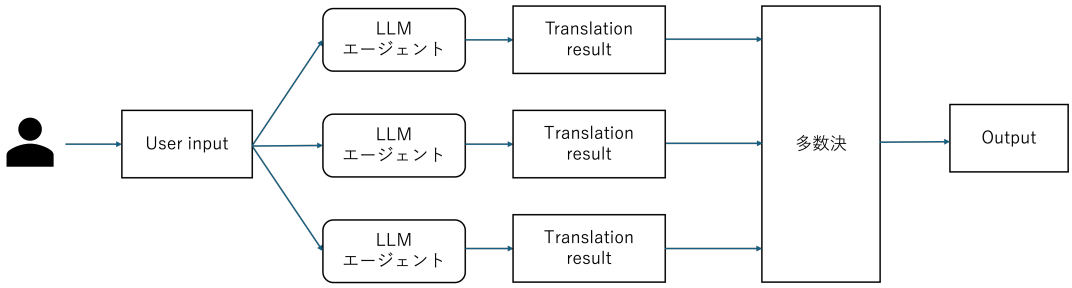

図 7: 並列型アプローチ

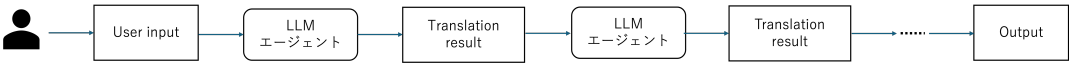

図 8: 直列型アプローチ

方式では，多様な翻訳候補を同時に生成できるため，誤訳の抑制や品質向上に加え，個々のエージェントのばらつきを低減することによる翻訳結果の安定性向上が期待できる．

(2) 直列型アプローチ　複数の LLM エージェントが順次翻訳を行い，後続エージェントが前段の翻訳結果を参照しながら修正や改善を加える方式である．スタイルの統一や専門知識の段階的付与など，逐次的な改善が求められる翻訳タスクに適している．

以上のような LLM エージェントによる協調翻訳は，高度な言語生成能力を活用しつつ，複数モデル間の知識補完を実現する有効な手法である．一方で，エージェント間の連携構造や協調相手の選択は事前に設計されたルールに依存しており，環境やデータ分布の変化に対して自律的に最適化されるわけではない．この点において，本研究で扱う強化学習を用いた連合翻訳の自動組織化手法との比較を通じて，両者の特性と有効性を明らかにする．

第7章　実験環境

7.1　モデルの構築

本研究で提案するシステムでは，最大5つの組織がそれぞれ1つの FL クライアントとして機能し，共同で1つの FL サーバを維持する構成を採用している．NMT モデルの構築には OpenNMT [22] を使用した．NMT モデルはバッチサイズ 32 で動作し，合計 5000 ステップの学習を行う．FL プロセス全体は5回のイテレーションで構成されており，学習の一定間隔ごとにサーバが 5 回の集約を実施する．

一方で，本研究で用いた LLM は *LLM-jp-3-13B* である．本モデルは国立情報学研究所によって開発された日本語処理に特化した大規模言語モデルであり，LLM-jp-3 シリーズの一つに位置づけられる．日本語データセットを用いた継続事前学習（continued pre-training）により構築されたベースモデルであり，日本語において高い性能を示すと同時に，英語に対する言語能力も維持している．

7.2　データセット

使用した日英対訳コーパスは，「Wikipedia 日英京都関連文書対訳コーパス」および「日英法令対訳コーパス - Graham Neubig」の 2 種類である．「Wikipedia 日英京都関連文書対訳コーパス」は 15 のカテゴリに分かれており，このうち「文化」「歴史」「偉人」の 3 カテゴリから 8 万件の対訳データを抽出し，それぞれのクライアントに割り当てた．同様に，「日英法令対訳コーパス - Graham Neubig」からは，「Law」および「Law2」カテゴリの各 8 万件を抽出した．これらのデータセットは短文で構成され，各データセットですべて同一の翻訳者によって作成されているため，整合性が高い．各クライアントは，自身のデータセットを NMT モデル構築のための学習データと，モデル性能を評価するための評価データに分割して使用した．

本研究では，データ分布の条件として，独立同一分布（iid）および非独立同一分布（Non-iid）の 2 つの設定を採用した．iid 設定では，各クライアントに割り当てられるデータが同一データセット内のドメイン（例：「文化」「歴史」「偉人」や「Law」「Law2」）で均一かつランダムに分布しており，すべてのクライアントが類似した特徴を持つデータで学習を行う．これにより，クライアント間の学習条件が均質であり，協調学習の性能を純粋に比較することが可能となる．

一方，Non-iid 設定では，各クライアントに異なるデータセットのドメイン (例：「文化」「歴史」「偉人」「Law」や「文化」「歴史」「偉人」「Law」「Law2」) のデータを割り当て，データ分布に偏りを持たせている．この設定では，クライアントごとに扱うデータの特性が異なるため，モデル間での知識共有や協調の仕方が翻訳性能に大きく影響する．このような非均質環境下での評価により，異なるドメイン間での協調戦略の有効性を検証することを目的とした．

7.3 評価指標

NMT モデルの精度評価には，機械翻訳の品質指標である BLEU (Bilingual Evaluation Understudy) スコア [26] を用いた．BLEU スコアは，機械生成翻訳と人手翻訳との間の n-gram 一致度を測定する指標であり，値が高いほど翻訳精度が高いことを示す．提案手法の有効性を検証するため，従来の連合学習手法である FedAvg を比較対象として翻訳精度を評価した．さらに，性能評価の信頼性を高めるために 5 分割交差検証（5-fold cross-validation） を実施した．これは，データセットをランダムに 5 つの部分集合に分割し，4 つを学習用，1 つを評価用として使用する方法である．この手順を 5 回繰り返し，得られた評価結果の平均値をモデル性能の指標とした．

7.4 実装

7.4.1 深層強化学習エージェントの実装

MARL を用いた組織化手法を評価するため，各組織に深層強化学習エージェントを追加した．エージェントの実装には，PyTorch 向けの深層強化学習ライブラリ PFRL を用いた．各エージェントは，対応するクライアントのデータを通じて環境状態を観測する．初期状態空間のサイズは，クライアントが管理する検証データ数に等しい．本実験では，状態空間のサイズを 1,000 に設定した．また，エージェントの行動空間のサイズは組織数に応じて変化する．具体的には，組織が 3 つの場合は 7，4 つの場合は 15，5 つの場合は 31 とした．強化学習の過程では，5 回の集約が実行される．各クライアントが指定された学習ステップを完了し，NMT モデルを構築するたびに，1 エピソードとしてカウントした (図 9 参照)．

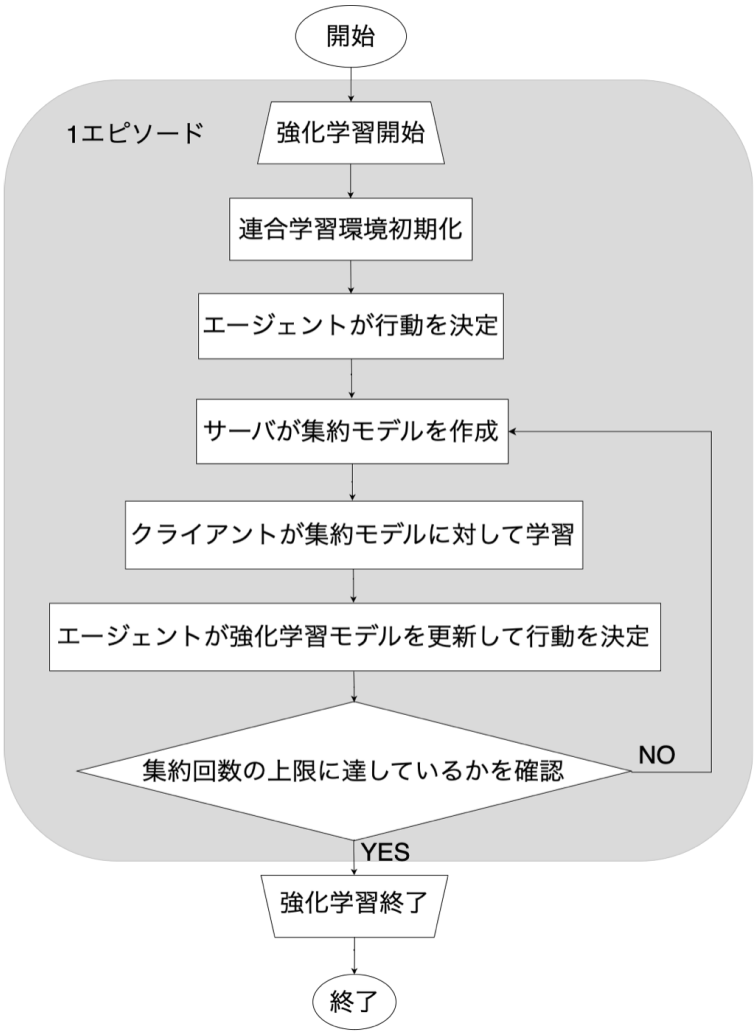

図 9: MARL 手法のフロー

7.4.2 LLM のファインチューニング

本研究では，本モデルを各データセットに対応した翻訳タスクへ適応させるため，プロンプトを用いたファインチューニングを実施した．ファインチューニングに先立ち，学習データの品質を確保するためにノイズ除去や不整合データの排除といったフィルタリング処理を行った．その後，翻訳タスク用に設計したプロンプトを用いてモデルのファインチューニングを行った．

本研究で使用した翻訳タスク用プロンプトを図 10 に示す．本プロンプトは，日本語文を入力として与え，余分な説明を含まない英語翻訳文のみを出力するよう明示的に指示している．なお，{input} は翻訳対象となるデータセット中の日本語文を指すプレースホルダであり，学習および推論の各段階において個々

```
### Instruction
Translate the following Japanese sentence into English.
Output only the translated text without any additional explanations.
Japanese: {input}
### Response
English:
```

図 10: 翻訳タスクに用いたプロンプト

の入力文が対応付けられる.

第8章　検証結果

8.1　ニューラル機械翻訳の連合学習

複数回の連合学習の反復を通じてNMTモデルを構築し，MARLを用いて合計460エピソードにわたって学習を行った．得られたポリシーはまず，iidのデータセットを用いた新たな連合学習環境で再評価し，翻訳精度を測定した．その後，Non-iid環境下で，異なるドメインのデータセットをクライアント間で混在させた条件で評価を行った．この二段階の評価設計により，均質なドメイン内および異質なドメイン間の両方において，異なる協調戦略が翻訳精度向上にどの程度寄与するかを検証した．

表2は，文化系および法務系コーパスを基に構築した2種類のiidデータセットにおける翻訳精度を示している．また，表3および表4は，全クライアントを対象としたNon-iid構成における結果を示している．

まず，iid・Non-iidのいずれの設定においても，他ドメインとの協調や情報共有を行わずに独立して学習を行うStand Alone方式は，すべてのカテゴリで最も低い翻訳精度を示した．この結果は，単一クライアントのデータのみに基づいて学習したNMTモデルでは，データ量および表現多様性の制約により，性能向上に明確な限界が存在することを示している．

これに対し，従来手法であるFedAvgは，すべてのカテゴリにおいてStand Aloneを上回る精度を達成しており，iid条件下ではモデル平均化による協調学習が一定の有効性を持つことが確認できる．しかしながら，その改善幅は限定

表2: 文化領域および法領域の2つのiidデータセットにおける，各手法の翻訳精度比較

ドメイン	カテゴリ	Stand Alone	FedAvg	greedy	利己的エージェント	協調的エージェント
文化系	文化	0.103	0.148	0.179	0.187	0.149
	歴史	0.107	0.110	0.121	0.128	0.125
	偉人	0.094	0.134	0.154	0.163	0.148
法務系	Law	0.099	0.131	0.145	0.172	0.159
	Law2	0.102	0.131	0.149	0.196	0.174

表 3: 4 クライアントの Non-iid データにおける，自己組織化手法ごとの翻訳精度比較

	FedAvg	greedy	利己的エージェント	協調的エージェント
文化	0.136	0.172	0.190	0.129
歴史	0.109	0.109	0.128	0.122
偉人	0.128	0.148	0.160	0.134
Law	0.117	0.145	0.146	0.112

表 4: 5 クライアントの Non-iid データにおける，自己組織化手法ごとの翻訳精度比較

	FedAvg	greedy	利己的エージェント	協調的エージェント
文化	0.139	0.170	0.187	0.128
歴史	0.106	0.113	0.121	0.121
偉人	0.120	0.145	0.154	0.121
Law	0.073	0.142	0.161	0.107
Law2	0.114	0.144	0.168	0.093

的であり，特に「歴史」カテゴリでは Stand Alone とほぼ同等の精度にとどまっている．これは，FedAvg が全クライアントの更新を一様に集約するため，各クライアントにとって必ずしも有益でない更新が混入し，性能向上効果が相殺されている可能性を示唆している．

一方，greedy 法は，iid 条件下のすべての文化系カテゴリにおいて FedAvg を明確に上回る性能を示した．具体的には，「文化」で 20.9%，「歴史」で 10.0%，「偉人」で 14.9% の精度向上が確認された．greedy 法では，各クライアントが自身のローカルデータに基づいて最も有益な協調パターンを選択するため，性能向上に寄与しないモデル更新の影響を抑制できる．このことから，FedAvg と比較して，より選択的かつ効率的な知識共有が実現されていると考えられる．

さらに，利己的エージェントは，すべてのカテゴリにおいて最も高い翻訳精

度を達成した．具体的には，「文化」で 26.4%，「歴史」で 16.4%，「偉人」で 21.6% の精度向上が確認され，特に「文化」および「偉人」カテゴリでは greedy 法を上回る顕著な改善が見られた．これは，利己的エージェントが自ドメインの検証データに基づいて報酬を設計し，自身の翻訳性能向上に最も寄与する協調関係を強化学習によって学習・選択できているためである．iid 条件下であっても，ドメイン内に存在する微細な分布差や表現傾向の違いを捉えられる点が，高い性能向上につながっていると考えられる．協調的エージェントもすべてのカテゴリで一定の性能向上を示したが，その改善幅は「文化」で 0.6%，「歴史」で 13.6%，「偉人」で 10.4% にとどまり，利己的エージェントほどの性能には達していない．しかしながら，FedAvg と同等あるいはそれ以上の精度を安定して示しており，性能の大きな劣化は観測されなかった．これは，協調的エージェントが全ドメインを均等に考慮した評価指標に基づいて行動を選択するため，特定カテゴリへの最適化は弱まる一方で，性能の偏りが抑制されるという設計思想と整合的である．

法務系ドメインに着目すると，利己的エージェントは「Law」で 6.87%，「Law2」で 27.5% の精度向上を達成しており，特に「Law2」ドメインにおいて顕著な改善が確認された．一方，協調的エージェントは「Law」で 9.9%，「Law2」で 17.6% の向上を示しており，「Law」ドメインでは協調的エージェントの方がやや高い精度を示した．この結果は，法務系ドメイン間の高い類似性により，ドメイン横断的な知識共有が有効に機能する場合があることを示唆している．

次に Non-iid 条件に着目すると，FedAvg の性能低下がより顕著に現れている．4 クライアントおよび 5 クライアントのいずれの設定においても，FedAvg は多くのカテゴリで greedy 法や利己的エージェントに大きく劣っており，特に 5 クライアント設定の「Law」では BLEU スコアが 0.117 から 0.073 へと大きく低下している．これは，クライアント数の増加に伴いデータ分布の不均一性が拡大し，一様集約による負の干渉が蓄積された結果であると考えられる．すなわち，FedAvg は Non-iid 環境においてスケーラビリティに課題を有することが示唆される．

これに対し，greedy 法および利己的エージェントは，Non-iid 環境下においても比較的安定した性能を維持している．特に利己的エージェントは，4 クライアント設定の「文化」カテゴリ（0.190）や，5 クライアント設定の「Law」「Law2」カテゴリにおいて顕著な改善を示しており，強化学習によって「自分にとって

有益な協調構造」を学習する仕組みが，分布不均一性の影響を効果的に緩和していると考えられる．

一方で，協調的エージェントは，iid 条件では安定した性能を示していたものの，Non-iid 条件では多くのカテゴリで性能が低下している．これは，全体最適や公平性を重視した報酬設計が，分布差の大きい環境では個々のクライアントにとって必ずしも有利に働かない場合があることを示している．さらに，クライアント数の増加に伴うスケーラビリティの観点では，greedy 法は全協調パターン（最大 $2^m - 1$ 通り）に対応するモデルを評価する必要があり，クライアント数の増加に伴って計算時間および通信コストが指数的に増大するという課題を有する．実験においても，クライアント数の増加に伴い greedy 法の 1 ラウンドあたりの処理時間が顕著に増大することが確認された．

これに対し，利己的エージェントおよび協調的エージェントは，強化学習によって有効な協調構造を逐次的に学習するため，全パターンを網羅的に評価する必要がなく，計算時間の増大を抑制できる．この点において，提案手法はスケーラビリティの観点でも優位性を有すると考えられる．

以上の結果から，iid 環境では選択的協調により FedAvg を上回る性能向上が可能であり，Non-iid 環境では一律集約ではなく自己組織化・選択的協調が本質的に重要であることが明らかとなった．特に，利己的エージェントによる強化学習型自己組織化は，分布不均一性に対して最も高い耐性を示し，連合ニューラル翻訳における実環境適応型手法として有効であることが示唆された．

8.2 LLM エージェント

LLM モデルをファインチューニングした後，各エポックにおけるモデルを用いてテストデータ上で推論を行い，翻訳精度の推移を評価した．その結果，図 11 に示すように，翻訳精度は学習初期から一貫して向上し，エポック 8 において最大値を示した．一方，それ以降のエポックでは精度改善が頭打ちとなり，図 12 に示される損失関数の推移からも，過学習の兆候が確認された．このことから，エポック 8 以降では訓練データへの適合が進む一方で，汎化性能の向上には寄与していない可能性が高いと判断できる．以上の結果を踏まえ，本研究ではエポック 8 のモデルを最終的な LLM エージェントとして採用し，LLM 間の連携実験を行った．

表 5 は，文化・歴史・偉人の 3 ドメインにおける翻訳精度を，ベースモデル，

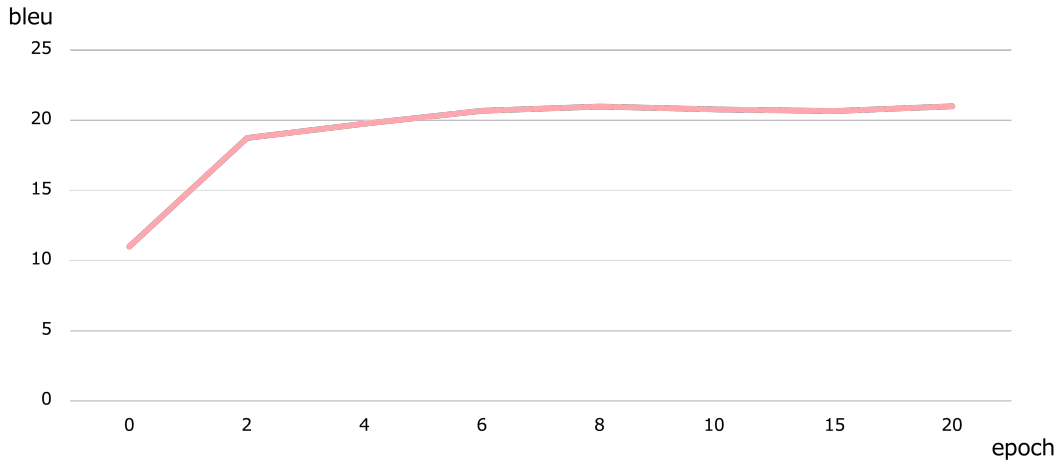

図 11: 翻訳精度の推移

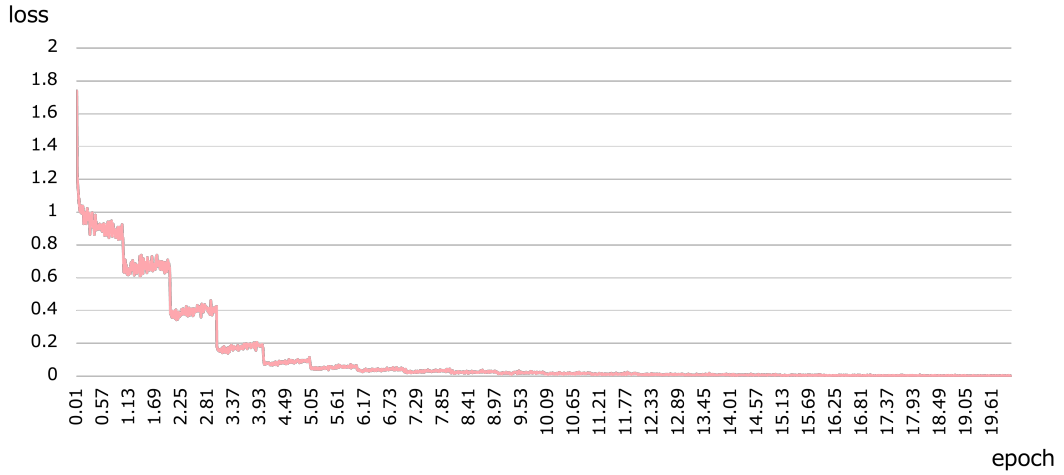

図 12: 学習過程における損失関数の推移

ファインチューニング後モデル，並列型アプローチ，直列型アプローチの 4 手法で比較した結果を示している．まず，全ドメインに共通して，ファインチューニング後のモデルはベースモデルを大きく上回る性能を示しており，ドメイン固有コーパスを用いたファインチューニングが翻訳精度向上に有効であることが確認できる．特に「歴史」ドメインでは，BLEU スコアが 0.032 から 0.123 へと大幅に向上しており，専門用語や文体の違いを学習する重要性が顕著に表れている．

次に，並列型および直列型アプローチに着目すると，いずれのドメインにおいても，ファインチューニング後モデルと同等，あるいはそれに近い性能を示

表 5: 翻訳精度比較

	ベースモデル	ファインチューニング後のモデル	並列型アプローチ	直列型アプローチ
文化	0.110	0.207	0.200	0.199
歴史	0.032	0.123	0.131	0.121
偉人	0.093	0.156	0.151	0.138

している．これは，複数の LLM を用いた協調的な翻訳生成が，単一モデルによる翻訳と比較しても一定の有効性を持つことを示唆している．「文化」ドメインでは，並列型アプローチが 0.200，直列型アプローチが 0.199，「偉人」ドメインでは，並列型アプローチが 0.151，直列型アプローチが 0.138 と，いずれもファインチューニング後モデルにはわずかに及ばないものの，高い翻訳精度を維持している．

一方，「歴史」ドメインでは，並列型アプローチが 0.131 と，ファインチューニング後モデル（0.123）を上回る結果を示している．これは，並列に生成された複数の翻訳結果を統合することで，単一モデルでは捉えきれない表現や文脈情報が補完され，翻訳品質が向上したためと考えられる．

第9章　考察

9.1　ニューラル機械翻訳の連合学習の考察

本節では，iid および Non-iid の両環境下において得られた翻訳精度の結果を基に，クライアント数およびドメイン特性が性能に与える影響について考察する．

まず，「文化」「歴史」「偉人」といった知識構造が比較的専門的でまとまりのあるドメインにおいては，クライアント数の増加に伴い翻訳精度が低下する傾向が確認された．特に，3 クライアント構成時に最も高い精度が得られており，それ以上のクライアント追加は必ずしも性能向上に寄与していない．これは，追加されたクライアントのデータが当該ドメインの言語的・概念的特徴と十分に整合せず，結果としてノイズとして作用した可能性を示唆している．

この結果は，これらのドメインにおいては，関連性の低いドメインとの広範な知識共有よりも，専門性の高い情報を集中的に活用する方が翻訳性能の向上に有効であることを示している．

一方で，「Law」および「Law2」といった法務系ドメインでは，クライアント数の増加に伴い翻訳精度が一貫して向上し，5 クライアント構成において最も高い精度を示した．これらのドメインでは，専門用語や文書構造，表現形式における共通性が高く，クライアント間での知識共有が相互に補完的に機能したと考えられる．

特に，法分野のように内容が複雑かつ多様なドメインでは，他クライアントが保持する関連知識が学習過程において有益に作用し，翻訳性能の向上につながることが示唆される．

以上の結果から，クライアント数の増加が翻訳精度に与える影響は一様ではなく，ドメイン間の類似性や知識構造の特性に強く依存することが明らかとなった．このような性能向上・低下の要因をより詳細に理解するため，各エージェントが選択した使用モデルおよび協調行動の傾向について分析を行う．

表 6〜11 には，利己的エージェントおよび協調的エージェントの協調結果を詳細に示している．ドメインは「文化」「歴史」「偉人」「Law」「Law2」をそれぞれ 0〜4 として略記し，複合モデルはこれらの組み合わせ（例：「01」は「文化」＋「歴史」）で表した．また，ドメイン間の類似度を定量的に評価するために，潜在ディリクレ配分法（Latent Dirichlet Allocation; LDA）[29] を用いた．

LDA は文書集合中の潜在トピックを確率的に推定するトピックモデルであり，各文書がどのトピックにどの程度属するかを定量的に評価できる．

本研究では，各クライアントのデータセットを 2,000 文書に分割し，図 13 に示すように分析を行った．文書番号 0〜1999 を「文化」，2000〜3999 を「歴史」，4000〜5999 を「偉人」，6000〜7999 を「Law」，8000〜9999 を「Law2」とした．LDA 図中の各色はトピック分布を示し，各色が各文書におけるトピックの出現確率を表す．表 12 には抽出された主要トピックを示す．

LDA の分析結果から，「文化」ドメインは「歴史」および「偉人」と強く関連しており，特に「偉人」との相関が高いことが確認された．同様に，「歴史」も「偉人」と高い関連性を示した．一方，「Law」と「Law2」は高い類似度を有していた．表 13 に示すように，これらのドメイン間類似度を LDA 結果から算出したところ，「文化」「歴史」「偉人」間では高い翻訳精度が得られた．これは，これらのドメイン間に強いトピック的共通性が存在するためであると考えられる．特に「文化」と「偉人」の間の高い類似性は注目に値し，両者の頻繁な協調が翻訳精度を大幅に向上させた．このことから，各エージェントは適切な協調相手を自律的に選択できていたことが示される．利己的な行動を取る場合でも，特定ドメイン内での精度向上が間接的に連携全体の性能向上につながることが確認された．

しかし，ここで「類似ドメイン同士を単純に結びつけるだけで十分か」という疑問も生じる．ドメイン類似度に関するメタデータが事前に利用可能な場合，それに基づいた組織化構築も考えられる．これを検証するため，LDA で最も類似度が高かった「文化」と「偉人」の 2 クライアントのみを用いて FedAvg を実行した．

その結果（表 14）では，提案手法よりも低い精度に留まった．つまり，単なる類似ドメイン間の協調よりも，適度に異なるドメインを含めた方が翻訳精度

表 6: 3 クライアントにおける利己的エージェントのクライアント選択結果

Aggregation Rounds	1	2	3	4	5
0: 文化	02	02	02	0	02
1: 歴史	0	1	0	2	02
2: 偉人	02	02	02	02	02

表 7: 3 クライアントにおける協調的ージェントのクライアント選択結果

Aggregation Rounds	1	2	3	4	5
0: 文化	01	01	12	01	01
1: 歴史	01	12	012	01	01
2: 偉人	01	02	01	01	2

表 8: 4 クライアントにおける利己的エージェントのクライアント選択結果

Aggregation Rounds	1	2	3	4	5
0: 文化	02	02	02	0123	013
1: 歴史	02	02	02	02	02
2: 偉人	02	02	02	0123	02
3:Law	02	02	02	23	3

表 9: 4 クライアントにおける協調的エージェントのクライアント選択結果

Aggregation Rounds	1	2	3	4	5
0: 文化	1	012	012	012	012
1: 歴史	012	012	023	1	012
2: 偉人	2	2	012	012	02
3:Law	012	012	012	023	13

の改善につながることが示された．これは提案手法の有効性を強く支持する結果である．

また，単純な類似度指標に基づく手法よりも，提案手法はより効果的な学習を実現できることが確認された．これは，組織化過程に強化学習を導入する意義を示しており，各エージェントが利己的に行動しつつも，全体として競合せず協調的に性能を高められることを意味している．

一般に，協調作業においてはメンバー間の協力行動が全体性能を高めると考えられる．しかし本研究では，利己的な行動を取るエージェントの方が全体性能を向上させるという興味深い結果が得られた．一方で，「文化」「歴史」「偉人」の 3 クライアントと，「Law」「Law2」の 2 クライアントとの間にはドメイ

表 10: 5 クライアントにおける利己的エージェントのクライアント選択結果

Aggregation Rounds	1	2	3	4	5
0: 文化	1234	013	124	1	134
1: 歴史	1	1234	01234	1234	1234
2: 偉人	1234	1234	1234	1234	1234
3:Law	1234	024	1234	1234	013
4:Law2	1234	1234	1234	1234	1234

表 11: 5 クライアントにおける協調的エージェントのクライアント選択結果

Aggregation Rounds	1	2	3	4	5
0: 文化	02	0124	02	3	02
1: 歴史	034	02	02	02	02
2: 偉人	02	02	02	0234	02
3:Law	024	02	124	2	0134
4:Law2	02	02	02	02	02

ン類似度が低く，両者を組み合わせると前者の精度は低下した．しかし，逆に後者は補完的な知識を得て精度が向上した．このため，5 クライアント全体で実験を行った場合に，総合精度が最も高くなるケースも見られた．これらの結果は，ドメイン間の知識補完性が特定条件下で翻訳精度を高める可能性を示唆している．

さらに，利己的エージェント間では類似ドメイン同士の強い協調が観察されたが，事前に類似度が把握できていたとしても，FedAvg のような単純な集約手法では十分な性能向上を得ることはできなかった．本研究では，強化学習の導入により，各エージェントが現在の状態に応じて動的に最適なモデルを選択できるようになり，高い翻訳精度と適応性を実現した．これは，単なるドメイン類似度の利用を超えた，より柔軟で動的な学習の有効性を示している．

利己的エージェントと協調的エージェントの行動を比較した結果，両者の協調戦略が学習結果にどのような影響を与えるかも明らかになった．利己的エージェントは常に自身の性能向上を最優先し，協調相手を選択した．表 6〜10 に

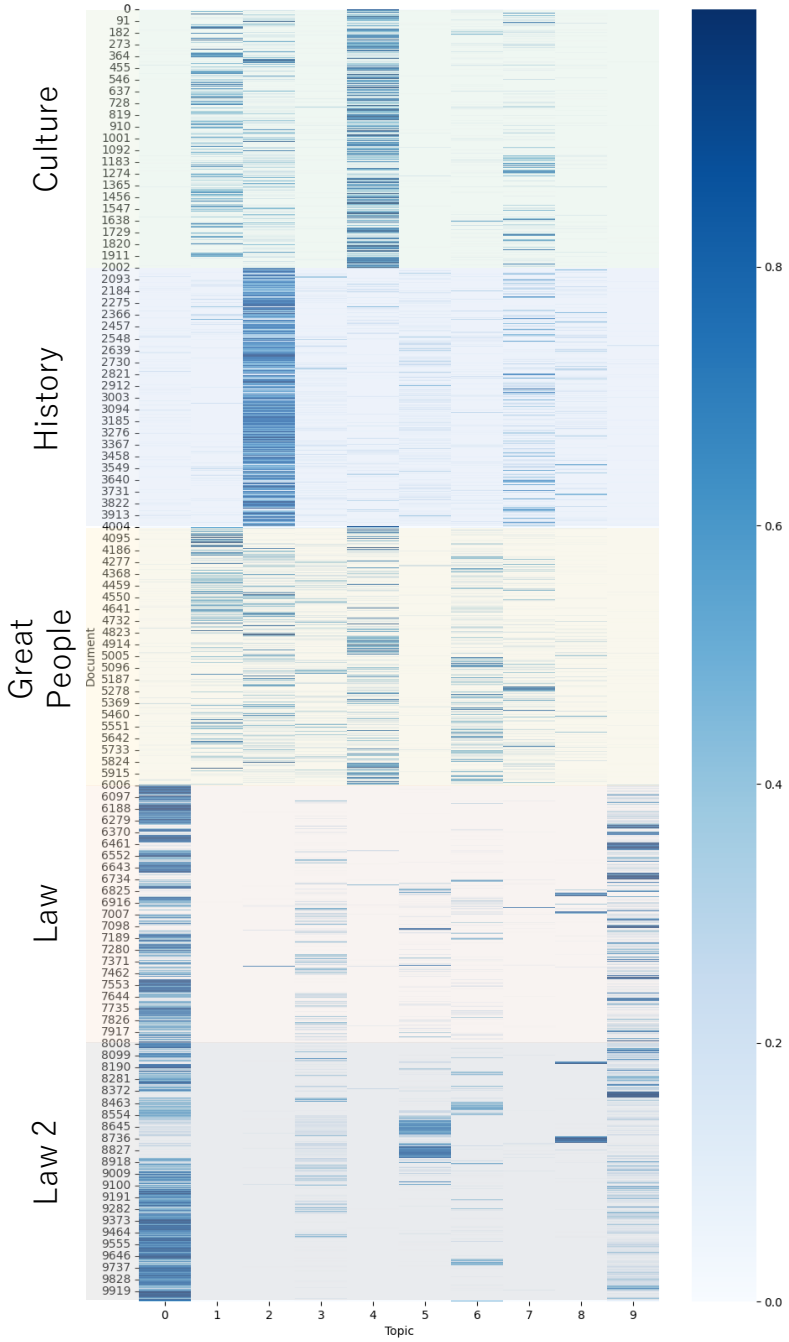

図 13: LDA 分析の結果

示すように，これらのエージェントは類似または同一ドメインのクライアントと繰り返し協調し，短期的な精度向上を達成した．一方，異質なドメインとの協調は限定的であり，長期的には知識の多様性が制限される可能性がある．

3 または 4 クライアント構成の場合，利己的エージェントは類似度の高いパー

表 12: LDA 分析により抽出された主要トピック

トピック	内容
0	法的文書の構成要素
1	日本の歴史的貴族階級
2	日本の文化と伝統
3	地域開発とインフラ
4	侍と戦国時代
5	労働安全と化学物質
6	裁判所と司法制度
7	教育と文化
8	地理と地域
9	企業と法的規制

表 13: ドメイン類似度 (LDA の結果)

	文化	歴史	偉人	Law	Law2
文化	-	0.27	0.82	0.01	0.01
歴史		-	0.51	0.03	0.04
偉人			-	0.08	0.09
Law				-	0.97
Law2					-

トナーを選ぶ傾向を示し，少数派ドメインのクライアントは，高性能かつ類似したモデルを戦略的に選択することで自身の翻訳精度を向上させた．また，自身のモデルを含む構成を選好する傾向も確認され，自己強化的な学習行動が見られた．

一方，協調的エージェントは，自身だけでなく全ドメインに有益となる協調相手を選択した．3 および 4 クライアント構成では，協調的エージェントは複数ドメインにまたがる幅広い協調関係を形成した．しかし，4 クライアント構成においては「3 対 1」の構造的不均衡が生じ，3 つのドメインに偏った選択が見られた．このため，残り 1 つのドメインの影響が限定され，全体への貢献が

表 14: 類似度が高かったドメインによる FedAvg

	Stand Alone	FedAvg	利己的 エージェント	協調的 エージェント
文化	0.103	0.164	0.187	0.149
偉人	0.094	0.141	0.163	0.148

制約されたと考えられる．また，協調的エージェントは他クライアントの学習状況も考慮する傾向があり，より多様な協調パターンを形成した．

これらの結果から，協調相手の選択には専門化と汎化のトレードオフが存在することが分かる．利己的エージェントは同質な知識空間内での密な協調により性能を最適化し，協調的エージェントは異質な情報源との交流を通じて汎化性能を高める．したがって，両戦略を適切に組み合わせることで，連合学習全体の翻訳性能をさらに向上させることが可能である．

さらに，パートナー選択の結果から，エージェントは無作為に協調相手を選んでいるのではなく，学習過程を通じて有益な関係を自律的に形成していることが示唆された．これは，提案手法における強化学習の導入が，事前定義された類似度指標に依存せず，動的かつ自己組織的な協調構造を実現していることを意味する．このような行動は，多様かつ分散した環境における効率的で適応的な協調学習の可能性を示している．

一方で，クライアント数の増加はエージェントの行動空間を拡大させ，最適行動の選択を困難にする可能性がある．これにより，一部クライアントの学習不足や精度低下を引き起こすことも考えられる．今後は，行動空間の次元削減や効率的な探索戦略の導入により学習効率を改善し，クライアント選択手法をさらに最適化することで，より安定した高精度の翻訳を実現できると期待される．

9.2 LLM エージェントの考察

並列型と直列型を比較すると，全体として並列型アプローチの方が安定して高い翻訳精度を示している．並列型アプローチでは，各モデルが独立に翻訳を行うため，多様な表現候補が生成され，それらを統合する過程で誤りの相殺や表現の補完が生じやすい点が特徴である．

並列型アプローチの利点は，生成される複数の翻訳候補が有効に機能する点

表 15: 歴史データの翻訳精度比較

	ファインチューニング後のモデル	並列型アプローチ
文化	0.128	
歴史	0.123	0.131
偉人	0.123	

にある．実際に，歴史のドメインのドメイン特化型データの各ドメインモデルを単独で用いた場合の BLEU スコアを測ってみると (表 15)，文化のモデルが 0.128，歴史が 0.123，偉人が 0.123 に留まっているのに対し，並列型アプローチでは 0.131 と最も高い精度を示している．この結果は，単一のドメインモデルが全ての入力文に対して常に最適な翻訳を生成できているわけではないことを示している．すなわち，特定ドメインのモデルでは不十分な翻訳となる場合であっても，他ドメインのモデルが生成した翻訳文の中に，より適切な表現や文脈解釈が含まれている可能性がある．並列型アプローチにおける多数決や選択機構は，こうした複数モデルの生成文の中から，相対的に品質の高い翻訳を選別できていると考えられる．また，並列型アプローチの性能向上は，各エージェントの翻訳能力が均質であることを前提としていない点も重要である．むしろ，各モデルが異なる強みと弱みを持つことで翻訳候補の多様性が確保され，結果として最終的な翻訳品質の向上につながっている．このことは，並列型アプローチが「最良のモデル」を選択するのではなく，「最良の翻訳文」を選択する枠組みとして機能していることを示している．

以上の考察から，並列型アプローチは，複数の LLM エージェントが生成した翻訳文の中から適切な翻訳を選択する能力に優れており，単一ドメインモデルでは達成しにくい安定した性能向上を実現していることが明らかとなった．ただし，本方式における選択機構は固定的なルールに基づいており，入力文やドメイン特性に応じて動的に最適化されるわけではないという制約も存在する．

一方，直列型アプローチでは，翻訳過程が逐次的であるため，初期段階で生じた誤訳や情報欠落が後続モデルに伝播しやすく，結果として性能の上限が制約される傾向が見られた．実際に，直列型アプローチにおける全てのエージェント組み合わせの翻訳精度を比較した結果，組み合わせによって精度にばらつき

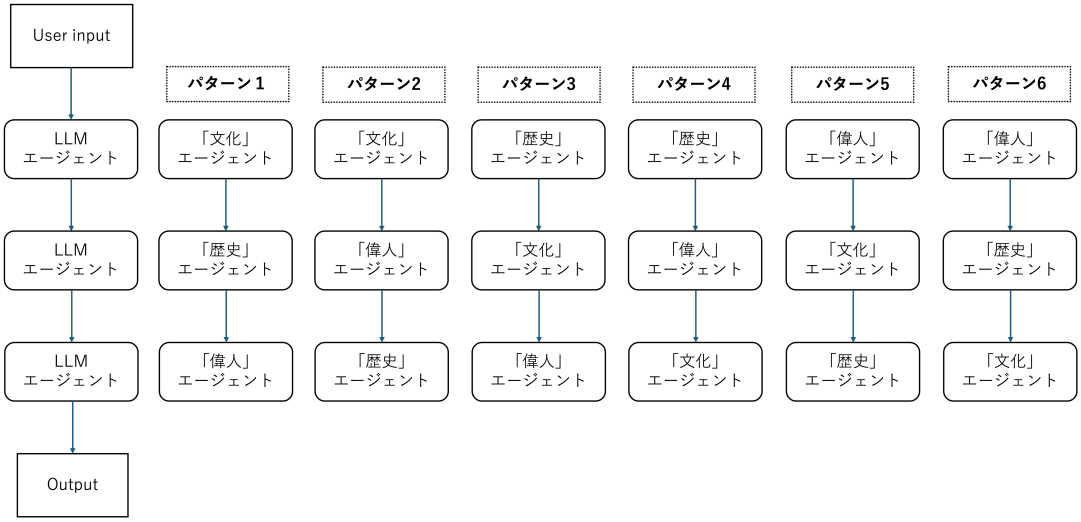

図 14: 直列型アプローチのすべての組み合わせ

が生じることが確認された (図 14，表 16)．特に，各ドメインに対応する自身のモデルが最初に配置された場合に，最終的な翻訳精度が最も高くなる傾向が見られた．この結果は，直列型アプローチにおいて，最初に生成される翻訳文が後続エージェントにとって強い参照情報として機能していることを示唆している．後続の LLM エージェントは，入力文そのものだけでなく，前段エージェントの出力を参照文として利用しながら翻訳を行うため，その内容に大きく影響を受けると考えられる．そのため，初段の翻訳品質が高い場合には，後続エージェントによる修正や洗練が効果的に働き，高い翻訳精度が得られる．一方で，初段に配置されたエージェントの翻訳品質が低い場合や，ドメイン適合性の低いモデルが担当した場合には，不適切な表現や誤訳が参照文として固定され，後続エージェントがそれを十分に修正できないまま最終出力に反映されるケースも確認された．これは，直列型アプローチにおいて誤り伝播の影響が顕在化しやすいことを示している．

表 16: 直列型アプローチのすべての組み合わせの結果

	パターン 1	パターン 2	パターン 3	パターン 4	パターン 5	パターン 6
文化	**0.199**	0.194	0.153	0.167	0.051	0.149
歴史	0.120	0.113	0.112	**0.121**	0.057	0.114
偉人	0.131	0.131	0.131	0.129	**0.138**	0.137

以上の結果から，直列型アプローチでは「どのエージェントを最初に配置するか」が翻訳精度に大きな影響を与える重要な要因であることが明らかとなった．自身のドメインに最も適合したモデルを初段に配置することで高品質な参照文を生成できる一方で，固定的な順序に基づく直列型協調では，入力文や状況に応じた柔軟な役割分担が困難であり，初期選択の誤りが全体性能を制約するという課題が残る．

第10章　おわりに

本研究では連合学習において，集約プロセスのたびに動的に連携相手を選択し，翻訳モデルを統合する協調的なエージェントを提案する．各データ所有者をエージェントとし，エージェントは深層強化学習を通じて連携相手を評価し，最適なモデルの統合先を選択する方策を学習する．提案手法によって構築されたニューラル機械翻訳モデルを用いて，評価データを翻訳し，BLEU スコアを用いてその精度の評価を行った．本研究の貢献は以下の通りである．

マルチエージェント強化学習に基づく自己組織化の設計

マルチエージェント深層強化学習の導入によって，連合学習の参加者が将来的のモデルの精度上昇を考慮しながら，最適な連携相手を動的に選択することで，従来手法によって構築されたニューラル機械翻訳モデルと比較して提案手法によって構築されたモデルの精度は平均で 21.5%高くなり，その有効性を検証した．

エージェントの内部モデル

その結果，利己的エージェントは，自身のドメインにおける翻訳精度を最大化することを目的として，短期的に最も大きな性能向上が見込める協調相手を優先的に選択する傾向を示した．このため，類似ドメイン間での強い協調が形成されやすく，高いドメイン特化性能を達成できる一方で，異質なドメインとの協調は抑制される傾向にある．一方，協調的エージェントは，全ドメインにわたる平均的な翻訳性能の向上を目標として行動を選択するため，特定ドメインへの最適化よりも，全体のバランスを重視した協調関係を構築する傾向を示した．その結果，単一ドメインにおける性能は利己的エージェントに及ばない場合があるものの，性能の極端な偏りが抑制され，より安定した汎化性能を示すことが確認された．このように，内部モデルにおける評価基準の違いが，エージェントの協調戦略および最終的な翻訳性能に直接的な影響を与えており，本研究の枠組みが「ドメイン特化性能」と「全体最適化」という異なる目的を柔軟に切り替え可能であることを示している．

提案手法におけるマルチエージェント深層強化学習は，従来の FedAvg 手法に比べて，ニューラル機械翻訳モデルの専門性を高め，その精度を大幅に向上させることができる．さらに，本研究では，各クライアントにおいて事前に LLM モデルをドメイン固有データでファインチューニングすることで，初期段階から

一定の翻訳能力とドメイン適応性を備えたエージェントを構築した．このファインチューニングにより，各エージェントは自身の専門分野における表現や語彙を適切に獲得しており，その後の連携アプローチにおいて，より有益な知識交換が可能となっている．

謝辞

本研究を行うにあたり，熱心なご指導，ご助言を賜りました村上陽平教授に深く感謝申し上げます．また，普段からお世話になっている社会知能研究室の皆様にも心より感謝申し上げます．

参考文献

[1] Bahdanau, D., Cho, K. and Bengio, Y.: Neural Machine Translation by Jointly Learning to Align and Translate, *CoRR*, Vol. abs/1409.0473 (2014).

[2] Tu, Z., Lu, Z., Liu, Y., Liu, X. and Li, H.: Modeling Coverage for Neural Machine Translation, *arXiv: Computation and Language* (2016).

[3] Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J. R., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G. S., Hughes, M. and Dean, J.: Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, *ArXiv*, Vol. abs/1609.08144 (2016).

[4] McMahan, H. B., Moore, E., Ramage, D., Hampson, S. and y Arcas, B. A.: Communication-Efficient Learning of Deep Networks from Decentralized Data, *International Conference on Artificial Intelligence and Statistics* (2016).

[5] Brown, P. F., Pietra, V. J. D., Pietra, S. A. D. and Mercer, R. L.: The mathematics of statistical machine translation: parameter estimation, *Comput. Linguist.*, Vol. 19, No. 2, p. 263 – 311 (1993).

[6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I.: Attention is all you need, *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, Red Hook, NY, USA, Curran Associates Inc., p. 6000 – 6010 (2017).

[7] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. and Amodei, D.: Language models are few-shot learners, *Proceedings of the*

34th International Conference on Neural Information Processing Systems, NIPS '20, Red Hook, NY, USA, Curran Associates Inc. (2020).

[8] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (Burstein, J., Doran, C. and Solorio, T.(eds.)), Minneapolis, Minnesota, Association for Computational Linguistics, pp. 4171–4186 (2019).

[9] Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D. and Chandra, V.: Federated Learning with Non-IID Data, *arXiv preprint arXiv:1806.00582* (2018). Submitted on 2 Jun 2018, last revised 21 Jul 2022.

[10] Arivazhagan, M. G., Aggarwal, V., Singh, A. K. and Choudhary, S.: Federated Learning with Personalization Layers, *ArXiv*, Vol. abs/1912.00818 (2019).

[11] Wang, K., Mathews, R., Kiddon, C., Eichner, H., Beaufays, F. and Ramage, D.: Federated Evaluation of On-device Personalization, *ArXiv*, Vol. abs/1910.10252 (2019).

[12] Lu, Z., Pan, H., Dai, Y., Si, X. and Zhang, Y.: Federated Learning With Non-IID Data: A Survey, *IEEE Internet of Things Journal*, Vol. 11, No. 11, pp. 19188–19209 (2024).

[13] Wang, H., Kaplan, Z., Niu, D. and Li, B.: Optimizing Federated Learning on Non-IID Data with Reinforcement Learning, *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, pp. 1698–1707 (2020).

[14] Sutton, R. S. and Barto, A. G.: *Reinforcement Learning: An Introduction*, The MIT Press, second edition (2018).

[15] Watkins, C. and Dayan, P.: Q-learning, *Machine Learning*, Vol. 8, pp. 279–292 (1992).

[16] Watkins, C.: Learning from delayed rewards (1989).

[17] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M. A., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran,

D., Wierstra, D., Legg, S. and Hassabis, D.: Human-level control through deep reinforcement learning, *Nature*, Vol. 518, pp. 529–533 (2015).

[18] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. and Riedmiller, M. A.: Playing Atari with Deep Reinforcement Learning, *ArXiv*, Vol. abs/1312.5602 (2013).

[19] Hasselt, H. V., Guez, A. and Silver, D.: Deep Reinforcement Learning with Double Q-Learning, *ArXiv*, Vol. abs/1509.06461 (2015).

[20] Zhang, S. Q., Lin, J. and Zhang, Q.: A Multi-agent Reinforcement Learning Approach for Efficient Client Selection in Federated Learning, *AAAI Conference on Artificial Intelligence* (2022).

[21] Wang, Y., Zhou, Y. and Huang, P.-Q.: A Novel Incentive Mechanism for Federated Learning Over Wireless Communications, *IEEE Transactions on Artificial Intelligence*, Vol. 5, No. 11, pp. 5561–5574 (2024).

[22] Klein, G., Kim, Y., Deng, Y., Senellart, J. and Rush, A.: OpenNMT: Open-Source Toolkit for Neural Machine Translation, *Proceedings of ACL 2017, System Demonstrations* (Bansal, M. and Ji, H.(eds.)), Vancouver, Canada, Association for Computational Linguistics, pp. 67–72 (2017).

[23] Yuan, T., Chung, H.-M. and Fu, X.: PP-MARL: Efficient Privacy-Preserving Multi-Agent Reinforcement Learning for Cooperative Intelligence in Communications, *IEEE Network*, Vol. 38, No. 5, pp. 196–203 (2024).

[24] Hebert, L., Golab, L., Poupart, P. and Cohen, R.: FedFormer: Contextual Federation With Attention in Reinforcement Learning (2023).

[25] Jin, H., Peng, Y., Yang, W., Wang, S. and Zhang, Z.: Federated Reinforcement Learning with Environment Heterogeneity, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics* (Camps-Valls, G., Ruiz, F. J. R. and Valera, I.(eds.)), Proceedings of Machine Learning Research, Vol. 151, PMLR, pp. 18–37 (2022).

[26] Reiter, E.: A Structured Review of the Validity of BLEU, *Computational Linguistics*, Vol. 44, No. 3, pp. 393–401 (2018).

[27] Singhal, K., Azizi, S., Tu, T. et al.: Large language models encode clinical knowledge, *Nature*, Vol. 620, pp. 172–180 (2023).

[28] Cheng, D., Gu, Y., Huang, S., Bi, J., Huang, M. and Wei, F.: Instruction Pre-Training: Language Models are Supervised Multitask Learners, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (Al-Onaizan, Y., Bansal, M. and Chen, Y.-N.(eds.)), Miami, Florida, USA, Association for Computational Linguistics, pp. 2529–2550 (2024).

[29] Blei, D. M., Ng, A. Y. and Jordan, M. I.: Latent dirichlet allocation, *J. Mach. Learn. Res.*, Vol. 3, No. null, pp. 993–1022 (2003).