卒業論文

マルチエージェント深層強化学習を用いたニューラル機械翻訳の連携

指導教官　　村上 陽平 教授

立命館大学　情報理工学部
先端社会デザインコース　4 回生
2600200131-3

北川 勘太朗

2023 年度（秋学期）卒業研究 3（CH）
令和 6 年 1 月 31 日

マルチエージェント深層強化学習を用いた
ニューラル機械翻訳の連携

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であるが，高品質な対訳データ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，翻訳モデルのみを統合でき，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，連合学習において非独立同一分布（Non-independent and identically distributed (Non-iid)）と呼び，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，各エージェントの選好に応じたモデルを集約する．また，エージェントの選考を反映させる手法として深層強化学習を用いる．各エージェントは深層強化学習によって方策を獲得し，その方策によって最適な連携相手を選択する．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．

訓練データの削減によるアクションの効率化

深層強化学習を用いて最善な連携相手の選択方策を獲得するには，エージェントは何度も試行を繰り返さなければならない．試行の度に，ニューラル機械翻訳の学習を行うため，方策獲得の所要時間を短縮するには，ニューラル機械翻訳の学習時間を短縮する必要がある．そこで，ニューラル機械翻訳の訓練回数や訓練データのサイズを削減しても，削減前と同様の方策を学習できているかを検証する．

状態空間の削減

ニューラル機械翻訳モデルの状態は，本来，獲得した重みパラメータであるが，パラメータ数が多く，状態を表す次元数も大規模化するため，深層強化

学習の状態に適さない．そこで，検証データの評価値群を用いて擬似状態を定義する．この次元数は検証データのサイズに依存するため，検証データのサイズを小さくし，状態空間を削減することで，各エージェントは最適な方策を獲得しやすくなる．一方で，状態空間の削減は実際の状態と疑似状態の乖離に繋がるため，状態空間の削減と精度の関係を検証する必要がある．

前者の課題に対しては，深層強化学習を進める中で，その中のニューラル機械翻訳の学習環境を変化させる．具体的には，エポック数を変えずに，訓練データのサイズと学習回数を削減した上で訓練する．削減前後の方策の同一性を新たな連合学習の環境下で検証する．検証方法としては，得られた方策を基に新たな連合学習の環境において，同じ条件で検証を試みて評価する．

後者の課題に対しては，前者の課題と異なり，いかに正確にモデルの状態を捉えるかに着目する．深層強化学習を進める中で，検証データのサイズを変化させる．検証データのサイズに応じて方策がどのように変化するかを新たな連合学習の環境下で検証する．検証方法としては，得られた方策を基に新たな連合学習の環境において，同じ条件で検証を試みて評価する．

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い，従来手法である FedAvg で構築されたニューラル機械翻訳モデルと比較することで，提案手法の有効性を検証した．本研究の貢献は以下の通りである．

訓練データの削減によるアクションの効率化

マルチエージェント強化学習を用いた自己的に組織化する手法は，従来手法である FedAvg より高精度のニューラル機械翻訳モデルを構築できることで提案手法の有効性を示した．また，ニューラル機械翻訳の訓練データの削減により，強化学習のタスクが 39.7%削減した．

状態空間の削減

マルチエージェント強化学習において状態を表す次元数を減らすことで各エージェントが最適な方策を獲得しやすくなった結果精度が 25.3%向上した．

Cooperative Neural Machine Translation with 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 distribution of the data owned by each data owner is different, 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.

Improving the efficiency of actions by reducing training data

Since neural machine translation is trained for each trial, it is necessary to reduce the training time for neural machine translation to shorten the time required for policy acquisition. Therefore, we have to verify whether the number of neural machine translation training runs, and the size of the training data can be reduced while still learning the same policies as before the reduction.

Reduction of State Space

The states of the neural machine translation model are essentially the acquired weight parameters, but the large number of parameters and the large number of states make them unsuitable for deep reinforcement learning. Since the number of states depends on the size of the validation data, reducing the size of the validation data and the state space makes it easier for

each agent to acquire the optimal strategy. On the other hand, since the reduction of the state space leads to a discrepancy between the actual state and the pseudo-state, it is necessary to verify the relationship between the state space reduction and accuracy.

To address the former issue, we changed the learning environment for neural machine translation in the process of deep reinforcement learning. Specifically, the number of epochs was not changed, but the size of the training data and the number of training cycles were reduced before training. The identity of the measures before and after the reduction was verified in the new environment of coalition learning.

For the latter task, unlike the former, we focused on how to accurately capture the state of the model. We varied the size of the validation data in the process of deep reinforcement learning. We examined how the strategy changes with the size of the validation data in a new environment of coalition learning.

Using the neural machine translation model constructed by the proposed method, we translated the evaluation data, evaluated its accuracy using BLEU scores, and compared it with the neural machine translation model constructed by FedAvg, a conventional method, to verify the effectiveness of the proposed method. The contributions of this study are as follows.

Improving the efficiency of actions by reducing training data

The effectiveness of the proposed method was demonstrated by the fact that the self-organizing method using multi-agent reinforcement learning can construct a neural machine translation model with higher accuracy than FedAvg, a conventional method. In addition, the reduction of training data for neural machine translation reduced the reinforcement learning task by 39.7%.

Reduction of State Space

In multi-agent reinforcement learning, reducing the number of states made it easier for each agent to obtain the optimal strategy, resulting in a 25.3% improvement in accuracy.

マルチエージェント深層強化学習を用いた
ニューラル機械翻訳の連携

目次

第1章 はじめに

ニューラル機械翻訳（Neural Machine Translation，NMT）[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，連合学習[1]を用いて，各データ所有者の下で学習したモデルを連携させる. そのようにすることで, ユーザ間での対訳データの共有を必要としなくなり，先ほどの問題で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)）と呼び，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，各エージェントの選好に応じたモデルを集約する．また，エージェントの選考を反映させる手法として深層強化学習を用いる．各エー

ジェントは深層強化学習によって方策を獲得し，その方策によって最適な連携相手を選択する．

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．

訓練データの削減によるアクションの効率化

深層強化学習を用いて最善な連携相手の選択方策を獲得するには，エージェントは何度も試行を繰り返さなければならない．試行の度に，ニューラル機械翻訳の学習を行うため，方策を獲得するのに要する時間を短縮するには，ニューラル機械翻訳の学習時間を短縮する必要がある．そこで，ニューラル機械翻訳の訓練回数や訓練データのサイズを削減しても，削減前と同様の方策を学習できているかを検証する必要がある．

状態空間の削減

ニューラル機械翻訳モデルの状態は，本来，獲得した重みパラメータであるが，パラメータ数が多く，状態を表す次元数も大規模化するため，深層強化学習の状態に適さない．そこで，テストデータの評価値群を用いて擬似状態を定義する．この次元数はテストデータのサイズに依存するため，テストデータのサイズを小さくし，状態空間を削減することで，各エージェントは最適な方策を獲得しやすくなる．一方で，状態空間の削減は実際の状態と疑似状態の乖離に繋がるため，状態空間の削減と精度の関係を検証する必要がある．

以下，本論文では，第２章においてニューラル機械翻訳と連合学習について紹介する．次に，第３章ではマルチエージェント深層強化学習による組織化手法について具体的に説明する．続いて，第４章は連合学習の効率化について２つの提案を行い，第５章で提案した手法に対する評価を行う．そして，第５章ではその評価結果について考察し，最後に第６章で本稿をまとめる．

第2章 ニューラル機械翻訳の連合学習

2.1 ニューラル機械翻訳

ニューラル機械翻訳は，言語間の翻訳を行うためのアルゴリズムである．ニューラルネットワークを使用しており，これは脳の神経細胞の情報伝達を模した計算モデルである．ニューラル機械翻訳は，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）に比べて高い翻訳精度を示している．ニューラル機械翻訳は，エンコーダー（encoder），アテンション機構（attention mechanism），デコーダー（decoder）の 3 つの構成要素から成り立っている．エンコーダーは翻訳される文書を入力文として順次に読み込み，実数ベクトルに符号化する．そして，アテンション機構は注目すべき部分を決定し，最後のデコーダーは出力文を生成する．ニューラル機械翻訳は，学習時に対訳コーパスと呼ばれる翻訳文のペアのデータを使用してモデルを学習する．その対訳コーパスが多ければ多いほど構築されたニューラル機械翻訳モデルの翻訳精度が向上する傾向があるため，ニューラル機械翻訳の品質が学習時に使う対訳データ量に強く依存している．しかしながら，現在では著作権やプライバシー，セキュリティ等の問題で，一つの組織で十分な量の対訳データを収集することは困難である．連合学習は，複数の組織がデータを共有せずにモデルを共同で学習する方法である．これにより，高品質な対訳コーパスの不足という問題を解決することが期待されている．

2.2 連合学習

連合学習（Federated Learning）は，2016 年に提唱された分散型の機械学習手法である．連合学習では，参加する各組織が所有する学習データを共有せずに，機械学習モデルの学習に必要な情報だけをローカルで計算し，共有する．つまり，各組織は自身のデータを保持しながら，学習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．一般的に連合学習は複数のクライアントと一つのサーバから構成され，図 1 のような流れで学習する．連合学習の利点は，データを一箇所に集めることなく，各組織の学習データの特徴を反映した機械学習モデルを生成できる点にある．各組織はローカルで学習を行い，学習済みのモデルのパラメータを集約することで，グローバルなモデルを構築する．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，連合学習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．

Federated Average(FedAvg)は，図 2 のような連合学習の代表的なアルゴリズムである[4]．以下の動作プロセスのように，サーバはランダムなクライアントを選択し，グローバルモデルを送り，学習させる．そして，クライアントからもらった更新用パラメータを各クライアントが持つサンプル数の重み付け平均によって統合している[5]．

1. パラメータサーバは値がランダムに設定されている初期モデル(グローバルモデル)を生成し，各クライアントへ配布する．
2. 各クライアントは，受信した初期モデルに対して，自身が保有する学習データを用いてローカルで学習を行い，モデルを更新する．
3. 各クライアントは，更新されたモデルからモデルパラメータを抽出し，更新用データとしてサーバへ送信する．
4. サーバは，クライアントからもらった更新用データを集約し，グローバルモデルのモデルパラメータを更新する．
5. サーバは，更新されたグローバルモデルを各クライアントへ配布し，2 の処理に戻る．

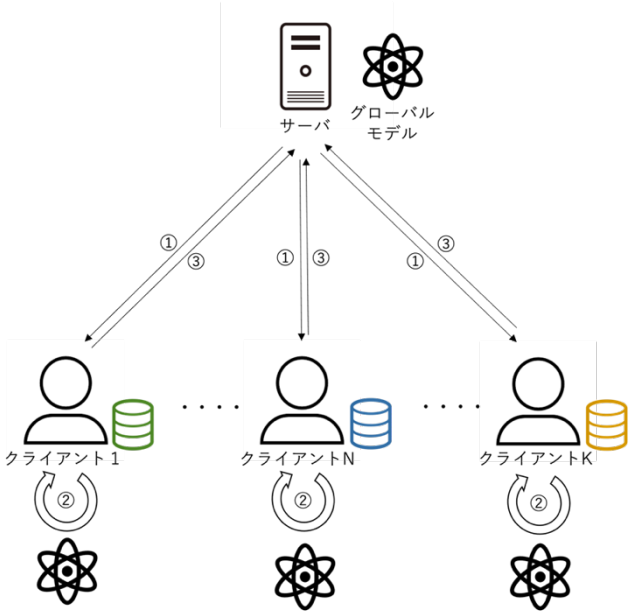

図 1:連合学習の流れ

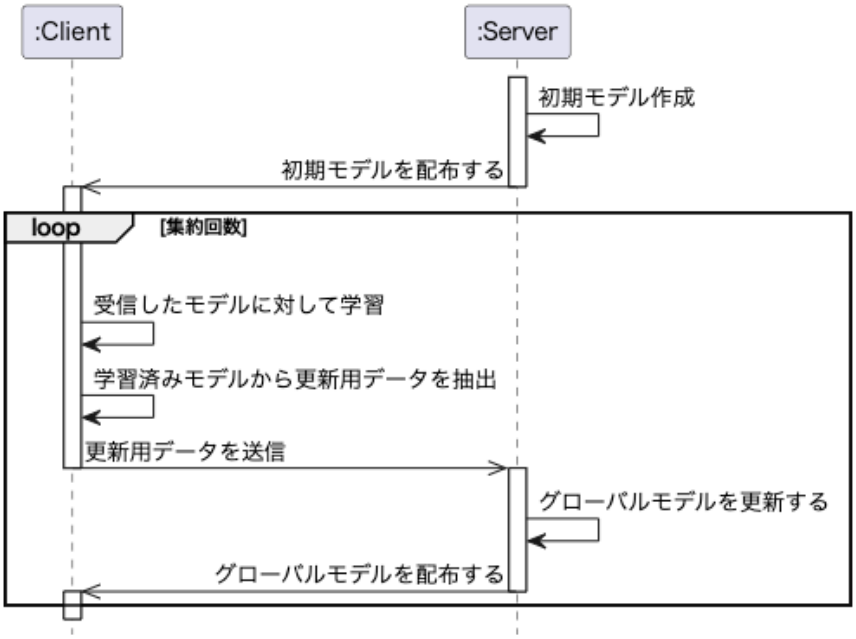

図 2:FedAvg のシーケンス図

第3章 マルチエージェント深層強化学習を用いた組織化

3.1 深層強化学習

強化学習[7]は，エージェントが環境の中で行う試行錯誤を通じて得られる報酬を頼りに，環境に適応するエージェントの方策を獲得する機械学習の一種である．心理学の中に刺激と応答の間に与えられる刺激（強化子）を元に特定の行動パターンの発見が増強されることを強化と呼ぶ．強化学習は，教師データを事前に与えられず，かつ観測可能な情報が行動によって変化する状況で，最適な行動の系列を見つけることを目指す．強化学習の一般的なケースは，自律的に動く主体が周囲に影響を与える場合であり，図 3 のように行動する主体をエージェント，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動することを行動と呼び，エージェントの行動によって変化する環境の要素を状態と呼ぶ．エージェントがどの行動を選択するかによってその後の状態が変わる．また，強化学習では，未知の環境でのエージェントの行動の良し悪しを評価する指標として報酬と呼ばれるスカラー値が利用されている．つまり，強化学習は置かれた環境の中で行動の選択を通して得られる報酬の総和を最大化することが目標である．

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[8][9][10]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

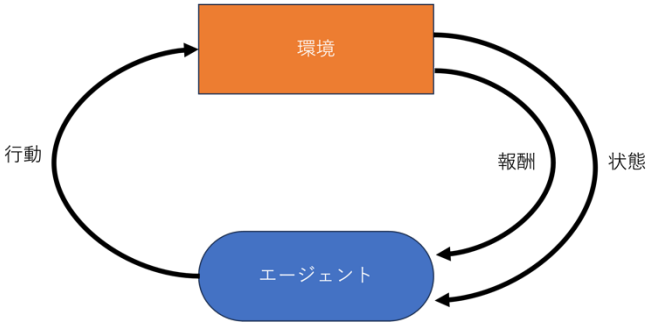

図 3:深層強化学習の様子

ばれる行動価値関数を学習し，制御を実現する．Q 関数は，時刻 t において状態 s のときに行動 a を行った場合に，その先でどれくらいの報酬 r が期待されるかを出力する関数である．Q 学習では，式(1)のようにエージェントが一つの行動を取った段階で得た報酬のもとに現状態のもとに取った行動の価値 Q を見積もり，Q テーブル$Q(s,a)$を更新する．

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha\left(r_{t+1} + \gamma \max_{a_{t+1}\in A} Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)\right) \quad (1)$$

ここで，α は学習率$(0 \le \alpha \le 1)$であり，Q の更新速度を制御するためのハイパーパラメータである．また，γ は期待報酬を見積もる際の不確かさを表現する割引率$(0 \le \gamma \le 1)$と呼ばれる係数であり，時間とともに価値を割り引いていくことで，Qテーブルが収束するように制御する．

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，強化学習における行動価値関数 Q 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が注目されている[6]．深層強化学習は，Mnih ら[6][11]によって提案された Q 学習の拡張である Deep Q-Network (DQN)などの手法があり，これらは主に画像認識で高い識別能力を発揮する[12][13][14]．式(1)の Q テーブルの更新において，右辺の第二項(TD 誤差)が 0 に近づくことで Q テーブルが完成する．そこで，式(2)に示すように損失関数を定義し，確率的勾配法などで最小化（Q テーブルの更新）を多層ニューラルネットワークで行う．

$$Loss = \left(r_{t+1} + \gamma \max_{a_{t+1}\in A} Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)\right)^2 \quad (2)$$

3.2 マルチエージェント深層強化学習

実際の世界では，ある環境内に複数のエージェントが同時に存在することは一般的である．図 4 のように同じ環境で複数の強化学習エージェントが同時に学習，行動し，お互いに影響し合う自律分散型システムをマルチエージェント強化学習(Multiagent reinforcement learning ; MARL)と呼ぶ．

マルチエージェント強化学習には，エージェントの利害関係に基づいて以下の種類がある．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．

また，学習方法によって三つに分類できる．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，不安定になる傾向がある．

3.3 ニューラル機械翻訳の組織化

マルチエージェント深層強化学習を用いることで，将来的なモデル精度の期待値を考慮し，最善な組織法を見つけ出すことができる．これを実現するために，連合学習システムをモデル化し，マルチエージェント深層強化学習システムに取り込む必要がある．本節でマルチエージェント深層強化学習を組み込んだ連合学習システムを説明する．

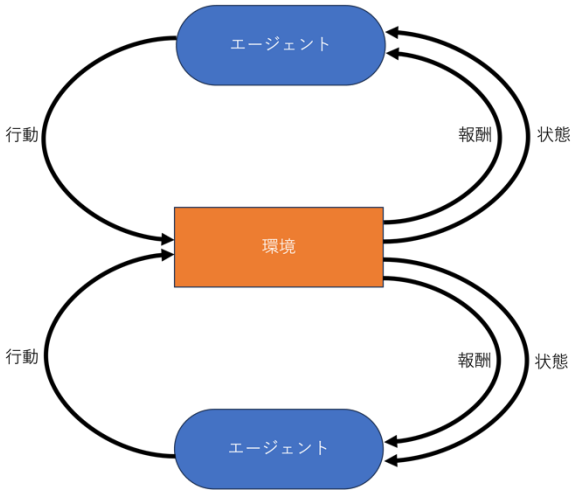

図 4:マルチエージェント強化学習(MARL)の様子

環境

連合学習システムをマルチエージェント深層強化学習の環境として設定する．この設定では，強化学習エージェントは連合学習のクライアントを介して環境である連合学習システムを観測し，影響を与える．この環境内には複数の連合学習クライアントと強化学習エージェントのペアが存在し，それらは一つの連合学習サーバと連携している．

サーバ

サーバは連合学習システムのサーバのことを指す．サーバは環境を初期化するときに，グローバルモデルである初期モデルを生成し，各クライアントに配布する役割がある．また，クライアントからモデルを受け取ったり，クライアントからのリクエストに応じて，モデルを集約し，その後指定されたニューラル機械学習モデルをリクエストしたクライアントに返す役割も担っている．

クライアント

クライアントは連合学習システムのクライアントのことを指す．クライアントは環境を初期化するときに，サーバが配布した初期モデルを受け取り，受け取ったモデルに対して自信が持つ独自のドメインのデータセットを用いて学習し，学習が完了したら，サーバに送信する．そして，サーバにどのクライアントのモデルを集約するのかをリクエストし，サーバから受け取った集約されたニューラル機械学習モデルを再度学習させる役割を持つ．また，クライアントは環境の状態を観測したり，報酬を算出する役割も持っている．

エージェント

エージェントはマルチエージェント深層強化学習システムにおける個々の強化学習エージェントを指す．サーバと各クライアントには 1 つのエージェントが割り当てられ，これらのクライアントとエージェントはペアとして存在する．特にサーバとペアになっているエージェントを集約エージェント，各クライアントとペアになっているエージェントを翻訳エージェントとする．翻訳エージェントは，ペアになっているクライアントを介して環境を観測し，現在の状態 s を取得する．次に，翻訳エージェントは現在の状態 s と保持しているニューラル強化学習モデルを使用して次の行動 a を選択する．選択された行動 a に基づいて，翻訳エージェントはクライアントがどのようなリクエストを送信するかを制御する．リクエストを受け取ったサーバ側の集約エージェントは指定された集約モデルを作成する．さらに，ペアになっているクライアントが報酬 r を算出する

と，現在の状態 s，選択した行動 a，得た報酬 r を使用して式 (1) のように Q 値を算出し，ニューラル強化学習モデルを更新する．

状態

強化学習エージェントがクライアントを介して環境を観測し，現在の状態 s を取得する．クライアントは学習済みのニューラル機械翻訳モデルを使用し，事前に用意された測定データの各センテンスを翻訳し，翻訳精度を評価する．式(3)のように各センテンスの翻訳精度を保持する行列が現在の状態を表す．また，環境は初期化され，クライアントが初期モデルに対して学習を完了した時点での状態が初期状態と呼ばれる．

$$s = \left[score(sentence_0), score(sentence_1), \cdots, score(sentence_k)\right] \quad (3)$$

行動

強化学習エージェントは状態 s のもとで，自分が保持しているニューラル強化学習モデルを使用して次の行動 a を決定する．ここでの行動 a は，どのクライアントを組織化し，集約されたニューラル機械翻訳モデルに対して既定の学習ステップで学習するかを指す．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，一方で単一のクライアントが組織になることも考えられる．組織になるクライアントは，自身とペアになっているクライアントを含まなくても構わない．すなわち，クライアント数が m の時，行動の種類は $\sum_{n=1}^{m} \frac{m!}{n!(m-n)!}$ となる．

報酬

強化学習エージェントが行動を実施し，その結果として環境に影響を与え，それに応じて報酬を得る．この報酬 r は，クライアントが前回の学習済みのニューラル機械翻訳モデルから現在の学習済みのニューラル機械翻訳モデルへの翻訳精度の向上量を示す(式(4))．ニューラル機械翻訳モデルの精度($score_t$)は，測定データの各センテンスの翻訳精度の平均で計算される(式(5))．

$$r = score_t - score_{t-1} \; (-1 \leq r \leq 1) \quad (4)$$

$$score = \bar{s} = \overline{score(sentence_0), score(sentence_1), \cdots, score(sentence_k)} \quad (5)$$

学習

図 5 のようにマルチエージェント深層強化学習を用いた連合学習システムでは，複数のニューラル機械翻訳モデルが存在し，異なる種類の学習が同時に行われており，複雑な仕組みになっている．このため，図 6 のシーケンス図に従っ

て，マルチエージェント深層強化学習による組織化手法の流れを説明する．

まず，サーバはランダムに設定された初期モデルを生成し，各クライアントに配布する．各クライアントは受信した初期モデルを使用して，ローカルで規定のステップの学習を行う．そして，各クライアントは学習を終えると，学習済みのニューラル機械翻訳モデルからモデルのパラメータを抽出し，サーバに送信する．同時に，各クライアントは各自が持つ測定データを使用して，学習済みのニューラル機械翻訳モデルの翻訳精度である $score_0$ を測定し，初期状態 s_0 を取得する．

次に，各クライアントはペアの強化学習エージェントに初期状態 s_0 を渡す．エージェントは状態 s_0 と自身が所持するニューラル強化学習モデルを使用して行動 a_0 を選択し，クライアントに通知する．各クライアントは行動 a_0 に基づいてサーバにリクエストし，サーバはリクエストに応じてクライアントたちを組

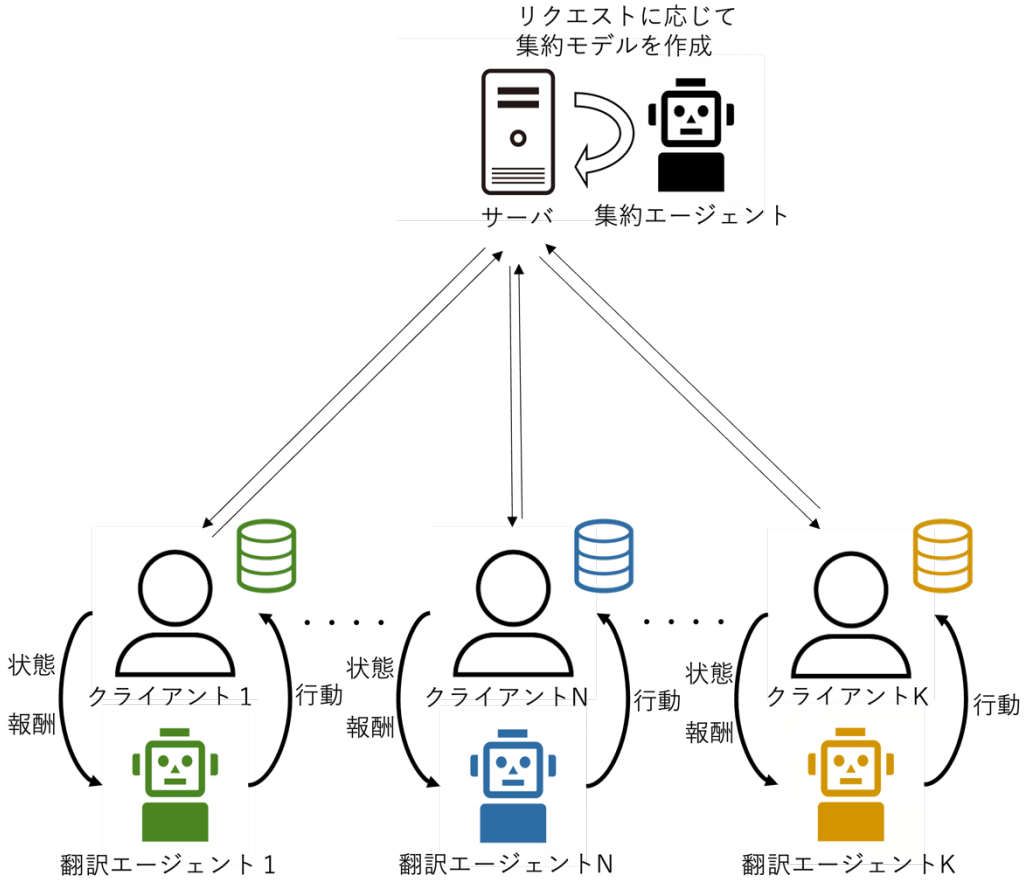

図 5:マルチエージェント深層強化学習を用いた連合学習の概要

織し，更新用データを集約して，集約されたニューラル機械翻訳モデルをレスポンスとしてクライアントに返す．

各クライアントは自身が受信した集約モデルを使用して，ローカルで規定のステップまで学習を続ける．学習が終了したら，学習済みのニューラル機械翻訳モデルから更新用データを抽出し，サーバに送信する．同時に，測定データを使用して学習済みのニューラル機械翻訳モデルの翻訳精度である $score_1$ と現在の

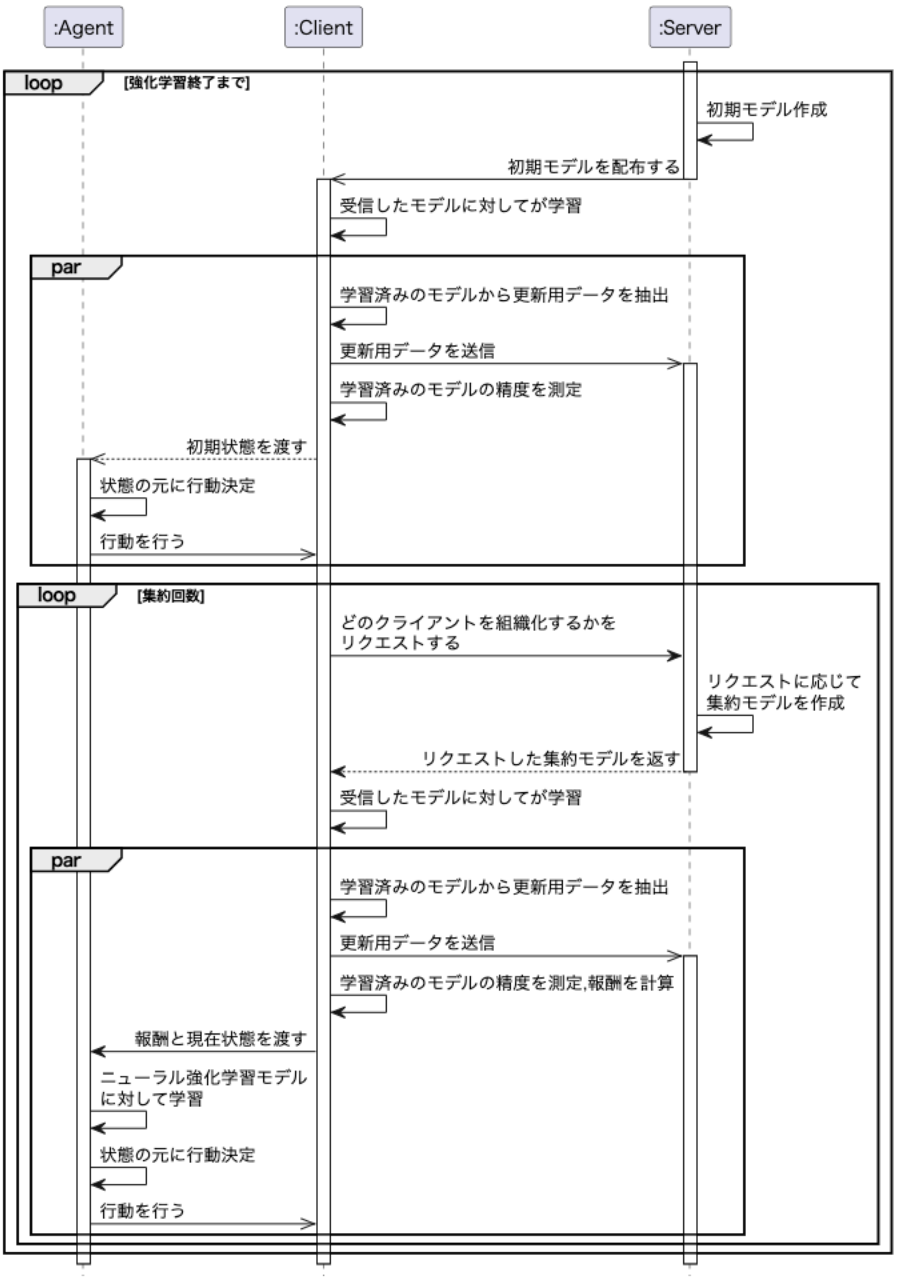

図 6:マルチエージェント深層強化学習を用いた連合学習のシーケンス図

状態 s_1を測定し，エージェントに与える報酬($r_0 = score_1 - score_0$)を算出する．

強化学習エージェントは現在の状態 s_1 と報酬 r_0 を使用して式(1)で $Q(s_0, a_0)$ を更新し，自身が持つニューラル強化学習モデルに学習させ，次の行動 a_1 を選択する．このプロセスを図 6 の loop の部分のように繰り返す．最終的に，各クライアントはエージェントの行動で得たニューラル機械翻訳の集約モデルを自身のグローバルモデルとして保存し，学習を終了する．

第4章 連合学習の効率化

4.1 訓練データの削除

深層強化学習を用いて最善な連携相手の選択方策を獲得するには，エージェントは何度も試行を繰り返さなければならない．試行の度に，ニューラル機械翻訳の学習を行うため，ニューラル機械翻訳の学習時間の長さが深層強化学習全体の学習時間の増大に繋がっている．したがって，方策獲得の所要時間を短縮するには，ニューラル機械翻訳の学習時間を短縮する必要がある．そこで，ニューラル機械翻訳の訓練回数や訓練データのサイズを削減しても，削減前と同様の方策を学習できているかを検証する．方法として，深層強化学習を進める中で，その中のニューラル機械翻訳の学習環境を変化させる．具体的には，エポック数を変えずに，訓練データのサイズ(data size)と学習回数(train steps)を削減した上で訓練する．エポック数は式(6)より求められる．なお式(6)より表 1 のように train steps と data size を (30,0000 回，70,0000 件)，(15,000 回，35,000 件)，(8,571 回，20,000 件)，(4,285 回，10,000 件)，(2,142 回，5,000 件)，(1,071 回，2,500 件) の計 6 回変化させて検証を行なった．

$$epock = train\ steps \div \frac{data\ size}{batch\ size} \tag{6}$$

そして，訓練回数や訓練データのサイズの削減前後の方策の同一性を新たな連合学習の環境下で検証する．検証方法としては，得られた方策を基に新たな連合学習の環境において，同じ条件で検証を試みて評価する．

表 1:検証を行なった train steps と data size

train steps	data size
30,000 回	70,000 件
15,000 回	35,000 件
8,571 回	20,000 件
4,285 回	10,000 件
2,142 回	5,000 件
1,071 回	2,500 件

4.2 状態空間の削除

ニューラル機械翻訳モデルの状態は，本来，獲得した重みパラメータであるが，パラメータ数が多く，状態を表す次元数も大規模化するため，深層強化学習の状態に適さない．そこで，本研究では検証データの評価値群を用いて擬似状態を定義している．したがって，検証データのサイズを小さくし，状態空間を削減することで，各エージェントは最適な方策を獲得しやすくなる．一方で，状態空間の削減は実際の状態と疑似状態の乖離に繋がるため，状態空間の削減と精度の関係を検証する必要がある．したがって，状態空間を 10,000 から 5,000，1,000，100 に削除することでいかに正確にモデルの状態を捉えるかに着目し，深層強化学習を進める中で，検証データのサイズ 10,000 から 5,000，1,000，100 に変化させる．そして，検証データのサイズに応じて方策がどのように変化するかを新たな連合学習の環境下で検証する．検証方法としては，得られた方策を基に新たな連合学習の環境において，同じ条件で検証を試みて評価する．

第5章 評価

5.1 実験環境

実際に評価を行った検証システムについて説明する．提案手法の有用性を検証するために，三つの組織が参加している連合学習システムを考案し，構築した．それぞれの組織が一個の連合学習クライアントを持ち，組織たちが共同で連合学習サーバを維持する．ニューラル機械翻訳モデルの構築は OpenNMT[15]というニューラル機械翻訳システムを利用する．

そして，対訳コーパスは「Wikipedia 日英京都関連文書対訳コーパス」を使用する．「Wikipedia 日英京都関連文書対訳コーパス」は 15 のカテゴリに分割されているのだが，本研究はその中の「伝統文化」,「歴史」,「人名」の三つのカテゴリからそれぞれ 80,000 件の対訳データを抽出し，データセットとして各クライアントに管理させる．表 2 に各データセット間のコサイン類似度を示す．

各クライアントは自分のデータセットから 70,000 件をニューラル機械翻訳モデルを構築するための学習データとして分け，残りの対訳データをクライアントたちが動的組織化する時にモデルの精度を測るための測定データに割り当てる．

各クライアントはニューラル機械翻訳モデルに対して規定のステップを学習するごとに，サーバで 1 回の連合学習の集約を行う．各クライアントはニューラル機械翻訳モデルに対してバッチサイズを 32，トータル学習回数はそれぞれ異なるステップに設定し，規定のステップごとに連合学習のサーバで計 5 回の集約を行う．

連合学習システムを深層強化学習の環境と設定し，学習の際は，機械翻訳モデルを状態とし受け取り，その状態をもとに自分が保持しているニューラル強

表 2:各データセットのコサイン類似度

	伝統文化	歴史	人名
伝統文化	1	0.593	0.825
歴史	-	1	0.815
人名	-	-	1

化学習モデルを通じ，次のモデル選択の行動を決定する．そして，翻訳精度の上昇量である報酬を受け取り，ミニバッチ学習を行う．

ニューラル機械翻訳モデルの精度は BLEU(Bilingual Evaluation Understudy) スコアで算出する. BLEU スコアとは同じ原文の自動翻訳と人間が作成した参考翻訳との違いの測定し計測する単位で，参考翻訳と近ければ近いほどその機械翻訳の精度は高くなる．各手法の有効性を評価する指標として，従来手法である FedAvg によりニューラル機械翻訳モデルを構築し, 翻訳精度を測定する. また,

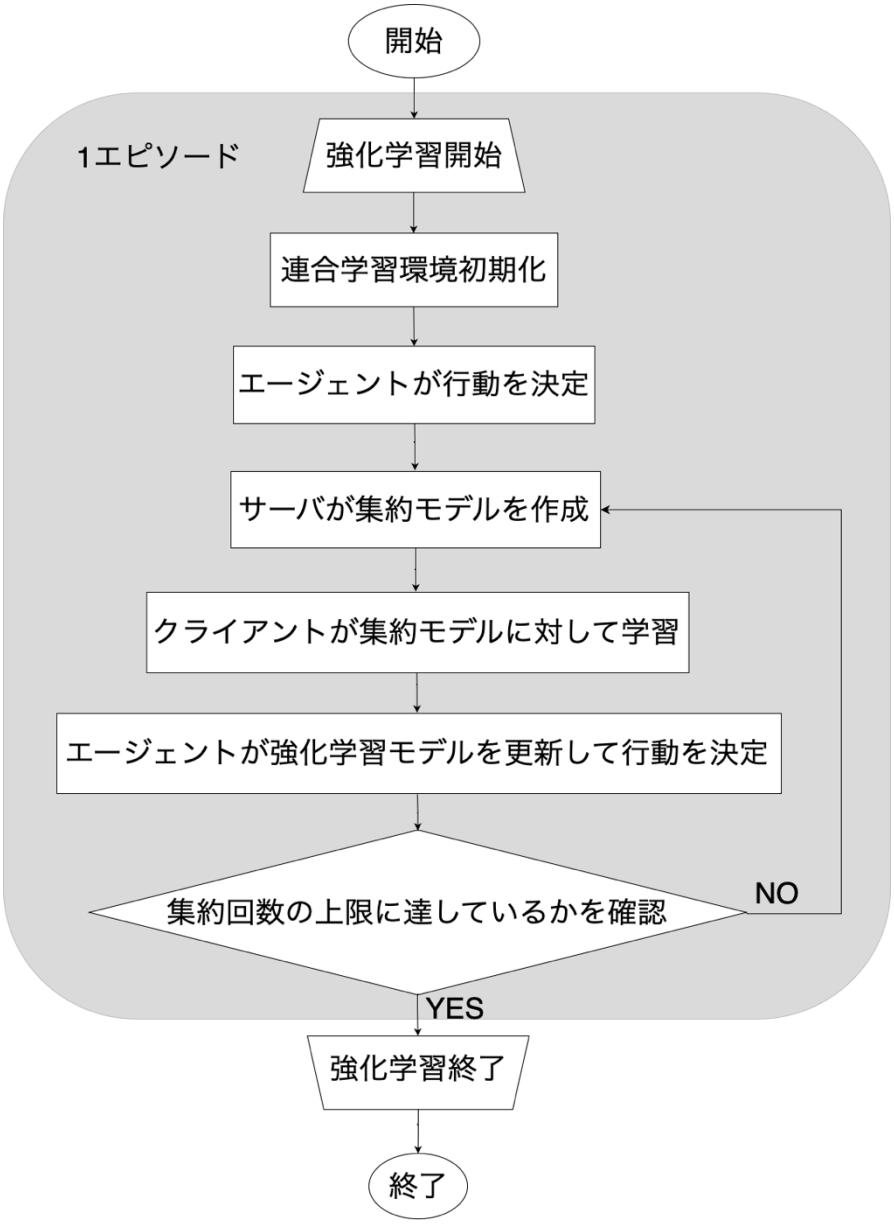

図 7:マルチエージェント深層強化学習を用いた連合学習の流れ

性能評価においてモデルを正確に評価するために，5-Fold 交差検定手法を用いる．5-Fold 交差検定では，データセットを 5 つのサブセットにランダムに分割し，そのうち 4 つのサブセットを訓練用データ，残りの 1 つをテスト用データとして使用する．この訓練とテストの過程を 5 回繰り返し，各回のテスト結果を平均することで，モデルの性能を正確に評価することができる．

マルチエージェント深層強化学習による組織化手法を評価するために，3.2 節に説明した検証システムのもとに，各組織に一個の深層強化学習エージェントを追加する．深層強化学習エージェントは PFRL という PyTorch 向けの深層強化学習ライブラリを利用して実装された．

各深層強化学習エージェントは同一組織のクライアントを通じて状態を観察するため，初期状態の状態空間のサイズは各クライアントが保有している測定データの件数と同じく 10,000 である．ただし，5.3 節における状態空間の再現性の検証においては，状態空間のサイズはその都度変更して検証を行なっている．また，検証システムに三つの組織が存在しているため，深層強化学習エージェントの行動空間のサイズは $\sum_{n=1}^{3} \frac{3!}{n!(3-n)!} = 7$ である．強化学習の中に連合学習で 5 回の集約を行なって，各クライアントが規定のステップの学習を完了し，ニューラル機械翻訳モデルを構築できたら 1 エピソードとしてカウントする(図 7).

5.2 方策の同一性の検証

図 6 と図 7 が示すように提案手法に従って，4.1 節で説明した条件のもと，複数回連合学習でニューラル機械翻訳モデルを構築し，マルチエージェント深層強化学習で 460 エピソードを学習させる．マルチエージェント深層強化学習の各エピソードで構築されたニューラル機械翻訳モデルの精度を記録し，折れ線グラフを作成し A.1 付録に掲載する．また，そこから得られた方策を用いて新しい連合学習の環境のもと再度検証を行い，その精度を比較した．それが表 3 である．greedy 法とは，集約ごとに全通りの組み合わせの集約モデルの中から各クライアントが精度が最も高い集約モデルを自身のグローバルモデルとして選択する学習手法である．詳しくは A.2 付録に掲載する．

30,000 回と比較して「伝統文化」で構築されたニューラル機械翻訳モデルが従来手法 FedAvg で構築されたニューラル機械翻訳モデルの精度より 35.9%上昇した．「歴史」の場合は 17.7%上昇した．「人名」の場合は 26.7%上昇した．15,000 回の場合は，「伝統文化」が 29.8%上昇，「歴史」は 1.28%減少，「人名」

が 29.4%上昇した．8,751 回の場合は，「伝統文化」が 22.3%上昇，「歴史」が 7.88%上昇，「人名」が 17.1%上昇した．4,285 回の場合は，「伝統文化」が 13.8%上昇，「歴史」が 12,8%上昇，「人名」が 0.92%上昇した．2,142 回の場合は，「伝統文化」が 12.0%上昇，「歴史」が 4.32%上昇，「人名」が 0.04%上昇した．1,071 回の場合は，「伝統文化」が 9.44%上昇，「歴史」が 7.56%上昇，「人名」は 1.64%減少した．

結果として，従来の FedAvg 手法と比較して，提案手法は 3 つのカテゴリーにおいて，一貫して精度が向上していた．最も顕著な改善は「伝統文化」のカテゴリーで，30,000 回の学習を経て 35.9%の精度向上を達成した．これは，特定のドメインに対するモデルの専門性が著しく向上したことを示唆し，マルチエージェント深層強化学習を用いた手法で構築されたニューラル機械翻訳モデルの専門性が従来手法より高いと言える．

表 3:各手法の検証結果

	FedAvg	greedy	1,071 回	2,142 回
伝統文化	0.150	0.193	0.164	0.168
歴史	0.105	0.123	0.113	0.109
人名	0.114	0.144	0.112	0.114

	4,285 回	8,751 回	15,000 回	30,000 回
伝統文化	0.171	0.184	0.195	0.204
歴史	0.118	0.113	0.103	0.123
人名	0.115	0.133	0.147	0.144

5.3 状態空間の再現性の検証

5.2 節と同様に，連合学習でニューラル機械翻訳モデルを構築し，各マルチエージェント深層強化学習に 460 エピソードを学習させる．また，そこから得られた方策を用いて新しい連合学習の環境のもと再度検証を行い，その精度を比較した．それが表 4 である．なお，本検証の train steps と data size は全て，(1,071 回，2,500 件)である．「伝統文化」においては状態を表す次元数 10,000 と比べて 100 の精度は 26.5%上昇した．「歴史」においては 19.1%上昇，「人名」においては 46.2%上昇した．同様に，1,000 の精度は「伝統文化」において 15.2%上昇，「歴史」においては 10.8%上昇，「人名」においては 47.6%上昇した．5,000

の精度は「伝統文化」において 3.71%上昇,「歴史」においては 10.5%上昇,「人名」においては 25.3%上昇した．同様に，1,000 の精度は「伝統文化」において 15.2%上昇,「歴史」においては 10.8%上昇,「人名」においては 47.6%上昇した.

表 4:検証結果

	100	1,000	5,000	10,000
伝統文化	0.208	0.190	0.171	0.164
歴史	0.134	0.125	0.124	0.113
人名	0.163	0.165	0.140	0.117

第6章 考察

まず始めに，訓練データの削除における方策の同一性の検証についてだが，表 3 のデータより，主に「伝統文化」と「人名」において訓練データの増加に伴い，精度の向上傾向が見られることから，訓練データが多いとモデルの学習が進むにつれて，より効果的な方策が得られていると考えられる．一方で，訓練データを減らせば方策が獲得しづらくなり，その方策を基に新たな連合学習の環境において，同じ条件で検証を試みると精度が下がるという結果となった．ただ「歴史」においては，訓練データの増加に伴って精度の向上傾向が見られなかった．

また，表 3 のデータと図 8 の各条件下のマルチエージェント深層強化学習にかかった時間を考慮すると，訓練データが増えるにつれて，モデル精度は向上する傾向にある一方で，必要な学習時間も著しく増加している．具体的には，1,071 回の学習で約 104 時間がかかり，30,000 回の学習では約 567 時間が必要であった．そこで，時間対効果を定義し，その観点から考察を行う．時間対効果を式(7)のように定義した．図 9~11 はそれぞれのカテゴリの時間対効果を表すグラフである．

$$時間対効果 = \frac{提案手法の精度 - FedAvg の精度}{学習時間} \tag{7}$$

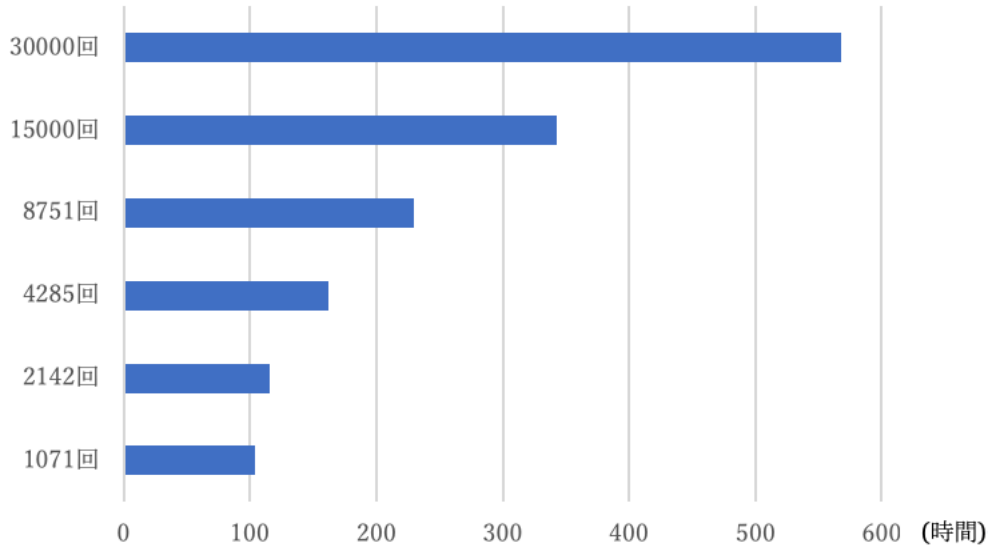

図 8:各検証の時間比較

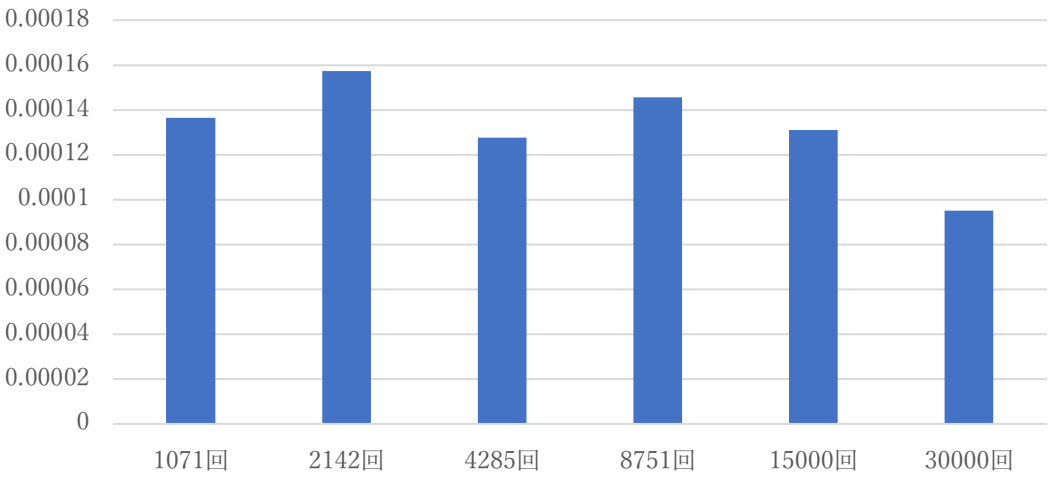

図 9:「伝統文化」における時間対効果

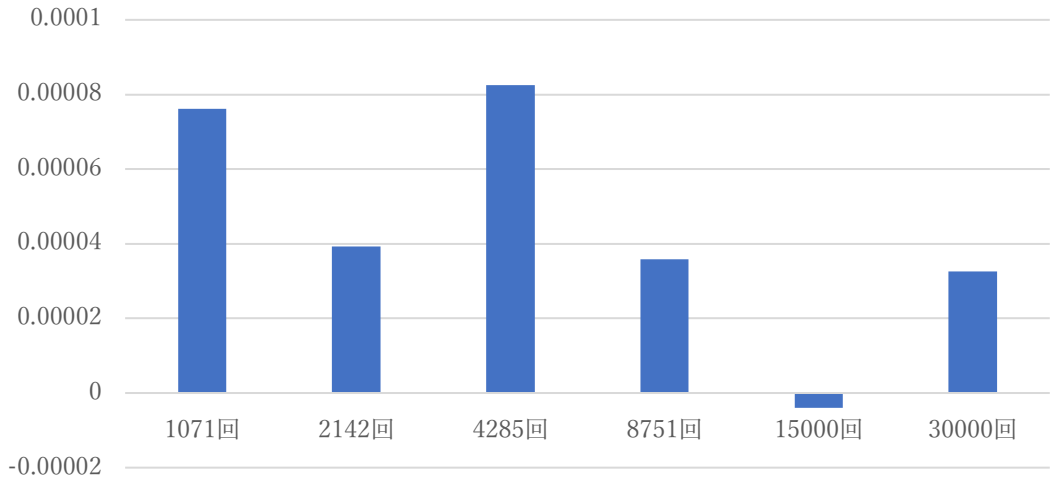

図 10:「歴史」における時間対効果

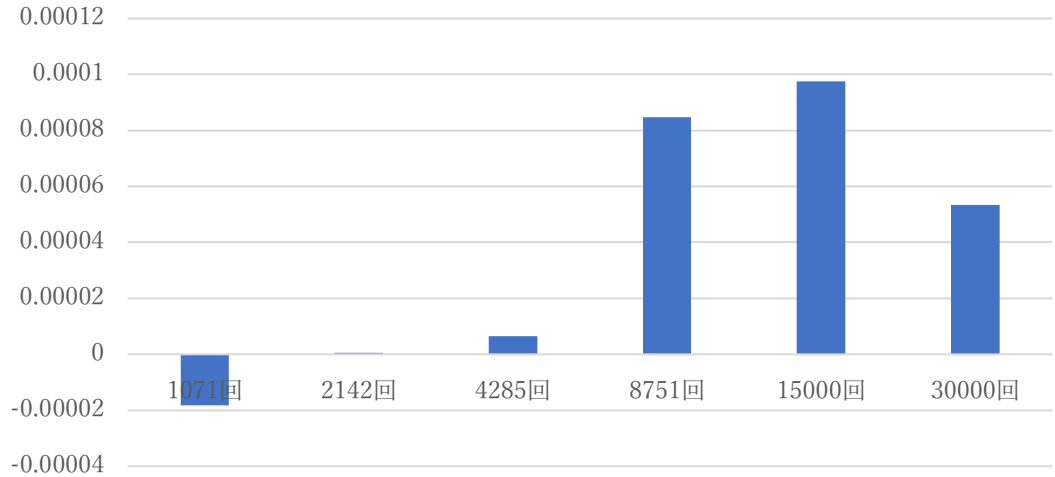

図 11:「人名」における時間対効果

モデル精度の向上が最も大きい「伝統文化」においては，15,000 回での学習が 29.8%の精度向上に対し，約 342 時間かかり，30,000 回での学習が 35.9%の精度向上に対し，約 567 時間かかったことから，訓練データを倍増させたにもかかわらず，精度向上率は約 6%しか増加しなかった．確かに精度は上昇しているが，時間対効果の観点からすると効果が低下している．同様に，「人名」での 15,000 回の学習では約 342 時間で 29.4%の精度向上を達成している一方で，30,000 回の学習を行った場合，約 567 時間での精度向上は 26.7%となり，時間対効果が低下していることがわかる．このことから，特定の訓練回数以降は学習時間に見合う精度向上が得られないことがわかる．また，「人名」の 30,000 回での精度が 15,000 回に比べて下がっていることから，特定の訓練回数を超えると，過学習や学習プロセスの収束などが原因で，追加の訓練回数がむしろ逆効果になる可能性があることがわかる．そのことは「歴史」においても明らかで，4,285 回の学習で 12.8%精度上昇したが，15,000 回で 1.28%精度減少したことから，ドメインによってはより短時間で高い精度向上が達成できることがあり，特定の訓練データ数の増加が必ずしもモデルの精度向上に寄与していないことがわかる．したがって，学習時間と訓練データ数を増やせば，精度は向上する傾向はあるが， 時間対効果の観点で見ると必ずしも効果が上がるとは限らない．また，単に学習時間と訓練データ数を増やしても精度向上につながらないことがあると言える．

次に，状態空間の削減による状態空間の再現性の検証についてだが，状態を表す次元数を 10,000 から 5,000，1,000，100 に減らすと各カテゴリで精度が向上する傾向になった．その理由として，深層強化学習では，エージェントは状態空間を通じて環境を理解し，その情報に基づいて行動を選択する．この次元数が多いと，学習に必要な計算量が増加し，また学習プロセスが複雑になるため，適切な方策を見つけるのに時間がかかる場合がある．一方で，次元数が少ない場合，エージェントは限られた情報からより迅速に学習を進めることが可能になるが，その代償として環境の理解が不完全になる可能性がある．本検証の状態を表す次元数の削減が深層強化学習におけるエージェントの意思決定プロセスを簡略化し，より関連性の高い特徴を捉え，その結果迅速に適切な方策を見つけることに成功したと考えられる．また，「伝統文化」に比べて「歴史」や「人名」において著しく精度が向上した理由については，「人名」のデータには特有のパターンや規則性が存在し，状態空間の削減によって深層強化学習がカテゴリ固有の

パターンや規則性をより効果的に捉え明確に識別できたためと考えられる．「歴史」のテキストにも同様の現象が起こり得るため，これらのカテゴリに適したモデルの学習戦略が形成されたと考えられる．

第7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，各エージェントの選好に応じたモデルを集約する．また，エージェントの選考を反映させる手法として深層強化学習を用いる．各エージェントは深層強化学習によって方策を獲得し，その方策によって最適な連携相手を選択する．本研究の貢献は以下の通りである．

訓練データの削減によるアクションの効率化

マルチエージェント深層強化学習を用いた自己的に組織化する手法は，従来手法である FedAvg より高精度のニューラル機械翻訳モデルを構築できることで提案手法の有効性を示した．また，ニューラル機械翻訳の訓練データの削減により，強化学習のタスクが 39.7%削減した．

状態空間の削減

マルチエージェント深層強化学習において状態を表す次元数を減らすことで各エージェントが最適な方策を獲得しやすくなった結果精度が 25.3%向上した．

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．そして，学習回数の増加に伴い，精度の向上傾向が見られることから，モデルの学習が進むにつれて，より効果的な方策が得られていると考えられる．一方で，学習に要する時間は大幅に増加することが明らかになった．そして，学習エピソード数が増えると精度の向上率が鈍化する傾向にあるため，学習の効率性について，特定のドメインに対する最適な学習エピソード数の決定や，学習時間の最適化は，今後の効率的なモデル構築のための研究課題である．

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., Bengio, Y.: Neural machine translation by jointly learning to align and translate, arXiv preprint arXiv:1409.0473 (2014).

[2] Tu, Z., Lu, Z., Liu, Y., Liu, X., Li, H.: Modeling coverage for neural machine translation, arXiv preprint arXiv:1601.04811 (2016).

[3] Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al.: Google's neural machine translation system: Bridging the gap between human and machine translation, arXiv preprint arXiv:1609.08144 (2016).

[4] McMahan, B., Moore, E., Ramage, D., Hampson, S., A. y. Arcas, B.: Communication-Efficient Learning of Deep Networks from Decentralized Data, Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (A. Singh and J. Zhu, eds.), PMLR, vol. 54, pp. 1273–1282 (2017).

[5] 清藤武暢: プライバシー保護技術としての連合学習の仕組みと最新動向，電子情報通信学会基礎・境界ソサイエティ Fundamentals Review, vol. 16, no. 3, pp. 196–204 (2023).

[6] Mnih, V., et al.: Human-level control through deep reinforcement learning, Nature, vol. 518, no. 7540, pp. 529 (2015).

[7] Sutton, R. S., Barto, A. G.: Reinforcement learning: An introduction, MIT press, vol. 1, no. 1 (1998).

[8] Watkins, C. J. C. H., Dayan, P.: Technical note q-learning., Machine Learning, vol. 8, pp. 279–292 (1992).

[9] Watkins, C. J. C. H., Dayan, P.: Q-learning, Machine learning, vol. 8, no. 3-4, pp. 279-292 (1992).

[10]Watkins, C. J. C. H.: Learning from delayed rewards, PhD thesis, Cambridge University (1989).

[11] Mnih, V., et al.: Playing atari with deep reinforcement learning, arXiv preprint arXiv:1312.5602 (2013).

[12]Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M.: Playing atari with deep reinforcement learning, NIPS Deep Learning Workshop 2013, arXiv:1312.5602 (2013).

[13]Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sardik, A., Antonoglou, L., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D.: Human-level control through deep reinforcement learning, Nature, vol. 518, Issue 7540, pp.529-533 (2015).

[14]van Hasselt, H., Guez, A., Silver, D.: Deep reinforcement learning with double q-learning, arXiv preprint arXiv:1509.06461 (2015).

[15]Klein, G., Kim, Y., Deng, Y., Senellart, J., Rush, A.: OpenNMT: Opensource toolkit for neural machine translation, Proceedings of ACL 2017, System Demonstrations, Association for Computational Linguistics, pp. 67–72 (2017).

付録

A.1 強化学習の学習過程

マルチエージェント深層強化学習の各エピソードで構築されたニューラル機械翻訳モデルの精度を記録し，図 12~16 のような折れ線グラフを作成した.

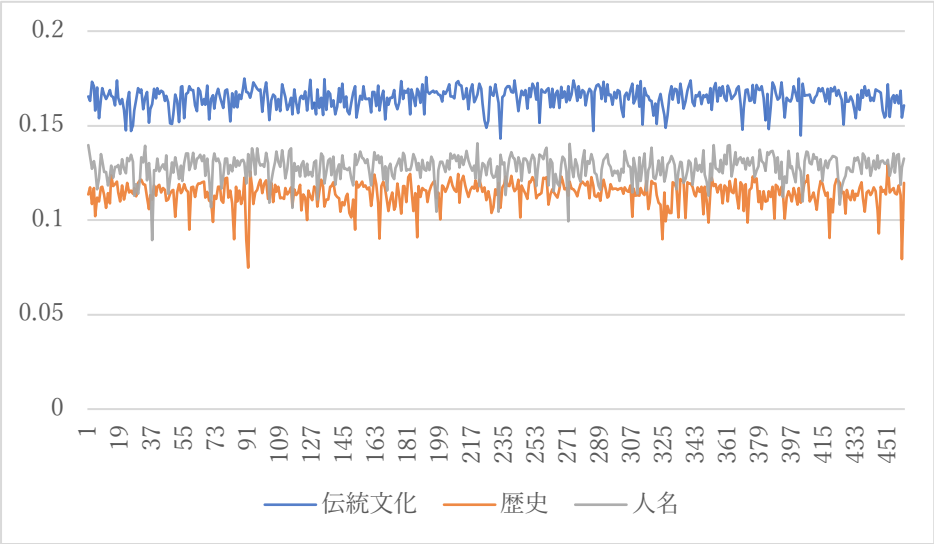

図 12:15,000 回の検証

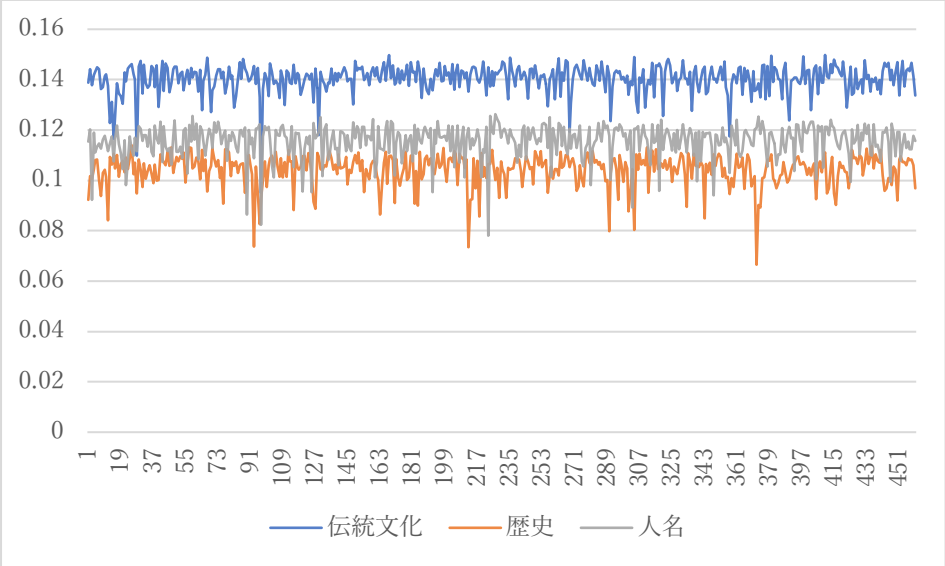

図 13:8,751 回の検証

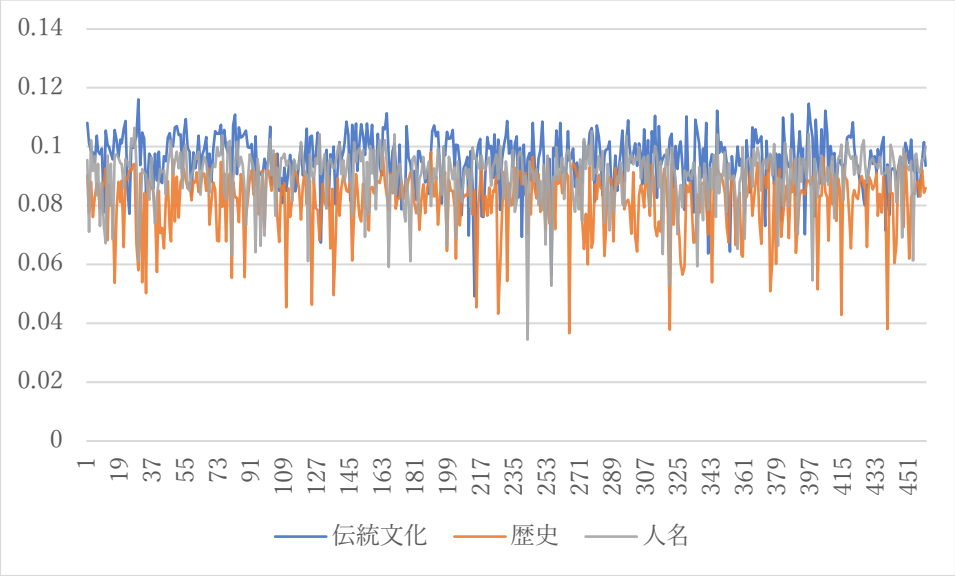

図 14:4,285 回の検証

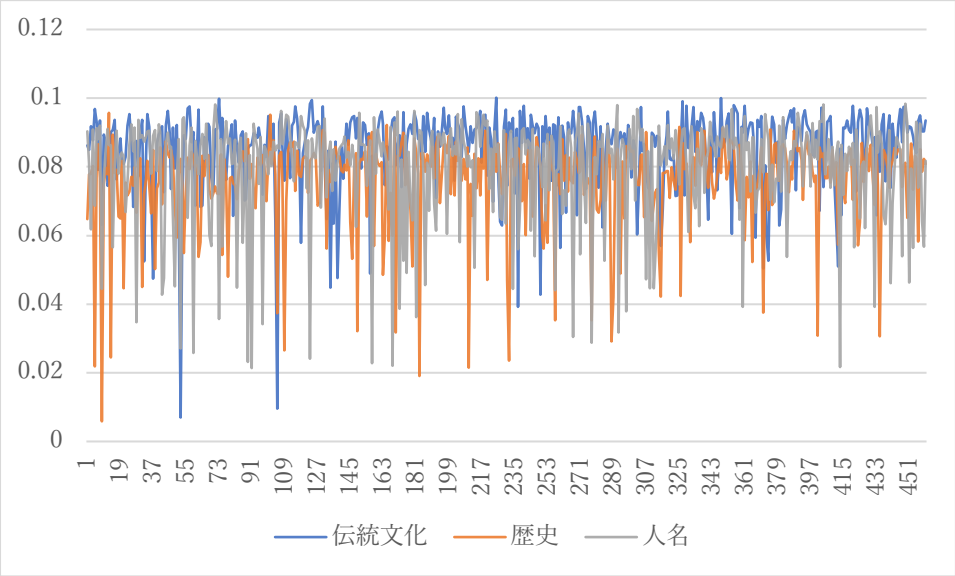

図 15:2,142 回の検証

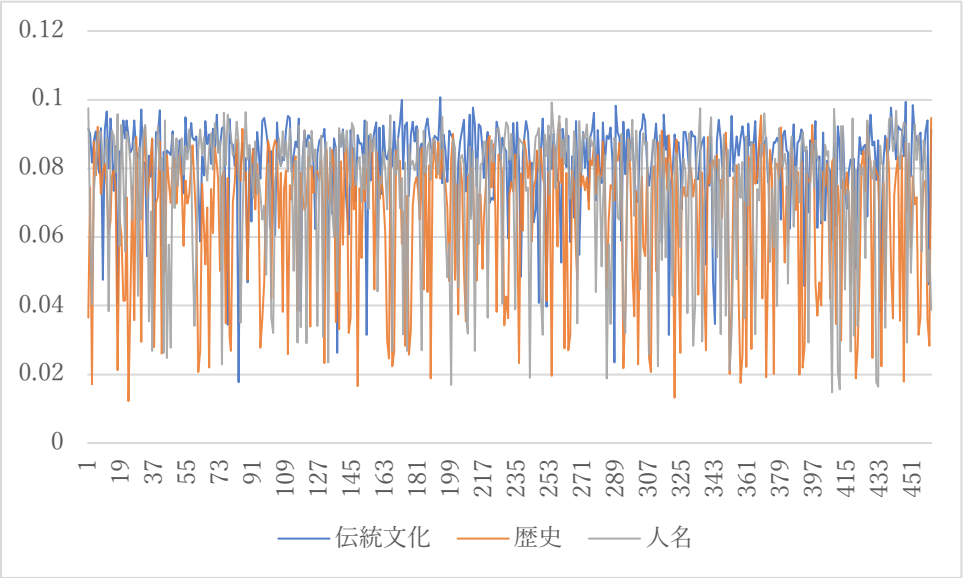

図 16:1,071 回の検証

A.2 Greedy 法

greedy 法とは，連合学習の参加した組織たちが自分の組織が管理している対訳コーパスのドメインに特化したニューラル機械翻訳モデルを作成するという目的に特化した手法である． 図 17 が示すように，まず集約エージェント(サーバ)が初期モデルを生成し，翻訳エージェント(各クライアント)に配布する．翻訳エージェントは受け取ったモデルを使用して学習を行い，既定の学習ステップに到達したら，学習モデルを集約エージェントに送信する．次に，集約エージェントは各翻訳エージェントから受け取った学習モデルを基に，全ての組み合わせのパターンの集約モデルを作成する．その後，翻訳エージェントが集約エージェントから全ての集約モデルを取得し，測定データを使用してローカルで集約モデルの翻訳精度を測定する．精度が最も高い集約モデルをグローバルモデルとして選択し，次の学習に進む．最後の集約後，翻訳エージェントは集約エージェントから全ての集約モデルを取得し，精度が最も高い集約モデルに対して，保有している学習データを使用して学習し，自身のドメインでの翻訳精度を向上させる． このように， greedy 法による組織化の手法を用いることで，各組織が参加する連合学習において，自身のドメインに特化したニューラル機械翻訳

モデルを構築することができる.

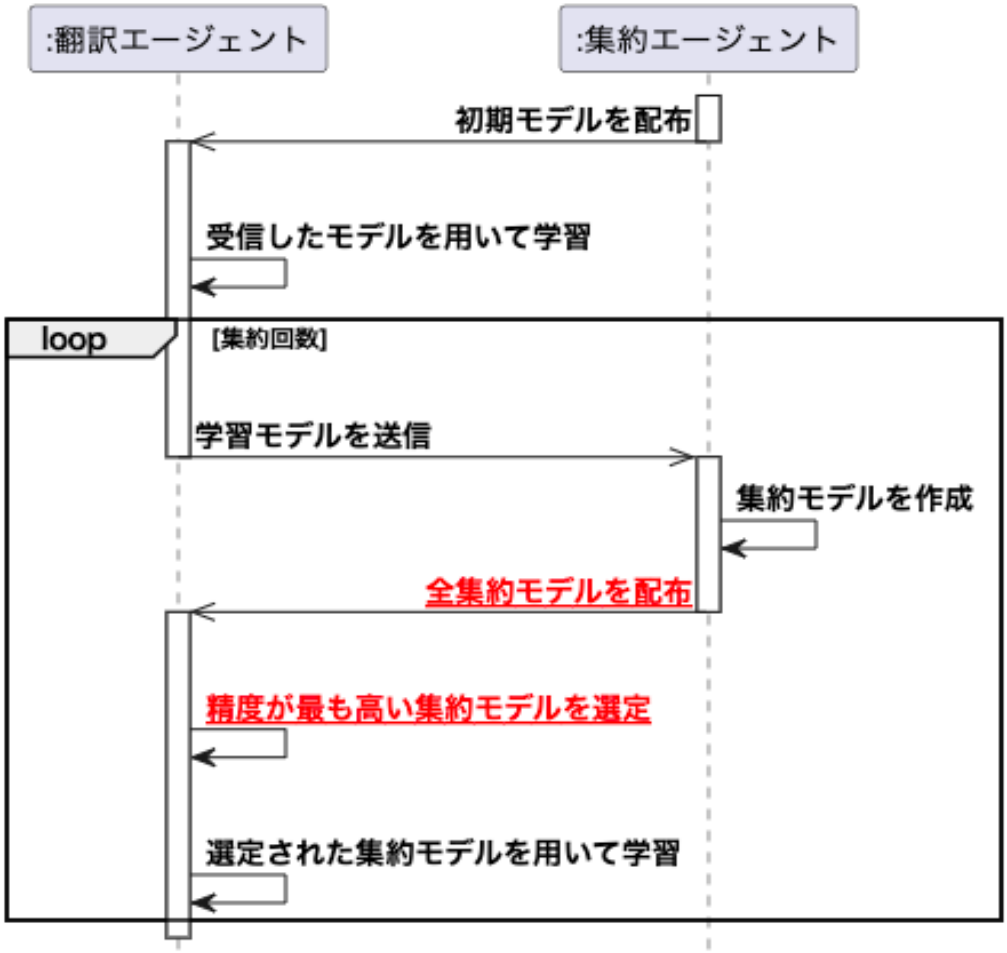

図 17:greedy 法のシーケンス図