観光オブジェクト認識モデルのユーザ参加型構築手法の提案

冨田 周作[1,2]　中村 優吾[3]　諏訪 博彦[1,2]　安本 慶一[1,2]

概要：近年，個人が撮影した写真などプライバシ情報を含むデータそのものをサーバに集約することなく，認識モデルの構築に必要なパラメータを用いて機械学習を行う手法として Federated Learning が注目を集めている．しかし，Federated Learning は，大きな計算能力と記憶容量を備えた集約サーバの設置が前提となっており，観光情報提供のシナリオにおいては，観光客が持つ端末と集約サーバ間の通信や端末上でのモデル更新が高頻度に行われるため，ユーザ端末の大幅な電力消費が発生する．そのため，これらの負担を抑制しつつ，効率的にモデル構築するための手法が必要となる．本研究では，観光客が持つ端末間での直接的な通信を活用した Federated Learning に基づくモデル構築の手法を提案する．また，より少ない通信回数で，認識モデルの精度を効率的に向上させるために，FedAvg(モデルパラメタを平均する方法) に基づくモデル統合手法の特性を網羅的に調査した．評価実験では，10 クラスのオブジェクトのデータセットである Cifar10 の一部を訓練した CNN モデルを 231 個構築し，各組み合わせのモデル統合で得られる accuracy を記録した．評価結果より，VGG16 の全体のパラメタ同士を単純平均で統合する場合，37315 パターンで最も多く accuracy が統合前の自身のモデルと比較して向上した．この結果に関して，FedAvg を適用する相手のモデルの accuracy が自身のモデルよりも高い場合 accuracy が向上する傾向にあり，相手のモデルパラメタの統合の判断が可能となることを示した．

Proposal of Participatory Training Method for Tourism Object Recognition Model

SHUSAKU TOMITA[1,2]　YUGO NAKAMURA[3]　HIROHIKO SUWA[1,2]　KEIICHI YASUMOTO[1,2]

1. はじめに

近年，医療や農業等の様々な分野で徐々に AI が普及するようになり，観光業界でも AI を活用した観光サービスの高度化が進んでいる [1]．このような AI の活用が今後増加すると予想されるが，AI の構築には大量のデータが必要となる．そのため，大企業等が保有するビッグデータや個人が所持するスマートフォン内の写真や検索履歴等も対象とできることが望まれる．しかし，これらのデータにはプライバシを含むデータも存在し，使用方法によってはプライバシの侵害に繋がる可能性がある．社会全体には大量のデータが存在するが，プライバシ保護のためにそれらを学習した AI の構築が困難となっている．社会全体の膨大なデータから汎化性能の高い AI を構築するには，データのプライバシ保護が課題となる．

上述の課題の解決に向けて，Google が Federated Learning を提案している [2]．この手法では，分散したエッジ端末のデータを端末上で訓練し，そのモデルを集約サーバに集約することで間接的に全デバイスのデータを訓練している．訓練データではなくモデルのパラメタや勾配のみが外部に送信されるため，訓練データのプライバシ保護が可能となる．しかし，十分な計算能力と記憶容量を備えた集約サーバの設置が前提となる．

Federated Learning の観光分野での活用において，観光地のコンテキスト認識のための観光地特有のオブジェクト (鹿，神社など) 認識モデルを観光客が持つ写真データから構築するシナリオを考える．観光客の所持データを使用するためには Federated Learning が有効であり，モデル構築に参加する観光客が多い程，多様な観光オブジェクトに対

[1] 奈良先端科学技術大学院大学，Nara Institute of Science and Technology
[2] 理化学研究所 革新知能統合研究センター (AIP), RIKEN, Center for Advanced Intelligence Project (AIP)
[3] 九州大学，Kyushu University
[†1] 現在，マルチメディア, 分散, 協調とモバイルシンポジウム Presently with DICOMO2021

応するモデルの構築が可能となる．しかし，モデル更新のために集約サーバとの高頻度・大容量の長距離無線通信が必要となり，観光客端末の電力を多大に消費するとともに，通信費を増加させる．端末の電力消費抑制とモデルの性能向上を両立する手法が必要となる．

本研究では，観光客の端末間での直接的な通信を活用した Federated Learning に基づくモデル構築手法を提案する．また，より少ない通信回数でモデルの精度を向上させるために，モデルパラメタのみ使用する FedAvg をベースとする複数の統合手法の特性を網羅的に調査した．提案手法における統合手法では，モデルパラメタを加算により統合するが，モデルの出力層とその他の層のパラメタをそれぞれ異なる比率または等しい比率で加算する手法を複数導入した．各統合手法による精度への影響を調査し，精度向上に有効な統合手法を検討した．提案するモデル構築手法としては，観光客同士がすれ違う時 (以降コンタクトと呼ぶ) に自分と相手のモデルパラメタを統合し自身のモデルを更新する．コンタクトの繰り返しで，観光客の移動と共にすれ違う複数の観光客端末のモデルとの統合が可能となる．また，相互のモデルの認識能力を維持できる FedAvg をベースとするモデルパラメタの統合方法を活用することで，少ない通信回数でモデルの精度を大きく向上できることが期待される．端末間の近距離直接通信及びモデル統合により，各観光客の端末の電力消費を抑制する効率的なモデル構築が可能になると考える．

評価実験では，モデルのタスクを画像分類とし，独自に構築した CNN14 層モデルと既存の VGG16 の 2 種類のモデルに，本稿で述べる 3 種類の統合手法を適用した場合の accuracy を取得した．また，鹿や車等の 10 クラスのオブジェクトのデータで構成される Cifar10 を使用し，クラスごとにデータ数を変えた 231 パターンのデータセットを別々に訓練したモデルを 231 個生成した．個々のモデルに対し，同一のモデルも含め全てのモデルの組み合わせの統合を行い accuracy が統合前より向上したパターンの総数で評価した．その結果，VGG16 のモデルを単純平均した場合，統合後の accuracy が向上する組み合わせが最も多く全 53361 パターンの中から 37315 パターン得られた．また，この組み合わせの結果に関して，FedAvg を適用する相手のモデルの accuracy が自身のモデルよりも高い場合，自身の accuracy が向上する傾向があった．この傾向より，コンタクト相手のモデルの accuracy 等に基づく統合の判断が可能となることが示唆された．

本稿の構成は以下の通りとする．第 2 章では，Federated Learning に関する関連研究について概説する．第 3 章では，観光分野における本研究の想定環境について述べ，その環境に沿ったシナリオと問題設定を述べる．第 4 章では，提案手法について述べ，統合時におけるパラメタの更新方法及び統合方法について概説する．5 章では評価実験の内容，6 章にその結果についての考察を概説する．7 章では，本稿のまとめと今後の展望について述べる．

2. 関連研究

本研究が対象とする Federated Learning の概念は，データセンタ以外の場所に分散しているデータを漏洩のリスクを最小限に抑えた上で機械学習モデルの訓練に用いるために Google により提案された．Yang[3] らは Federated Learning の先行研究について調査しており，その概念，分類，定義，応用事例を述べている．Yang[3] らによると，Federated Learning は Horizontal Federated Learning，Vertical Federated Learning 及び Federated Transfer Learning の 3 種類に分類される．これらの手法は訓練対象のデータとその内部に存在する特徴量の取りうる範囲の違いによって分類されるが，いずれもモデルを集約するサーバが存在するという共通点を持ち，全体の処理に重要な役割を担っている．

Lee[4] らは集約サーバを使用せずにエッジ端末間の通信で Federated Learning を実施する手法を提案している．Lee らの提案した Opportunistic Federated Learning[4] は，エッジ端末のユーザ毎に認識対象が異なる場合でも，それぞれの認識対象に合わせた訓練を可能とする手法である．この手法では，通信可能な距離に位置するエッジ端末間でのみ訓練を実行し，それ以外の端末とは訓練しない．実験では，画像分類タスクにおいて認識対象のラベルが異なるモデルを多数生成し，エッジ端末の移動シミュレーションに基づいて対象のモデルを訓練する実験を行なっている．エッジ端末間での近距離通信を活用しているため，通信による電力消費が抑制される手法となっている．しかし，モデルの更新手法として FedSGD[2] に関連した手法を採用しており，訓練データをモデルに入力して得られる勾配を主に通信している．そのため，相互の端末内のデータに対する勾配が更新に必要となり，データ数が多い程互いの勾配を通信する回数が増加する．また，十分な更新のために多くの端末との勾配の通信が必要であり，通信の回数が制限されていないため近距離通信における電力消費が多くなる．

Lalitha[5] らは，集約サーバが設置されないネットワークのノード間で行う分散学習について，ネットワークの任意のノードの近傍ノードに対し，各近傍ノード内のデータ分布に関する情報を活用することにより，対象ノードに適したモデルを選択する手法を提案している．しかし，この手法では前提としてネットワークのノード間の通信経路が固定されており，エッジ端末等の移動が予想されるノード間で構成される環境に対しては適用することが困難である．

Chen[6] らは，オリジナルの Federated Learning[3] において重要な集約サーバへの依存度を低下させる手法を提案している．この手法では，集約サーバへのアクセスが困難

な端末も訓練への参加が可能になるように，互いに近距離に位置するエッジ端末間のローカルネットワークでモデルを集約している．各端末にも集約処理を実行させることで，集約サーバが全てのエッジ端末を網羅できない状態でも全体の端末のモデルの訓練を可能としている．しかし，この手法は，近隣の通信可能な全てのエッジ端末と通信するため，端末数に応じて通信回数が増加する．そのため，モデル集約の際に生じる電力消費等のコストが増大する．

Kim[7] らは，ブロックチェーンを活用することで特定のサーバにおけるグローバルモデルの構築を必要としない手法として，BlockFL を提案している．BlockFL は，集約サーバをブロックチェーンで代用し，各端末のモデルの更新情報を記録したブロックを生成する．そのブロックの情報をもとに各端末内で共通のグローバルモデルを構築し，再度ローカルで訓練を繰り返す．これにより，各ローカル端末がグローバルモデルを持つことになるため，集約サーバの必要性が無くなる．しかし，グローバルモデルの構築には，全端末のモデル更新後のパラメタの差分が必要があり，多数の端末が BlockFL に参加する場合，グローバルモデル構築のためのパラメタの差分が揃うまでに時間がかかる．そのため，BlockFL のアルゴリズム全体の処理時間が遅くなる．

これらの関連研究では，Federated Learning に関する概念や手法について研究されている．基本的に Federated Learning ではモデルを集約するサーバが必要となるが，集約サーバへの依存性を低減する手法も複数提案されている．しかし，端末間のネットワークの固定が必要な手法や，間接的に集約サーバが必要となる手法，訓練全体で多くの端末との複数回の近距離通信が必要な手法等，通信環境が比較的安定する中での手法が多く，通信環境が全体的に不安定な環境での適用は困難である．本研究で提案する手法では，端末間の直接的な通信によって通信環境に依存しない Federated Learning を実施し，端末間の通信回数の制限により電力消費を抑制しながら各モデルの精度を最大にすることを目的としている点で，既存研究と異なっている．

3. 問題設定

本章では，本研究におけるユーザ間のオブジェクト認識モデルの構築手法を検討する前提となる想定環境とシナリオ，その中で対象となる課題について述べる．

3.1 想定環境と問題定義

本研究では，Federated Learning に基づいて観光オブジェクト認識モデルをユーザ参加型で構築するための手法の実現を目指している．**表 1** に想定環境の要素とその説明を示す．

想定環境である観光エリアの集合を A とする．それぞれの観光エリア $a \in A$ には認識したい観光オブジェクトの集合 O_a が存在している．O_a には移動するオブジェクト（動物，群衆，屋台など）や固定されたオブジェクト（建物，鳥居，樹木など）が含まれる．A の観光エリア間及び内部には，複数の観光客が滞在及び移動をしており，各観光客はスマートフォン等のモバイル端末を所持している．モバイル端末の集合を端末 C_{mobile} とする．また，各観光客はモバイル端末を 1 台のみ持つ．モバイル端末の他には，サイネージ等の固定位置にある端末の集合 $C_{stationary}$ がある．以上より，想定環境内に分散する端末の集合 C を式 (1) で表す．

表 1 想定環境に登場する記号一覧

要素	定義
A	観光エリアの集合
A_c	認識能力を強化したいエリア
O_a	認識オブジェクトの集合
O_c	認識強化オブジェクトの集合
$C_{stationary}$	固定端末の集合
C_{mobile}	移動端末の集合
C	全端末の集合
R	端末が通信可能な範囲
D_c	端末 c が持つデータの集合
D	全データ
M_c	端末 c が持つモデル
W_c	モデル M_c のパラメタ
T	想定環境内の時間 t の集合
$pos(c,t)$	端末 c の時刻 t の位置
$cn(c,c',t)$	時刻 t での端末 c，c' のコンタクト
CN	全端末のコンタクト

$$C = C_{stationary} \cup C_{mobile} \tag{1}$$

想定環境におけるタイムスロット (時刻) の集合を T とすると，$c \in C_{mobile}$ は，$t \in T$ によって位置が変化する．そのため，c の t における位置を $pos(c,t)$ と表記する．時刻 t にある端末 c と別の端末 c' が互いに通信可能な範囲 R にいる場合，それらの端末が互いにコンタクトしているとみなし，$cn(c,c',t)$ と表記する．全てのコンタクトの集合 CN を式 (2) として定義する．

$$CN = \bigcup_{c,c' \in C, t \in T} cn(c,c',t) \tag{2}$$

各モバイル端末 $c \in C_{mobile}$ が収集・所持しているデータを D_c と表記する．固定端末はデータを収集しない．そのため，$c \in C_{mobile}$ においては $D_c \neq \phi$，$c \in C_{stationary}$ においては $D_c = \phi$ となる．以上より，想定環境内に存在する全てのデータを D として式 (3) で定義する．

$$D = \bigcup_{c \in C_{mobile}} D_c \tag{3}$$

各端末 $c \in C$ は独自のデータ D_c で学習した観光オブジェクト認識モデル M_c を保有しており，M_c のパラメタ W_c を Federated Lerning に基づく方法で他のパラメタ $W_{c'}$

と統合する．ここで，各端末 c が持つモデル M_c には認識能力を強化したい観光エリアが設定されており，$A_c \subseteq A$ とする．A_c 内のオブジェクトの集合 O_c を式 (4) で定義する．

$$O_c = \bigcup_{a \in A_c} O_a \tag{4}$$

時刻 t で任意の端末と他の端末の 1 対 1 のコンタクトが発生した場合，両方の端末間でパラメタの通信が可能とする．複数の端末と通信が可能な状況にある場合においては，その中の 1 つのみと通信可能とする．この制約を二値変数 $x_{cn(c,c',t)}$ を用いて式 (5) として定義する．

$$\sum_{cn(c,c',t) \in CN, c \neq c'} x_{cn(c,c',t)} \leq 1, \forall c \in C, \forall t \in T \tag{5}$$

式 (5) において，$x_{cn(c,c',t)}$ は端末間の通信の有無を表す変数であり，端末 c と任意の端末 c' との間で通信を行う場合は $x_{cn(c,c',t)} = 1$，通信しない場合は $x_{cn(c,c',t)} = 0$ となる．

消費電力抑制のため，各端末が他の端末のパラメタを統合できる最大回数を L と表記する．この制約は式 (6) として記載できる．

$$\forall c \in C, \sum_{cn(c,c',t) \in T} x_{cn(c,c',t)} \leq L \tag{6}$$

本研究の目的は，通信回数を抑制しながら，モデルの改善度合いを最大化するようように，モデルパラメタの交換を行うコンタクトを選択することである．M_c に $M_{c'}$ のパラメタを統合することによるモデル精度の改善度合いを $Improve(M_c, M_{c'})$ と表記する．この時，本問題の目的関数は以下の式 (7) のように表現することができる．

$$\begin{aligned} Maximize \sum_{c \in C} \sum_{c' \in C\{c\}} \sum_{t \in T} \sum_{cn(c,c',t) \in CN} \\ x_{cn(c,c',t)} \cdot Improve(M_c, M_{c'}) \\ subject\ to\ (5), (6) \end{aligned} \tag{7}$$

3.2 想定シナリオ

本研究における先行研究 [8] で，前節の想定環境に基づく想定シナリオについて述べている．想定シナリオの概略図を**図 1** に示す．シナリオ内では奈良県奈良市を対象としており，奈良公園，春日大社，東大寺を観光エリアとしている．観光エリアにはそれぞれ観光オブジェクトが存在しており，観光エリアを特徴づけるオブジェクトであるとする．これらのオブジェクトの集合を以下の $O_{奈良公園}$,$O_{東大寺}$,$O_{春日大社}$ とし，一部のオブジェクトは観光エリア間で重複している．

$O_{奈良公園}$ = { 牡鹿, 牝鹿, 小鹿, 鹿煎餅の売店,...}

$O_{東大寺}$ = { 仏像, 池, 鯉, 桜, 牡鹿, 牝鹿...}

$O_{春日大社}$ = { 藤, 鳥居, 池, 鯉, 社,...}

観光エリア間及び内部では，観光客が所持するスマートフォン等のモバイル端末 $c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9$ が移動し，互いの観光オブジェクト認識モデルのパラメタの通信を行う．図 1 に示す通り，各端末が所持する観光エリア毎の写真データの数は均一でない．各端末のモデルはそれぞれが保存している写真データについて訓練済みであるとする．データを所持していない端末に関しては，未訓練の状態であるとする．また，各観光エリアには，土産店等の敷地内にサイネージが 1 台設置されており，$c_{奈良公園}$,$c_{東大寺}$,$c_{春日大社}$ とする．これらのサイネージには写真データは保存されていないが，観光エリア内の一部のオブジェクトに対する認識能力があるモデルを持つ．

想定シナリオでは，端末 c_1 のユーザの行動に着目する．各モバイル端末のユーザは最初**図 2** のように点在しており，時間の経過と共に移動する．対象の c_1 のユーザは図 2 の左側に位置する近鉄奈良駅付近から図 2 の右下側にある春日大社まで移動する．その途中で，奈良公園，東大寺，春日大社を観光し，**図 3** の通りに $c_3, c_4, c_{奈良公園}, c_7, c_2, c_{東大寺}, c_5, c_6, c_9, c_8, c_{春日大社}$ のユーザ及びサイネージとコンタクトし，端末間での通信が可能なタイミングを複数回設ける．他のユーザ及びサイネージとのコンタクトを活用し他のモデルにアクセスしながら c_1 の観光オブジェクトモデルが更新される．また，c_1 の他のモバイル端末やサイネージにおいても同様に，他の端末とのコンタクトが生じた場合に相互のモデルにアクセスし更新を行う．想定シナリオでは，それぞれの端末が他の端末とのコンタクトを活用してモデルを更新する．

3.3 課題とアプローチ

3.3.1 本研究における課題

本研究で対象とする課題として以下の 2 点が挙げられる．

本研究の想定環境に関して，各観光客のモデル更新に従来の Federated Learning[2][3] を活用する場合，全端末がアクセス可能な集約サーバが必要となる．しかし，観光客の数が膨大になると，集約サーバやそのアクセスに必要なネットワークに負荷が集中する．そのため，従来の Federated Learning の集約サーバを活用する手法では，モデルの訓練に参加可能な端末数の制限や，更新頻度の低下が発生する．そのため，**（課題 *1*）集約サーバを経由しない他の端末へのアクセス**を解決しなければならない．

また，本研究では観光オブジェクト認識モデルとして CNN モデルを採用しており，CNN モデルは数十 MB 以上のパラメタを持つ場合が多いため，1 回のモデル更新におけるモデルパラメタの通信に消費する電力が多くなる．Lee[4] らの手法では，端末間での訓練で直接的な通信を複数回行っており，通信毎に多くの電力が消費される．観光客は端末の充電が可能な場所に滞在しているとは限らない

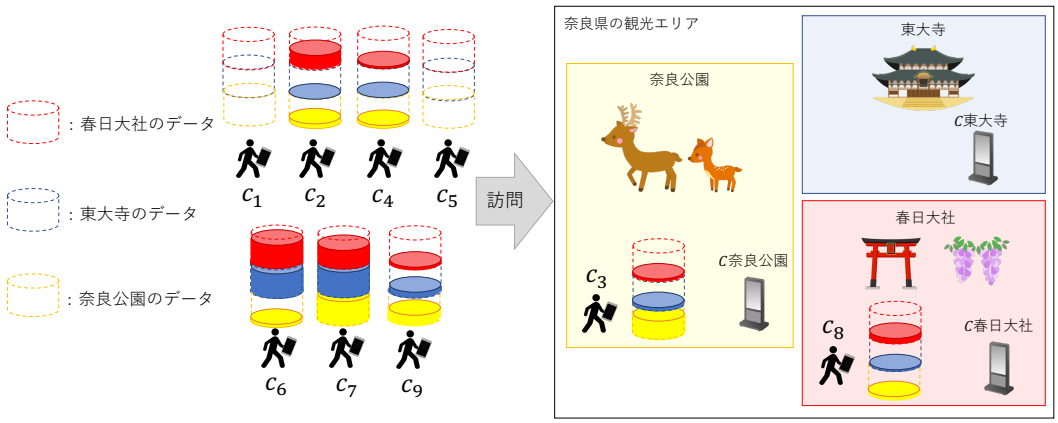

図 1　シナリオの概略図

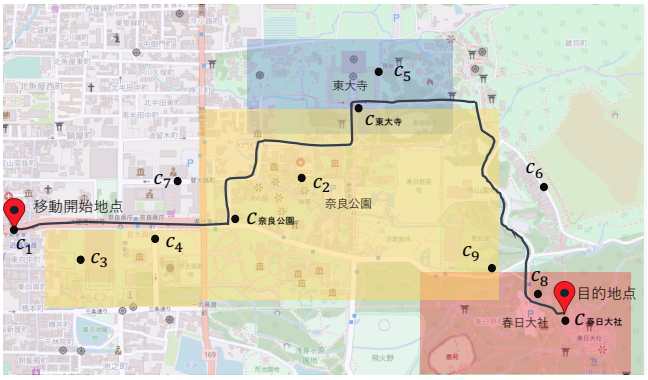

図 2　各ユーザ及びサイネージの初期位置

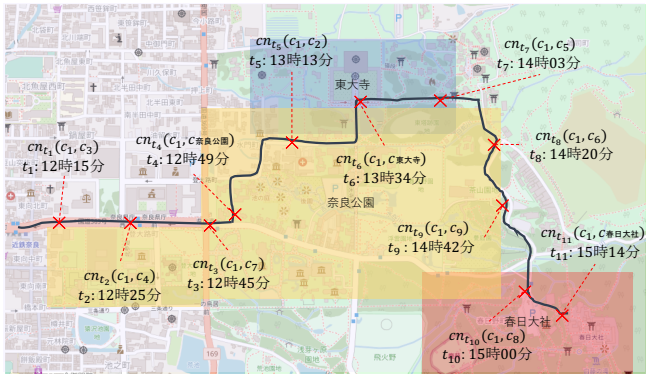

図 3　各端末の位置とコンタクト

ため，端末の電力消費の抑制が求められる．電力消費の抑制のため，**(課題 2) モデルパラメタの通信回数を最小限に抑えながら精度を最大にする手法**を考案する必要がある．

3.3.2　各課題に対するアプローチ

課題 1 に対して，集約サーバの介在なしで他の端末のモデルにアクセスする方法として，端末間で直接通信する方法を採用する（WiFi Direct や BLE の使用を想定している）．端末での直接通信を主なアクセス方法とすることでネットワークや集約サーバに依存することが無くなり，より多くの観光客がモデルの訓練に参加できると考える．

課題 2 に対して，最小の通信回数で最大の精度を得る方法として，統合後のモデルの精度を予測する方法を挙げる．精度の予測により，統合後の精度が向上するモデルが選択できるため，通信対象を限定した精度が最大化が可能となる．また，モデル統合においても，精度向上に効果的な統合方法を活用することで，モデルの精度をより少ない通信回数で最大化できると考える．

4.　提案手法

前章の課題 (1)，(2) を解決する手法として，ユーザの端末間での直接的な通信を活用した Federated Learning に基づくモデル構築の手法を提案する．

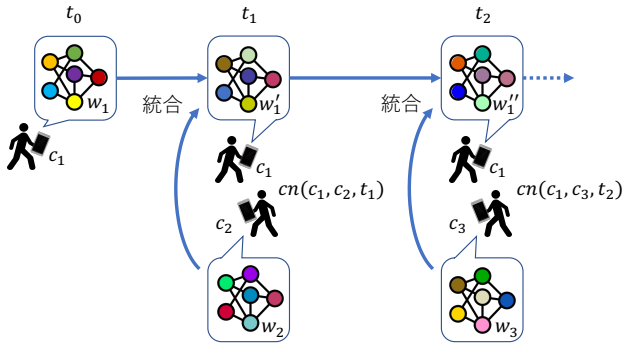

図 4　モデルパラメタ統合の概略図

4.1　モデルパラメタの更新処理

本稿で想定するモデルの統合の概略図を図 4 に示す．本提案手法では他の端末とコンタクトしたタイミングで端末間で通信し，モデルのパラメタを取得する．パラメタの取得後，自らのパラメタと Federated Learning に基づく方法で統合する．Federated Learning には主に FedSGD[2] と FedAvg[2] が提案されている．FedSGD では，データをモデルに入力した際の勾配の平均を計算しており，データ数に応じて集約サーバとの通信回数が増加する．一方で FedAvg は，各エッジ端末で訓練した後のモデルパラメ

表 2　モデル間の統合におけるパラメタの比率

	出力層以外	出力層
統合方法 (1)	1	0.5
統合方法 (2)	1	ラベル毎のデータ数の比
統合方法 (3)	0.5	0.5

タを集約し平均するため，FedSGD と比較して通信回数が少ない．そのため本研究では，通信コストの抑制を考慮し，FedAvg に基づくモデルパラメタの統合に着目する．図 4 より，c_1 に着目した場合，移動中に他の観光客の端末 $c_2, c_3, ...$ とコンタクトし，その際に相手のモデルパラメタ $w_2, w_3, ...$ を取得する．c_1 のモデルパラメタ w_1 と取得したパラメタを FedAvg に基づく方法で統合し，w_1 を更新して新たなパラメタ w_1' を得る．以降のコンタクトでは，統合後のパラメタ $w_1', w_1'', ...$ を自身のパラメタとして更新していく．他の端末のモデルパラメタの取得を通信可能なタイミングで実行することで，ネットワークの状態に依存しないモデル更新が可能となる．また，端末間の直接的な近距離通信を主に活用することで，広域ネットワーク内の遠距離通信で消費する電力の抑制ができる．

4.2　モデルパラメタの統合方法

本稿で提案するモデルパラメタの統合方法について述べる．提案する統合方法では，出力層と出力層以外の層で異なる割合または同様の比率での統合をしており，**表 2** に示す比率で 2 個のモデル間のパラメタを統合している．

また，統合方法の概要を**図 5** に示す．統合方法 (1) による統合では，出力層のパラメタの単純平均を計算し，その他の層はパラメタの和を計算している．統合方法 (2) による統合では，各出力クラスに対応するパラメタ (図 5 における紫，緑，赤のエッジの重み) に対して，出力クラスに応じたラベルのデータ数の加重平均で統合しており，その他の層は統合方法 (1) と同様に和を計算している．データ統合方法 (3) は FedAvg に最も近い統合方法であり，パラメタ全体の単純平均を計算している．

4.3　統合対象の選択方法

本研究の課題 2 に関して，統合後の精度予測によって，統合対象を精度向上に有効なモデルに限定し精度の最大化できると述べた．本節では，精度向上に有効なモデルを選択する手法について提案する．

統合対象の選択には，オブジェクト認識モデルとは別に，統合後の精度の変化 (向上または低下) を予測するモデルを使用する．統合対象の選択方法としては，端末間のコンタクト時に精度変化予測モデルを使用し，精度の向上が予測された場合，相手のモデルパラメタを統合対象として選択する．精度変化予測モデルの使用により，精度向上に有効なモデルの選択が可能となる．精度変化予測モデルの入力は，コンタクトした 2 端末間のオブジェクト認識モデルの精度や画像枚数などのデータであり，互いにこれらの情報を共有することで統合後の精度の変化を予測する．これらの精度や画像枚数などのデータの共有にも端末間で直接通信するが，モデルパラメタと比較し非常に小さいデータとなると予想されるため，データ共有時に消費する電力は考慮しないものとする．

5.　評価実験

前章で概説した提案手法に対して，独自の 14 層モデルと VGG16 の場合でそれぞれ評価実験を実施し，2 種類のモデルにおける各統合手法による統合後の accuracy の変化を評価した．実験では，10 クラスのオブジェクトを内包した Cifar10 データセットに関して 231 個のモデルを訓練し，モデル間で各統合手法を適用した際の accuracy を記録した．本章では，評価実験の内容及び評価方法について述べる．

5.1　初期モデルの生成

初期モデルの生成には，CNN モデルを使用しており，独自に構築した 14 層モデルと VGG16 の 2 種類を使用した．独自の 14 層モデルについては，9 層の畳み込み層，1 層の全結合層，4 層の最大プーリング層で構成されており，パラメタは Glorot の一様分布に基づいて初期化している．VGG16 については，全結合層を除いた層の fine tuning を行っており，全結合層は本来の 3 層 [9] から 1 層に変更した．そのため，VGG16 のモデルは 13 層の畳み込み層，1 層の全結合層，5 層の最大プーリング層で構成される．VGG16 の fine tuning では，ImageNet データセットを学習したパラメタを使用した．これら 2 種類のモデルに関して，最適化関数は確率的勾配降下法 (stochastic gradient descent)，損失関数は categorical crossentropy を使用して訓練を実施した．

本研究では，観光オブジェクトのデータを対象とするが，評価実験では Cifar10[*1] を使用して提案手法の評価を行った．Cifar10 の 60000 枚のデータセットの内 50000 枚を訓練用，残りの 10000 枚をテスト用に割り当てた．訓練用のデータから，ラベル毎にデータ数の多さが異なるデータセットを複数用意し，それぞれのデータセットを訓練したモデルを生成した．ラベル毎のデータ数の多さは，少数，中程度，多数で分けており，少数では 0〜1000 枚，中程度では 1001〜3000 枚，多数では 3001〜5000 枚の範囲でランダムにデータを割り当てている．これら 66 通りの各組み合わせに対して 3 パターンのデータセットを作成し，198 パターンのデータセットが作成されそれぞれを訓練したモデルを生成した．また，accuracy が著しく低いモデル

[*1] https://www.cs.toronto.edu/ kriz/cifar.html

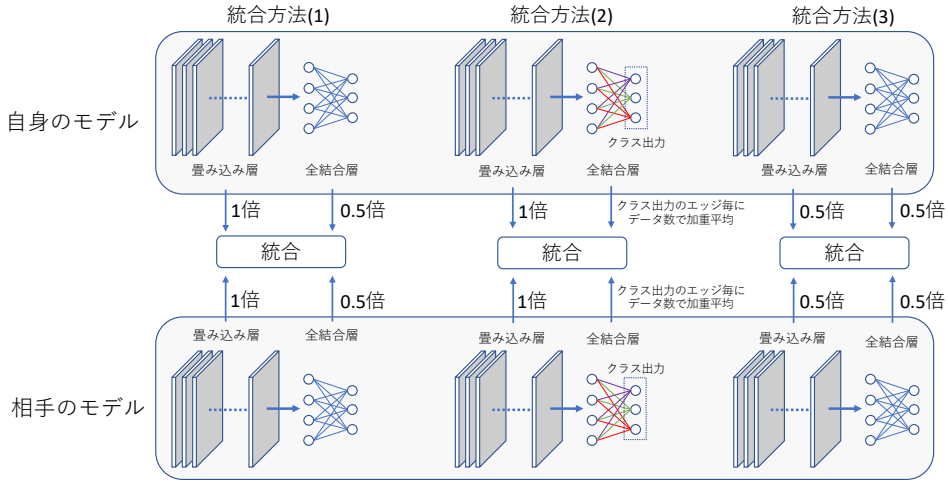

図 5　出力層と出力層以外の層のパラメタの統合方法

を統合した場合の評価のために，ラベル毎のデータ数の多さが 0〜10 枚と 11 枚〜100 枚の範囲となる組み合わせに対し全 33 パターンのデータセットを訓練したモデルも生成した．実験では，2 種類の CNN モデル (14 層独自モデルと VGG16) に対しこれらの 231 個のモデルを生成し統合を行った．各モデルが訓練したデータセット間には重複するデータも含まれる．各モデルの訓練にあたりエポック数は 100 回を上限としたが，訓練データに対する損失値が向上したエポックがある場合，そのモデルの訓練を終了するようにした．

5.2　実験方法と評価方法

図 6 に評価実験の概要を示す．実験では，複数のモデルから 1 個を選択し，自身のモデルも含め全モデルとの統合処理を実施した．この統合処理を全モデルで実施したため，53361 パターンのモデル統合を各統合手法で実施した．この評価実験を 14 層モデルと VGG16 に基づく CNN モデルで別々に行い，各パターンの統合モデルで cifar10 から抽出した 10000 枚のテストデータを分類した際の accuracy を記録した．

評価方法としては，統合前の自身のモデルと統合後のモデルのテストデータに対する accuracy の差分を求め，統合後で accuracy が向上したパターン数で評価した．

5.3　評価結果

評価実験の結果，**表 3** の通りに accuracy が向上したパターンが見つかった．表 3 より，統合方法 (1) と統合方法 (2) では，両モデルの場合で accuracy が向上したパターンが統合方法 (2) の方が多い結果となった．また，統合方法 (3) に関しては，14 層モデルについては評価結果の中では 3277 パターンと最も少なく，VGG16 の方では 37315 パターンと最も多い結果となった．VGG16 で統合方法 (3) を適用した場合に関して，accuracy が向上したパターンと低下したパターン，変化しなかったパターンの散布図を**図 7** に示す．図 7 において，ownAccuracy と otherAccuracy はそれぞれ自身のモデルとコンタクト相手のモデルのテストデータに対する accuracy である．図 7 では，コンタクト相手の accuracy が自身より高い場合に統合後の accuracy の向上が見られる．反対に，自身の accuracy より低いモデルを統合すると統合後の accuracy が低下する傾向にあるが，一部では向上したパターンも含まれる．

表 3　accuracy が向上したパターン数

	14 層モデル	VGG16
統合方法 (1)	11750	15442
統合方法 (2)	11960	15516
統合方法 (3)	3277	37315

6.　考察

本章では，評価実験で得られた結果と，図 7 の散布図についての考察を述べる．

6.1　accuracy が向上したパターン数のモデル毎の差

評価結果において，全体的に VGG16 のモデルで統合した方が accuracy が向上するパターン数が多い結果となった．モデルによってこのような差が生じた要因として，VGG16 の fine tuning が考えられる．VGG16 の方では，ImageNet の 1000 クラスのオブジェクトを訓練したパラメタを初期パラメタに使用しており，訓練前でもある程度特徴抽出が可能な状態であった．そのため，Cifar10 の訓練では，各モデルが同じ特徴を抽出するパラメタを各データセットに合わせて調整し，統合時には互いのパラメタの調整前と調整後の変更分が統合されたと考える．統合前のパラメタに対

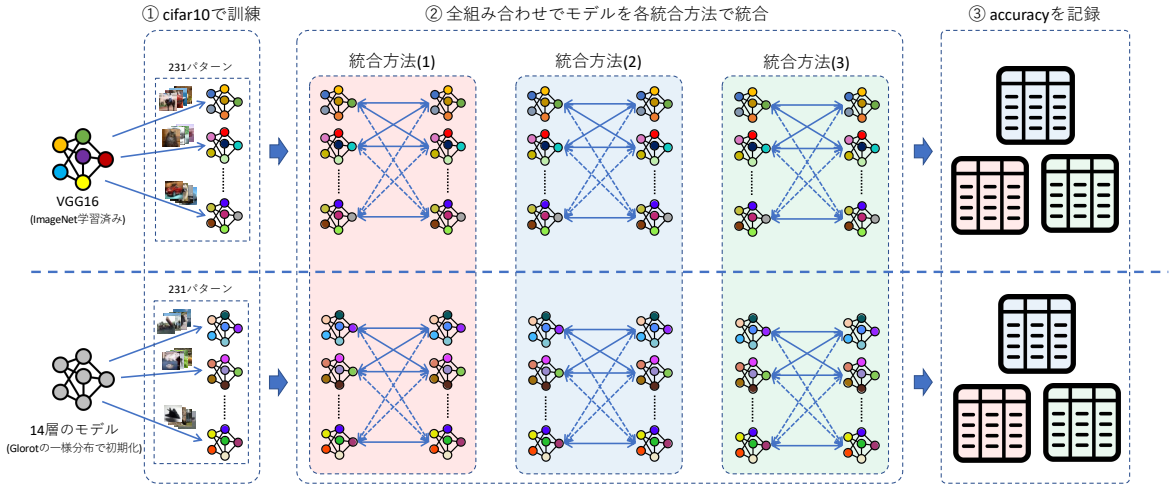

図 6　評価実験の概略図

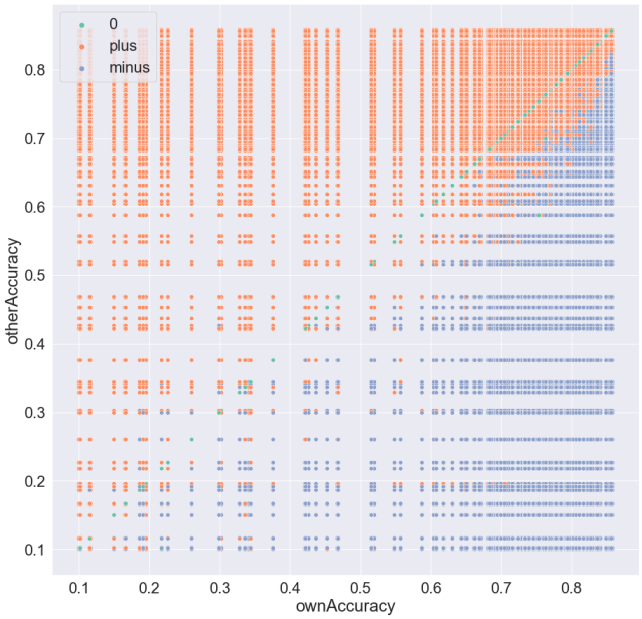

図 7　各モデル統合後の accuracy の変化

して，統合された変更分のみパラメタが変化するため，基盤となる特徴抽出に基づいた分類がされ accyracy の向上が多く見られたと考える．一方で，14 層モデルは Glorot の一様分布で初期化したため，各モデルの訓練でそれぞれのデータセットに基づいた独自の特徴抽出をするパラメタが生成される．このようなパラメタが統合されると，互いの特徴抽出の過程に悪影響を与え，分類に有効な特徴抽出が困難になると考える．

6.2　統合方法によるモデルパラメタへの影響

表 3 の評価結果より，統合方法 (1) と統合方法 (2) を比較すると，accuracy が向上したパターン数に変化が見られたが，全体の統合パターン数に対して小さい変化であったため，これら 2 つの方法についてはほぼ同等の結果となったと考える．これらの方法では，出力層となる全結合層の統合比率のみが異なっており，それ以外の層となる畳み込み層については同じ比率で統合していた．各モデルは畳み込み層が 9 層もしくは 13 層であるのに対し，全結合層は 1 層のみであったため，出力層の統合比率の変化による影響が少なかったと考える．

統合方法 (3) を適用した結果については，他の方法と比較すると 14 層モデルと VGG16 の間で accuarcy が向上したパターン数の差が最も大きく，評価結果における最小値と最大値が得られた．統合方法 (1) と比較して統合方法 (3) は出力層以外の層の統合比率のみが異なっているため，本稿の実験においては畳み込み層の統合比率が影響していると考える．統合方法 (1) では，パラメタ間で和を計算しており畳み込み層による空間方向の特徴抽出の効果が直接影響するが，統合方法 (3) では比率が 0.5 であり，それらの特徴が半減された状態で統合されたと考える．VGG16 では，前述のようにパラメタの差分が統合されているとすると，統合方法 (3) でも基盤となる特徴抽出には差分の平均だけ影響する．統合方法 (1) と比較して，畳み込み層における差分の影響が緩和された状態で統合されるので accuracy が向上するパターンが多くなったと考える，しかし，14 層モデルの統合では，独自の特徴抽出を行うパラメタが統合されており，互いの特徴抽出処理全体に影響を与える．統合方法 (3) により，相互のパラメタの抽出する特徴が半減される状態で統合されるため，有効な特徴抽出がより困難となり accuracy が向上するパターンが減少したと考える．

6.3　accuracy が向上するモデルパラメタの組み合わせ

図 7 では，統合手法 (3) で VGG16 を統合する場合について，モデルパラメタの組み合わせ毎の accuracy の変化を散布図で表した．この散布図において，自身のモデルよりも accuracy が高いモデルのパラメタを統合すると，

自身のモデルの accuracy も向上する傾向が見られた．また，自身のモデルの accuracy が高い程，相手のモデルの accuracy が自身と比較し低い場合でも accuracy の向上パターンが見られた．全体的に accuracy が向上しており，$Improve(M_c, M_{c'})$ が向上するパターンが多くなるため，式 (7) での最大化に有効であると考える．精度変化予測モデルの構築においても，図 7 の結果を訓練データとすれば，自身と相手の accuracy から統合後の accuracy を予測するモデルの構築が期待できる．

7. おわりに

本稿では，観光オブジェクト認識モデルのユーザ参加型構築に向けて，端末間での直接的な通信を活用した Federated Learning に基づくモデルの構築手法を提案した．手法の概要としては，2 人の観光客のコンタクトが発生した際に，互いの端末がモデルパラメタを通信及び統合し，統合後のパラメタを自身のモデルパラメタとして更新する手法である．また，最小限の通信でモデルの精度を最大化する統合方法及びモデルの検討のために，FedAvg に基づいた統合手法の特性を評価実験で網羅的に調査した．評価実験では，Cifar10 で構成される 231 パターンのデータセットを訓練した 2 種類の CNN モデルに対して，それぞれ 3 種類の方法で統合した．統合方法は，FedAvg に関連した方法であり，出力層とその他の層で統合する比率が異なる手法を 3 種類用意し評価実験を行った．評価方法としては，各モデルと統合方法の組み合わせ毎に，accuracy が向上したパターン数で評価した．結果としては，VGG16 で fine tuning を実施したモデルの全パラメタを単純平均した場合，53361 パターンのモデルの組み合わせの中から 37315 パターンで accuracy が向上し最大のパターン数を得た．この結果に関して，各統合後の accuracy の差分を散布図に表し，その傾向を図 7 として可視化した．この散布図より，自身のモデルに対して相手のモデルの accuracy が高い場合，統合後の accuracy が向上する傾向にあった．また，図 7 の結果を活用して，統合前のモデルの精度から統合後の精度の予測が可能となることが示唆された．

今後は，評価実験で得られた統合後の accuracy の変化の傾向をベースに，モデルパラメタ統合後の精度変化予測モデルの構築を検討する．更に，実際の観光オブジェクトに対する認識モデルをモバイル端末等に搭載し，提案手法に基づいた観光オブジェクト認識モデルの構築実験を行う．

謝辞 本研究成果の一部は国立研究開発法人情報通信研究機構 (NICT) の委託研究「スマートコミュニティを支える高信頼ネットワーク構成技術の研究開発」により得られたものです．

参考文献

[1] 国土交通省観光庁. AI(人工知能) 等導入による旅行サービスの高度化事業調査報告書. `https://www.mlit.go.jp/kankocho/content/001330607.pdf`, 2019. Accessed: 2021-03-12.

[2] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20 th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Vol. 54, , 2017.

[3] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, Vol. 10, No. 2, pp. 1–19, 2019.

[4] Sangsu Lee, Xi Zheng, Jie Hua, Haris Vikalo, and Christine Julien. Opportunistic federated learning: An exploration of egocentric collaboration for pervasive computing applications. *arXiv preprint arXiv:2103.13266*, 2021.

[5] Anusha Lalitha, Osman Cihan Kilinc, Tara Javidi, and Farinaz Koushanfar. Peer-to-peer federated learning on graphs. *arXiv preprint arXiv:1901.11173*, 2019.

[6] Mingzhe Chen, H Vincent Poor, Walid Saad, and Shuguang Cui. Wireless communications for collaborative federated learning in the internet of things. *arXiv preprint arXiv:2006.02499*, 2020.

[7] Hyesung Kim, Jihong Park, Mehdi Bennis, and Seong-Lyun Kim. Blockchained on-device federated learning. *IEEE Communications Letters*, Vol. 24, No. 6, pp. 1279–1283, 2019.

[8] 冨田周作, 中村優吾, 諏訪博彦, 安本慶一ほか. Federated learning over dtn によるオブジェクト認識モデルの地域間での共有手法の検討. 2020 年度 情報処理学会関西支部 支部大会 講演論文集, Vol. 2020, , 2020.

[9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.