希少性と応答コストに基づく生活行動アノテーション要求手法の検討

Examining the method of requesting activities of daily living annotation based on scarcity and response cost.

佐久間 隆友 † 松井 智一 † 諏訪 博彦 † 安本 慶一 †
Shigetomo Sakuma Tomokazu Matsui Hirohiko Suwa Keiichi Yasumoto

1. はじめに

近年，次世代の住宅としてあらゆる家電をネットワークに接続し，集中的に自動化･制御が可能なスマートホームが注目されている．しかしながらその利用は未だ限定的であり，省電力化やヘルスケア，セキュリティ監視にとどまっている．そのため最近では，スマートホームの利便性を上げるために，スマートホームに接続された IoT デバイスから様々な情報を収集・蓄積し，料理支援や個人に最適化された環境制御のような，よりコンテキストアウェアなサービスを提供することが求められている．スマートホームによるコンテキストの理解を実現するために，教師あり機械学習手法を用いて居住者の行動を分析し，構築された行動分類モデルを用いて認識を行う手法が主流である．

カメラや環境センサから得られる画像や照度といったデータに正解値を付与していくアノテーション作業は，モデルを構築するための教師あり学習で利用するデータセットの構築に非常に重要なプロセスであり，高品質であることが求められる．宅内行動認識分野においては，生活習慣やデータを取得する環境の差異に起因する個人差が大きいため，どの居住者にとっても適合する汎化性能を持つ生活行動認識モデルの構築が難しい．そこで居住者自身がその家庭のデータのアノテーションを行い，その環境に適合したモデルを構築することでその問題を解決できる．その際，その場でセンシングされたデータからなる，ラベル付けがなされてないデータセットのすべてをアノテーションできることが理想であるが，居住者が日常生活の中でアノテーションに使用できる可処分時間は限られており，応答コストの高いタイミングで要求を行ってしまうと，回答を拒否されるまたは適当に答えるなどデータセットの品質低下につながる．そのため，適切なアルゴリズムを用いて，要求価値の高いデータのみを取捨選択して居住者にアノテーションを依頼することが求められる．

そこで本提案では，宅内生活行動認識において，居住者自身が継続的に高品質なアノテーションを行うことを目的として，図 1 に示すような，ユーザの状況や価値を考慮する適切なアノテーション要求モデルを検討する．さらに，本提案の部分的な検証として，アノテーション要求が限定的なシナリオにおいて，希少性に限定してアノテーション要求のデータ選択を行い，ランダムサンプリングよりも効率的であることを示す．

本稿の構成は以下のとおりである．2 章では，宅内行動認識や介入戦略，能動学習などの既存研究を概説し，関連研究の課題を述べるとともに本提案の位置づけを述べる．3 章では本手法の問題設定やアルゴリズムについて述べ，4 章ではアノテーション要求に制限があるシナリオで希少性の考慮が有効であるかの検証を行う．5 章ではまとめと展望を述べる．

2. 関連研究

本章では，宅内生活行動認識や介入に関する研究，大規模言語モデルと能動学習分野の既存研究における課題を抽出し，本研究の位置づけを明らかにする．

2.1 宅内生活行動認識

居住者に異常検知や生活支援を提供することを目的として，部屋に設置されたカメラや，居住者自身が着用しているウェアラブルデバイスのデータを分析し，行動を認識する研究が行われている．Bianchi ら [1] は，ウェアラブルセンサの慣性データから，宅内で起こりうる 9 個の行動について，95%以上の精度で認識に成功している．データの収集は，ベルトにセンサーを装着した被験者が，一般的な宅内行動について，事前に設定された順序で行うことにより集められている．また，Annalisa ら [2] は，骨格と画像データを組み合わせ，機械学習手法を用いることによって多クラス行動認識を行い，98.8%の精度を達成している．訓練データおよび評価データは公開データセットである CAD-60 および CAD-120 を用いており，より一般的な宅内環境への拡張は，さらなる分析が必要と示されている．

宅内生活行動認識に用いられるデータセットは実験ベースで一時的に集められたものや，公開されているデータセットを利用しているものが多い．しかしながらこの手法は，継続的なアノテーションを居住者自身が行っておらず，その家庭に特化したモデルの長期的な構築が難しいといった課題がある．この課題を解決するために，ア

† 奈良先端科学技術大学院大学，Nara Institute of Science and Technology

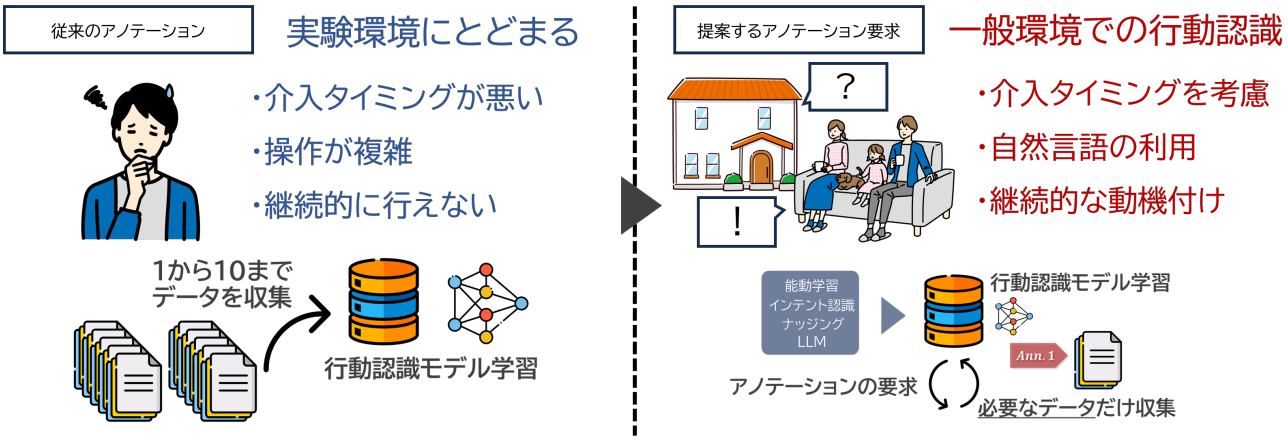

図 1: 提案手法概要

ノテーションの負荷の低減やインタフェースによる動機づけなど，実際の家庭で継続的に運用するためのユーザビリティの考慮が必要である．

2.2 介入に関する研究

Chen らの研究 [3] では，VR (Virtual Reality) コンテンツ中における通知の介入について，休憩中や動きの少ない動画の鑑賞中は，通知の受諾率が比較的高いことが示されている．さらに，VR ヘッドマウントディスプレイのセンサーから得られる体動や視線情報から，深層学習手法を用いて行動を認識し介入タイミングを制御している．

また，Züger らの研究 [4] では，実際の労働環境における労働者の介入可能性を機械学習手法を用いて予測している．これにより，介入可能性を同僚に示し，介入可能性が低い場合はインタラクションを避けるようにすることで，日々の中断によるストレスを最小限にできる．この手法により 75.3%の精度で介入可能性を予測することが可能であるが，評価のためのアノテーション要求のインタフェース自体が介入可能性に影響を及ぼすとも言及している．

Kobayashi ら [5] は，電子メールの配信を一時調停し，推定された介入可能性が一定の配信基準に届いた場合を見計らい，電子メールを配信するシステムを構築している．アプリケーションの切り替えやクリップボードの使用をモニタリングしながら，ルールベースで介入可能性を推定している．

従来の介入タイミングに関する研究では，特定のタスク遂行を目指すための行動変容や，通知介入タイミングの推定をユースケースとしており，実際にアノテーション要求インタフェースとして，行動ラベルの価値を定量化し，宅内行動中のどのタイミングに何を聞くことが望ましいのかを明らかにする研究は少ない．

2.3 能動学習に関する研究

アノテーションの戦略としては能動学習と呼ばれる分野で様々な手法が提案されている．能動学習は，モデルの高精度化に寄与する度合いの高いデータから順に学習する技術であり，アノテーション工数の削減や，学習データの作成コストの削減に寄与する Uncertainty Sampling [6] では，分類結果について，モデルの推論確率の低いデータから優先的にラベル付けを行うことにより，効率的に高精度化を行う手法である．実際に行動認識分野で能動学習を行っている例として，Hossain らの DeActive [7] がある．DeActive はデータ密度をベースに，ラベルが付けられた場合に現在の学習モデルに最も大きな変化をもたらすデータを選択している．その結果，能動学習手法が，手作業によるアノテーションの労力の削減に寄与し，行動認識の精度を劣化させずに学習速度を速めることを実証している．この手法は，本提案で多数のラベル付けされていないデータから，アノテーションするべきサンプルを選択し要求する際の，サンプリングアルゴリズムの設計の参考となる．

2.4 大規模言語モデル

現在，急速に進化している GPT-3 [8] をはじめとする大規模言語モデルは，文章全体の意味を捉えて自然言語を生成して応答することが可能であり，文章要約やチャットボットのような，自然言語を扱う分野での活用が期待されている．Ye らの研究 [9] では，ChatGPT を活用することによって，自然で直感的な人間とロボットのインタラクションを実現し，作業の信頼性を大幅に改善している．このように大規模言語モデルを活用することによって，効果的なヒューマンコンピュータインタラクションが実現できる．

2.5 本研究の位置づけ

関連研究で，特定の環境やタスクに対する行動認識や介入タイミングの研究が行われている．一方で，各家庭に適合する宅内行動認識モデルの構築や，アノテーションを目的とする介入可能性の推定が行われていない点が問題点として挙げられる．そうした問題点を踏まえ，本研究では，生活行動認識におけるラベルの希少性や応答コストを定義し，それらが人間の感性に即したものか，日常生活行動中の適した介入タイミングかを明らかにする．

3. 提案手法

本章では，まず提案システムの概要について述べ，要件定義を行う．続いて要求価値を定式化した上で，要求価値の要素となるラベル希少性と応答コストについて定義し，要求を最適化する手法について検討する．最後に，アノテーション要求を行うインタフェースについても検討を行う．

3.1 提案手法の概要

提案するアノテーション要求インタフェースの全体図を図 2 に示す．図上部に示すように，従来の宅内行動認識では，居住者がアプリなどを用いてセンサデータに対応する行動をアノテーションし，あらかじめ汎化されたモデルに対し転移学習を適用することによって家庭ごとに適合するモデルを作成していたが，アノテーションの継続的な動機付けは困難であった．そこで図下部に示すように初期から適合させたい家庭のデータを収集し，その居住者自身がアノテーションを継続的に行うことによってその家庭に特化した行動認識モデルを構築するアノテーション要求インタフェースを提案する．このアノテーション要求インタフェースは，要求価値率モデル，要求最適化，会話による介入，フィードバック機構の 4 要素から構成される．

適切なアノテーション要求を行うためには，優先順位付けを行うために要求価値を推定するモデルの構築を行う必要がある．要求価値を定量化し実際にアノテーション要求を行う際，連続した応答はユーザの認知資源を消費する．認知資源が枯渇するとアノテーションの品質低下を引き起こすため，枯渇しない程度に最大限要求を行うための要求最適化を行う．この要求は会話によって行われる．さらに居住者の反応に応じて介入アルゴリズムを更新するフィードバック機構を作成することによって，最終的には個人に適合した要求インタフェースの構築を目指している．

本稿はその中でも要求価値率モデルと要求最適化にフォーカスする．従来のデータセット構築におけるアノテーションは行動する度にアプリなどによるアノテーションが必要であった．継続的に行うためにはアノテーション要求の頻度を下げる必要があるが，そのためには要求価値推定モデルを使い要求に優先順位をつける必要がある．また，その際ユーザの認知資源の消費を最小限にすることが求められる．さらに，様々な家庭の情報を収集し，誰でも参加できるインタフェースであることが重要であるため，簡易に扱えるインタフェースの構築が課題である．以上から，提案するシステムは，以下の要件を満たすように設計される．

(要件 1) 宅内行動認識におけるラベル希少性及び応答コストを定量化し，要求価値率推定モデルを構築すること．

(要件 2) ユーザの認知資源を最小限にするような最適化手法を構築すること．

(要件 3) アノテーション要求および応答のインタフェースは一般の居住者が簡易に扱えるものであること．

3.2 要求価値の定式化

アノテーション要求数に制限がある中で，適切なタイミングかつ精度を最大化するためのアノテーション要求を行うためには，ラベル希少性と応答コストを見積り定量化する必要がある．一例として，以下のようなアノテーション要求モデル RV を定義する．

$$RV_{req} = \text{ラベル希少性 } LS_{req} \times \text{応答コスト } AC_{req} \quad (1)$$

ここで，LS と AC はそれぞれラベル希少性と応答コストを示しており，req はベースとなるモデルから得られる推論結果が不確実なデータであり，アノテーション要求の候補としてプールされる．実際のシステムでは，この RV が高い req を優先的にアノテーション要求の対象とする．

3.3 ラベル希少性の定義

宅内行動認識でよく行われるマルチクラスの分類タスクにおいて，不平衡データを深層学習の入力とすると，少数派のクラスは十分なサンプル数を確保できないため，モデルは多数派のクラスに過剰に適合してしまう問題がある．そのため，サンプリング方法の工夫やアルゴリズムの工夫により，データの不均衡性を是正したり，不均衡学習を効率的に行うことが一般的である．アノテーション要求においても，可能な限り平衡データになるように，データセット中のラベル割合が少ないものを希少性が高いとし，優先的に要求することで効率的なアノテーションが実現できる．

また，希少性が高いクラスのクラスタの中でも，疑似ラベル付けを行った際の疑似ラベルの確信度を考慮する

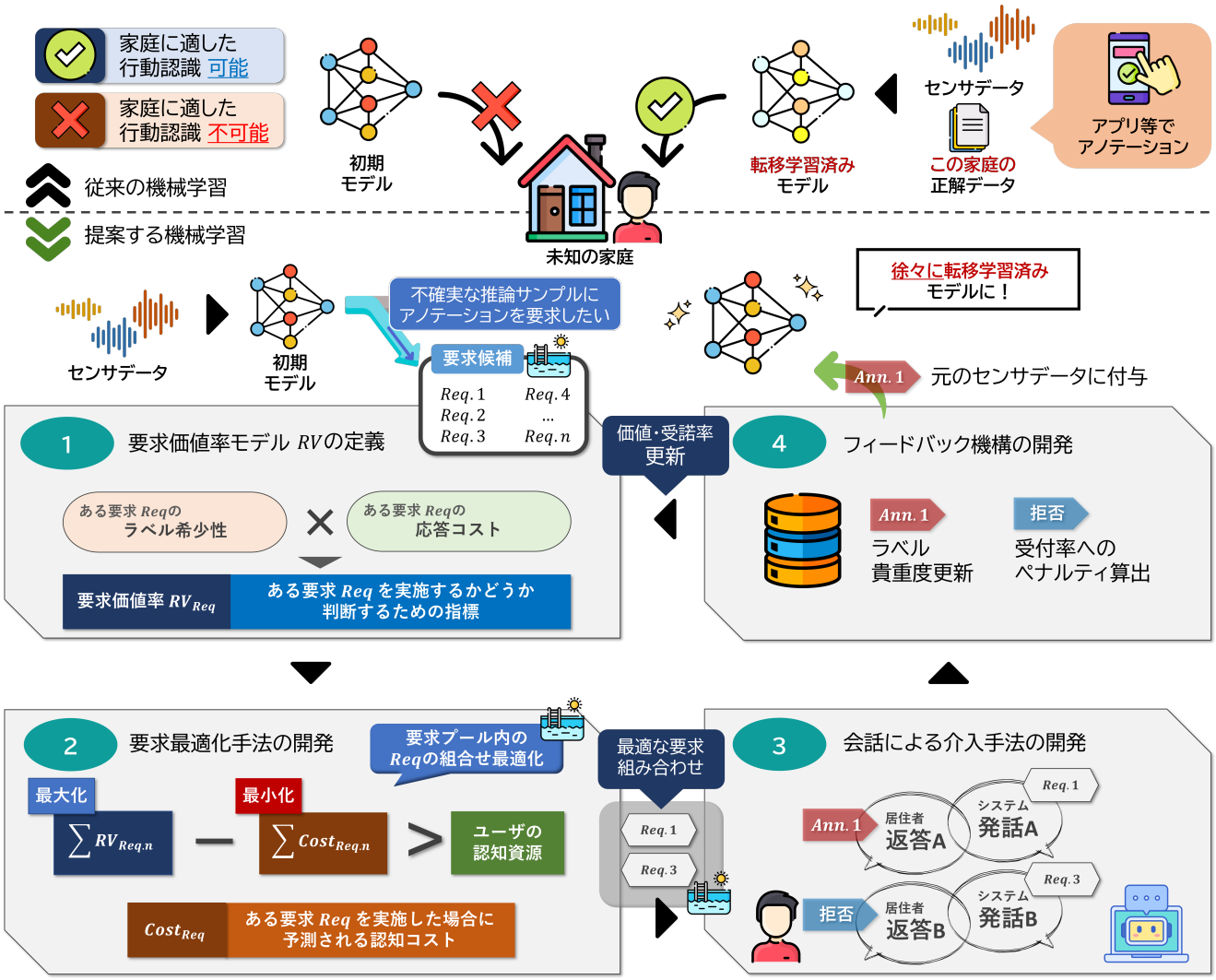

図 2: アノテーション要求インタフェースの全体図

必要がある．確信度が低いデータはデータセットの中に類似例が少ないため，優先的にアノテーション要求することによって，全体の精度向上につながる．具体的には，前述の Uncertainty Sampling [6] という能動学習の一手法を用いる．Uncertainty Sampling では，サンプル x_i を y に分類する確率 $P(y|x)$ について，モデルの推論確率のないサンプル x^* を選択する手法であり，式 2 で表される．

$$x^* = \arg\max(1 - P(y|x)) \tag{2}$$

マルチラベル分類タスクにおいて，確信度は一般的に活性化関数である Softmax から得られる．しかし Softmax 関数から得られる確信度は実際の正解率に準じてはいない可能性があり，Temperature Scaling などの手法を用いて校正する必要性がある [10]．校正後，出力される疑似ラベルの調整済み確信度を希少性として定義する．

3.4 応答コストの定義

居住者に実際の生活の中でアノテーションを要求するうえで，応答コストにおいても考慮する必要がある．応答コストは，ユーザへアノテーションを要求した際に受諾する度合いであり，認知負荷量も含まれる．認知負荷の例として以下のアノテーション要求を考える．

(1) 「現在何をしていますか？」

(2) 「現在は料理をしていますか？」

ここで，(1) の質問は正解値を直接尋ねることが可能であるが，自由形式の回答を行わなければならず，(2) の質問のような「はい」「いいえ」で回答する確認クエリよりも認知負荷が高くなる．Wu らが行った調査 [11] では，12 人の対象者の内 1 人を除いた全員があまりにも多くの自由形式の質問に苛立ちを感じると回答している．また，応答コストのもう一つの要素として，ユーザの応

答可能性についても考慮する．想定される応答可能性が低いケースとして，朝方に遅刻しそうな際にアノテーションを要求しても，拒否もしくは適当な回答をする恐れがある．このようなユーザビリティを考慮することは，データセットの品質を高めるとともに，継続的なインタラクションを行うために重要である．具体的に応答可能性を定量化する手法として，カメラやその他センサから画像などの情報を取得し，深層学習手法を用いてユーザの応答可能性を予測する手法が考えられる．また，応答コストが高い場合はその行動が行われる時ではなく事後的にアノテーションを要求することも考えられる．そのようなケースでは経過時間に対する忘却率を適用し，相応のペナルティを与えるといったような工夫をすることによって，信頼性の低い不正確なアノテーションに対してもロバストな要求価値予測モデルを構築することが可能である．

3.5 要求最適化手法の開発

要求を連続して行うことは，居住者の認知資源を消費する．認知資源が枯渇すると居住者はストレスを感じ，結果としてアノテーション品質の低下を引き起こし，継続的なアノテーションが行われなくなる．したがって，ある要求 req に対して，要求価値率 RV と応答コストを考慮した最適化手法の開発が必要である．具体的には，ある期間における居住者の認知資源を $Budget$，要求によって消費するコストを $Cost$ としてそれぞれ定義する．式 3 のように，$Buget$ が枯渇しない範囲で要求価値の最大化と $Cost$ の最小化を行う．最終的には，実際に問い合わせる要求データの組み合わせを，動的計画法や強化学習といった手法により最適化する．

$$\sum RV_{req \cdot n} - \sum Cost_{req \cdot n} > Budget \tag{3}$$

そのように最適化されたモデルを再利用してアノテーション価値を再計算し，再度アノテーションを要求する．この一連のサイクルを繰り返すことによって，局所的な環境に適合した宅内行動認識モデル及びアノテーション要求モデルを構築できる．本手法の評価には，アンケートなどを用いてアノテーションへの動機付けが適切に行われたかを明らかにしたうえで，ベースラインと比較し，実際の行動認識タスクに対して高い精度を得られたことを accuracy や F1-Score のような分類タスクにおける評価指標を用いて行う．

3.6 アノテーション要求インタフェースの検討

実際にアノテーション要求を行うインタフェースとして，自然言語を用いた対話形式が有効である．実際に Alexa をはじめとする音声アシスタントは誰もが扱えるスマートスピーカーとして普及しつつある．また，自然言語の発話によるアノテーションは，場所依存性が低く，スマートホンのアプリを操作するよりも簡易で子供や高齢者でも扱いやすい．

しかし問いかけに対して応答をすることは認知負荷が高いため，将来的には ChatGPT をはじめとする大規模言語モデルや行動経済学におけるナッジング [12] を用いて，日常会話の中から無意識的にアノテーションを実施させることが望ましい．

4. 希少性に基づく要求データ選択の検証

本提案の実現可能性を検証するために，提案に対する予備実験を行った．予備実験ではラベル付けされていないデータに対してアノテーション要求数が制限されているシナリオで，希少性を考慮したアルゴリズムが有効かを評価するために，応答コストやアノテーション回答の信頼度を考慮せずに単純化している．

希少性がどの程度アノテーション価値として扱えるのか，実際に深層学習モデルを構築しながら，アルゴリズムを用いない場合と比較して検証する．

4.1 実験条件

実際の家庭でセンシング及びアノテーションされた宅内行動認識データセット [13] を使用して，本提案手法における希少性に基づくアノテーション要求のデータ選択が，ランダムサンプリングモデルよりも有効であるか検証を行った．データセットの詳細を表 1 に示す．データセットは実際に家庭で入浴・調理・食事・外出・就寝の 5 つの行動について 1 か月間記録されたものであり，人感センサや環境センサ，またドアセンサの情報に対して，5 つの行動のいずれかがラベル付けされている．

4.2 行動認識モデルの構築

データの分割方法を図 3 に，使用したネットワークを図 4 にそれぞれ示す．

検証では，2 週間分のデータに加え，ランダムサンプリング・希少性ベースのサンプリングそれぞれの戦略でアノテーション要求されたものをラベル有り，されなかったものをラベル無しとして，半教師あり学習を行った．希少性ベースのサンプリングでは，シャッフルされた 3 週間目のデータから Softmax の出力による確信度の低い順に 10%データを取り出し，ランダムサンプリングでは，データを重複なしで同数ランダムで抽出し，それぞれアノテーションを行っている．半教師あり学習を行うためにはモデル仮定が必要であり，様々な手法があるが今回は Self-Training [14] を使用した．Self-Training は，まずラベル付けされたデータのみを用いて教師あり学習を行い，生成されたモデルでラベル付けされてないデータを推論し，その結果を疑似的な正解値としてラベリングし，再学習する手法である．Self-Training で構築され

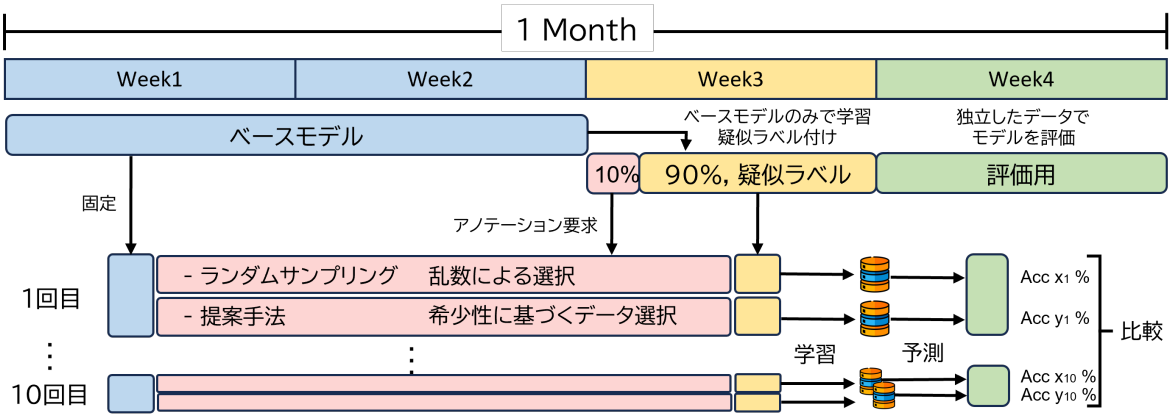

図 3: データの分割方法

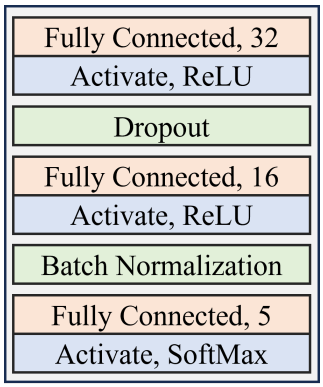

図 4: 使用したネットワーク

表 1: データセット詳細

データ種別	個数	特徴量	期間
既知データ	2045	80	1,2 週間目
未知データ	1120	80	3 週間目
評価データ	1225	80	4 週間目

たモデルは，独立したデータで評価し二つの戦略を比較した．

4.3 評価方法

1 か月間収集されたデータセットを最初の 2 週間は既知のデータセットとして，3 週間目のラベル付けされていないデータセットの中から少量のみアノテーション要求としてラベル付けを行った．そうして構築されたデータセットを，半教師あり学習を用いてモデルを構築し，学習に使用していない 4 週間目のデータを用いてモデルを評価した．未知のデータに対する行動認識の精度を評価するため，深層学習の入力となる訓練・評価・テストデータはホールドアウト法を用いて分割した．また，ランダムサンプリングの乱数の偏りによる影響を抑えるため，ランダムサンプリングおよび検証モデルそれぞれ乱数のシードを重複なしで変更しながら accuracy の平均を 10 回求め比較した．

表 2: 検証結果

試行回数	ランダムサンプリング [%]	提案手法 [%]
1	51.4	56.6
2	58.7	58.5
3	56.2	58.0
4	49.7	58.0
5	49.6	53.7
6	53.5	49.0
7	52.3	55.1
8	54.4	46.5
9	55.1	62.9
10	45.5	60.3
平均値	52.6	55.9

4.4 検証結果

検証結果を表 2 に示す．10 回の平均を求めた結果，ランダムサンプリングによってアノテーションされたモデルの accuracy は 52.6%，対して希少性を考慮したモデルの accuracy は 55.9%の精度を示した．これにより，希少性を考慮するモデルはランダムサンプリングと比較し，約 3pt の精度の改善を確認した．ランダムサンプリングよりも精度を改善したことから，希少性を考慮したデータ選択がアノテーション要求戦略に有効であることが確認できた．将来的には，応答コストを求めるための行動認識や，要求価値率モデルに基づく介入のための応答インタフェースを設計し，実際の居住環境で検証することによってアノテーション要求が適切であるか評価する実

験を行うことにより本提案を実証する．

5. おわりに

本提案では，宅内行動認識におけるアノテーション要求の最適化を目的として，希少性と応答コストに基づくアノテーション要求モデルの検討および一部検証を行った．検証の結果，希少性に基づくデータ選択が，ランダムサンプリングと比較して 3pt 改善し，希少性の考慮が有効であることを確認した．これにより実現されるシステムによって，一般家庭における継続なアノテーションを実現できる．今回は希少性のみを考慮した実験により有効性を示したが，応答コストも考慮しかつ，実際の家庭環境で実証する必要がある．

謝辞

本研究は JSPS 特別研究員奨励費 JP22KJ2287 の助成によって行うものである．

参考文献

[1] Valentina Bianchi, Marco Bassoli, Gianfranco Lombardo, Paolo Fornacciari, Monica Mordonini, and Ilaria De Munari. Iot wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment. *IEEE Internet of Things Journal*, Vol. 6, No. 5, pp. 8553–8562, 2019.

[2] Annalisa Franco, Antonio Magnani, and Dario Maio. A multimodal approach for human activity recognition based on skeleton and rgb data. *Pattern Recognition Letters*, Vol. 131, pp. 293–299, 2020.

[3] Liwei Chan Kuan-Wen Chen, Yung-Ju Chang. Predicting Opportune Moments to Deliver Notifications in Virtual Reality. In *ACM Conference on Human Factors in Computing Systems 2022*, 2022.

[4] Manuela Züger, Sebastian C Müller, André N Meyer, and Thomas Fritz. Sensing interruptibility in the office: A field study on the use of biometric and computer interaction sensors. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–14, 2018.

[5] Yasumasa Kobayashi, Yuichiro Fujimoto, and Kinya Fujita. Development of e-mail delivery mediation system based on interruptibility and its evaluation in daily office work scenario. *IEEE Access*, Vol. 7, pp. 94084–94096, 2019.

[6] David D Lewis and Jason Catlett. Heterogeneous uncertainty sampling for supervised learning. In *Machine learning proceedings 1994*, pp. 148–156. Elsevier, 1994.

[7] H. M. Sajjad Hossain, MD Abdullah Al Haiz Khan, and Nirmalya Roy. Deactive: Scaling activity recognition with active deep learning. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 2, No. 2, jul 2018.

[8] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, Vol. 33, pp. 1877–1901, 2020.

[9] Yang Ye, Hengxu You, and Jing Du. Improved trust in human-robot collaboration with chatgpt. *IEEE Access*, 2023.

[10] Yu Sun Kilian Q. Weinberger Chuan Guo, Geoff Pleiss. On Calibration of Modern Neural Networks. In *Proceedings of the 34 th International Conference on Machine Learning*, pp. 1321–1330, 2017.

[11] Jason Wu, Chris Harrison, Jeffrey P Bigham, and Gierad Laput. Automated class discovery and one-shot interactions for acoustic activity recognition. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020.

[12] 佐藤佑磨, 松井智一, 中村優吾, 諏訪博彦, 安本慶一ほか. 宅内行動認識モデル最適化のためのナッジを用いたアノテーション行動誘導方法の検討. マルチメディア, 分散協調とモバイルシンポジウム 2021 論文集, Vol. 2021, No. 1, pp. 11–19, 2021.

[13] Tomokazu Matsui, Kosei Onishi, Shinya Misaki, Manato Fujimoto, Hirohiko Suwa, and Keiichi Yasumoto. Salon: Simplified sensing system for activity of daily living in ordinary home. *Sensors*, Vol. 20, No. 17, p. 4895, 2020.

[14] Xiaojin Jerry Zhu. Semi-supervised learning literature survey. 2005.