一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
SC2018-38 (2019-03)

# コグニティブAPIによる特徴量を用いた宅内コンテキスト認識手法の提案

陳　　思楠† 　佐伯　幸郎† 　中村　匡秀†,††

† 神戸大学　〒657–8501 神戸市灘区六甲台町 1–1
†† 理化学研究所・革新知能統合研究センター　〒103-0027 東京都中央区日本橋 1-4-1
E-mail: †chensinan@ws.cs.kobe-u.ac.jp, ††sachio@carp.kobe-u.ac.jp, †††masa-n@cs.kobe-u.ac.jp

**あらまし**　近年，深層学習技術の発展によって，マルチメディアデータを活用したコンテキスト認識技術が有望視されている．我々は，この技術をスマートホームにおける宅内のコンテキスト認識に使うことを考える．一般に宅内コンテキスト認識では，世帯ごとに部屋のレイアウトや環境，認識したいコンテキストの要求が異なるため，世帯ごとに独自の認識モデルを構築する必要がある．このとき，深層学習を利用するアプローチでは，非常に多くの訓練データが必要になり，一般家庭で実施するのは事実上不可能である．本研究の目的は，一般家庭で導入可能な画像を活用したコンテキスト認識手法を開発することである．提案手法では，汎用的な画像認識を行うコグニティブ API を活用し，画像内に認識される情報をテキストとして取り出す．これを特徴量として，通常の教師あり機械学習にかけることで，コンテキストを分類する．深層学習を用いるアプローチに比べて，提案手法では，コグニティブ API が提供する汎用的な画像認識と軽量な機械学習を利用する．これにより，各世帯に特化したコンテキスト認識をはるかに少ないエフォートで実現できる．

**キーワード**　コンテキスト認識，画像，コグニティブ API，機械学習

# Proposal of home context recognition method using feature values of cognitive API

Sinan CHEN†, Sachio SAIKI†, and Masahide NAKAMURA†,††

† Kobe University　Rokkodai-cho 1–1，Nada-ku，Kobe，Hyogo 657–8501 Japan
†† Riken AIP　1-4-1 Nihon-bashi, Chuo-ku, Tokyo 103-0027 Japan
E-mail: †chensinan@ws.cs.kobe-u.ac.jp, ††sachio@carp.kobe-u.ac.jp, †††masa-n@cs.kobe-u.ac.jp

**Abstract**　The emerging deep learning technology is a promising means for context recognition with multimedia data. We are interested in using the deep learning with images for context recognition in smart homes. In the home context recognition, the room layout, the environment, and the contexts to be recognized are different from one household to another. Therefore, a unique recognition model is required for every different household. For this, if we take a naive approach that uses the deep learning directly, a huge amount of labeled images are required, which is practically impossible for general households. The goal of this research is to develop an image-based context recognition method that is affordable at home. In the proposed method, we exploit a cognitive API which performs general image recognition, and retrieve the information within the image as text. By using the text as features, we classify the context with ordinal supervised machine learning. Compared with the expensive approach with deep learning, the proposed method uses generic image recognition of the cognitive API, and light-weight machine learning. As a result, the context recognition customized for every household can be achieved with much less effort.

**Key words**　Context recognition, Image, Cognitive API, Machine learning

## 1. はじめに

IoT（Internet of Things）技術の急速な発展に伴い，物理空間の様々な情報を収集し，付加価値サービスに活用することが可能になっている．スマートホームの分野では，宅内の居住者や環境についての様々な状況（総じて**宅内コンテキスト**と呼ぶ）を認識する研究が盛んにおこなわれている．宅内コンテキストの例としては，食事をしている，寝ている，テレビを見ている，本を読んでいる等といった居住者の日常生活行動によって定義される状況や，電気が消えている，部屋に誰もいない，部屋が散らかっている等の宅内の環境状態によって定義される状況等が挙げられる．

従来の宅内コンテキスト認識では，建物内あるいは身体に装着されたセンサや，家電から得られる数値データを使うこと主流であった．例えば，家電の消費電力と居住者の位置から日常行動を認識する研究 [1] や，スマートフォンのセンサを用いて同様の認識を行う研究 [2] が存在する．また，宅内の温度や湿度，照度などの環境変化値の測定から，宅内の環境状態を学習・推定する研究もある [3].

その一方で，近年，深層学習（Deep Learning）の発展によって，画像や音声，動画，テキストなどの**マルチメディアデータ**の学習・認識技術が格段に進歩している．そのため，我々はマルチメディアデータ（特に**画像データ**）を活用した新しい宅内コンテキスト認識手法の開発に興味をもって取り組んでいる [4] [5].

一般的に，画像による機械学習によって宅内コンテキストを認識しようとする場合，**世帯ごとの個別の違い**が問題となってくる．部屋のレイアウトや存在するオブジェクト，環境の状態は，世帯ごとに異なるため，同じコンテキスト（例えば食事をしている）であっても，画像に映る情報は大きく異なる．また，システムに認識させたいコンテキストも世帯ごとに異なる．よって，**世帯ごとに個別の認識モデル**を構築する必要がある．単純なアプローチとして，宅内で収集する画像を直接深層学習にかけて，高精度な認識モデルを構築する方法が考えられる．しかしながら，このアプローチは非常に多くの訓練データと強力なマシンパワーを必要とするため，個々の一般家庭で実施するのは現実的ではない．

そこで本研究では，画像を活用した宅内コンテキスト認識を一般家庭で実施可能にすることを目的とする．そのためのアプローチとして，まず一般家庭における**宅内コンテキスト認識のための機械学習フレームワーク**を提案する．このフレームワークでは，ユーザが自宅で任意のコンテキストを定義し，データを収集し，機械学習を用いて，コンテキストを認識するための汎用的な枠組みを提案する．フレームワークでは，データの種類や機械学習アルゴリズムは規定せず，深層学習を適用することも可能である．

次に，上記のフレームワークの一実装として，**コグニティブ** *API* による特徴量を活用した，新たな宅内コンテキスト認識手法を提案する．コグニティブ API とは，画像や音声等のマルチメディアデータを高度に認識するクラウドサービスの API (Application Programming Interface) である．提案手法では，宅内で収集した画像を汎用的な画像認識 API に送信し，API が認識する画像に含まれる情報（**タグ集合**）を抽出する．次に，抽出した全てのタグ集合に対して，テキストマイニングを行い，各画像をベクトル化する．最後に，これらのベクトルを軽量な教師あり機械学習にかけ，多値分類モデルを構築する．提案手法では，深層学習を利用せずに軽量な機械学習を用いるため，各家庭に特化したコンテキスト認識モデルをはるかに少ないエフォートで実現できる．

提案手法の有効性を評価するため，研究室内の画像を収集して，コンテキストを認識する実験を行った．研究内に定点カメラを設置し，5 秒間隔で画像データを 2 週間撮影・蓄積した．認識したいコンテキストとして，全体ミーティング，掃除，食事，無人，個別ミーティング，遊び，仕事の 7 つを定義した．各コンテキストに対して，蓄積した画像からそのコンテキストに対応する 100 枚を選別してラベル付けした．選別した 700 枚の画像を *Microsoft Azure Computer Vision API* [6] に送信し，API 認識結果からタグ集合を抽出した．抽出したタグ集合を文書とみなし，*TF-IDF (Term Frequency - Inverse Document Frequency)* [7] 法を用いてベクトル化した．最後に，ベクトル化したタグ集合と対応するコンテキストのラベルを *Microsoft Azure Machine Learning Studio* [8] に導入し，Multiclass Neural Network によって認識モデルを構築した．

評価実験の結果に関して，本研究で構築した認識モデルの全体精度 (Overall accuracy) は 0.929 であり，平均精度 (Average Accuacy) は 0.980 という結果になった．そして，各コンテキストのテストデータのうち，正しく認識できたデータの割合 (Micro-averaged precision) は 0.929 であり，全体のテストデータのうち，正しく認識できたデータの割合 (Macro-averaged precision) は約 0.924 であった．混同行列（Confusion Matrix）を用いた結果によって，全体ミーティングの認識精度は 95.3%，掃除は 90.9%，食事は 83.3%，無人は 100.0%，個別ミーティングは 96.0%，遊びは 82.2%，仕事は 100.0% となった．7 つのコンテキストのうち，無人と仕事の認識精度が最も高く，食事と遊びの認識精度が低いことがわかった．

## 2. 準　　備

### 2.1 宅内コンテキストの認識

宅内コンテキストとは，宅内の居住者や環境に関するあらゆる**状況情報**を指す．宅内が現在どんな状況にあるかは，提供すべきサービスの内容や提供タイミングに対して重要な意味を持つ．よって，宅内コンテキストをいかに精度良く認識するか，コンテキストに応じたサービスをいかに提供するか（**コンテキスト・アウェアサービス**) は重要な研究課題とされ，特にユビキタス・コンピューティングの分野で長年研究されてきた．

ユビキタス・コンピューティングでは，ウェアラブルセンサや環境センサ，スマートフォン等から得られる数値データを利用して，コンテキストを認識する研究が主流であった．具体的な例として，ユーザ位置情報と家電消費電力に基づいた宅内生活行動認識システム [1] や，スマートフォンを用いた生活行動認識 [2]， 宅内の環境変化と声掛けに基づく在宅高齢者の日常

生活行動センシングシステム [3] などが存在する．これらは各種センサから得られる数値データに対して，ルールや機械学習を適用して，居住者の行動や宅内の状態を推定・判別する．残念ながら従来研究の多くは，専用のセンサを必要とすることや運用の複雑さから，一般家庭に広く普及するには至っていない．

宅内コンテキスト認識は，昨今実用化が著しいスマートホームでの応用が期待される．代表的な応用例として，独居高齢者の見守りや居住者の日常生活リズムの維持向上などがある．

### 2.2 画像データを用いた宅内コンテキスト認識

昨今，Web カメラを代表とするカメラデバイスは，低廉化や小型化によって，一般家庭にも容易に導入・設置できるようになっている．また，カメラから得られる画像データの情報量は，センサの数値データの情報量に比べて大きい．これらのことから，カメラ画像を用いることで，より強力で導入が簡単な宅内コンテキスト認識を実現できる可能性がある．

一般に画像データを認識・理解するためには，高度な画像認識技術が必要であるが，昨今の深層学習の発展・普及によって，実用に耐えうる十分な精度の認識が可能になっている．しかしながら，深層学習を用いて，各家庭に特化したコンテキスト認識モデルを 0 から構築することは，訓練データの準備や計算リソースの観点から，現実的ではない．

### 2.3 画像に基づくコグニティブサービス

コグニティブサービスは，画像や音声，テキストといったマルチメディアデータを認識するクラウドサービスである．一般的にはクラウドの豊富な計算資源を生かして構築された学習済み機械学習モデルによって実装されている．コグニティブ API は，コグニティブサービスを外部アプリケーションから呼び出して利用するための API である．これによって，大規模で複雑な認識処理をアプリケーションに容易に組み込める．

画像を認識するコグニティブサービスは，与えられた画像を解析して，様々な情報を認識・抽出して返す．有名なサービスとして，Microsoft Azure Computer Vision, IBM Watson, Google Cloud Vision, Amazon Rekognition 等が存在する．認識・抽出される情報には，顔，年齢，性別，髪型，物体，テキスト，背景，カテゴリ，場所，色等がある．

### 2.4 先行研究：画像に基づくコグニティブ API の宅内センシングへの適用可能性 [4] [5]

我々は先行研究 [4] [5] において，商用のコグニティブ API によって宅内コンテキストの認識が実現できるかを考察した．具体的には，研究室に定点カメラを設置し，ミーティングや食事，遊び等のコンテキストが映り込んだ画像を取得して API に送信した．そこから抽出された情報をによってコンテキストを推定できるかを調べた．実験では，Microsoft Azure，IBM Watson，Google の 3 種類の画像認識 API を利用し，各 API が出力した画像を説明する単語の集合（**タグ集合**）を分析した．

分析では，タグ集合がオリジナルのコンテキストを反映しているか，同じコンテキストの凝集，異なるコンテキストの分離が可能かを，文書間類似度によって評価した．その結果，API が出力するタグ情報を，そのまま研究室のコンテキスト認識に適用しても，十分な性能が得られないことが分かった．

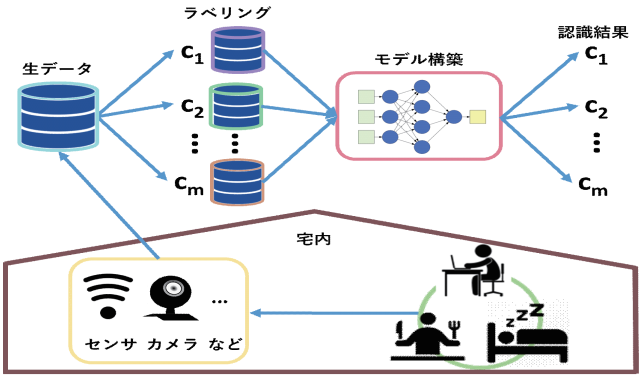


図 1　機械学習フレームワークの全体の流れ

### 2.5 機械学習プラットフォーム

ユーザが自分の用途に応じた機械学習モデルを，クラウド上に構築，配備できるプラットフォームが登場している．ユーザは自らの目的に合わせてデータをアップロードし，多種多様なアルゴリズムを組み合わせて独自の機械学習モデルを実験・構築できる．例えば，Microsoft Azure Machine Learning Studio, Google Cloud Machine Learning Engine [9]，Amazon SageMaker [10] 等が知られている．本稿では主に Azure Machine Learning Studio を活用する．

## 3. 宅内コンテキスト認識のための機械学習フレームワーク

### 3.1 目　　的

宅内コンテキスト認識を行う際には，世帯ごとの個別性，すなわち部屋のレイアウトや環境の違いを考慮する必要がある．そのため，機械学習で認識を実装する場合には，世帯ごとにカスタマイズされた認識モデルを構築する必要がある．また，モデル構築の際にも，利用するデータや認識したいコンテキストの種類，機械学習アルゴリズムも世帯ごとに自由に選べるようにするために，本節では宅内コンテキスト認識を機械学習で実現する際の大まかな流れをフレームワークとして定義する．

フレームワークでは，モデル構築の流れのみを規定し，具体的なデータの種類やコンテキスト，機械学習アルゴリズムは規定しない．これによって，世帯ごとに異なる様々な要求や制約に柔軟に対応することを目的とする．

### 3.2 全体の流れ

提案するフレームワークは次の 5 ステップから成る (図 1)：

**STEP1：データの取得**

観測対象となる空間に，データを取得するためのデバイス（センサ，カメラ等）を設置し，適当な間隔でデータを収集し，観測期間内でデータの蓄積を行う．

**STEP2：データセットの作成**

**I. コンテキストの定義**：空間内で認識したい $m$ 個のコンテキスト $C = \{c_1, c_2, \cdots, c_m\}$ を定義する．

**II. データの選択**：各コンテキスト $c_i \in C$ に対して，STEP1 で蓄積したデータの中から $c_i$ を特徴的に反映している $n$ 個のデータ $data(c_i) = \{d_{i1}, d_{i2}, ..., d_{in}\}$ を選別し，それぞれに $c_i$

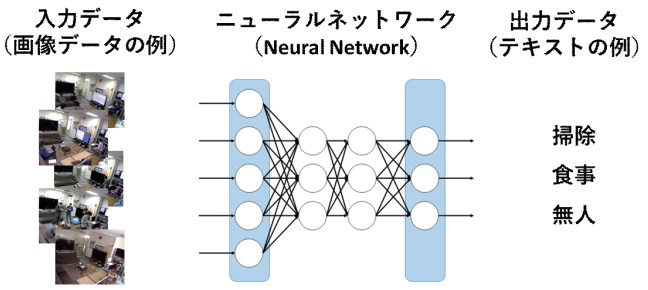


図 2　深層学習に基づく画像を活用した宅内コンテキスト認識の流れ

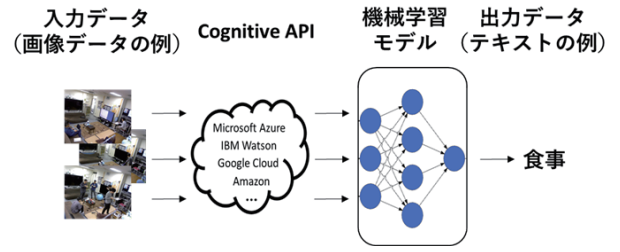


図 3　提案手法の流れ

のラベルを付ける．本ステップでは，合計 $m \times n$ 個のデータからなるデータセットが得られる．

**III．データの分割**：作成したデータセットを訓練用とテスト用に分割する．具体的には，各 $data(c_i)$ の $n$ 個のデータを $\alpha$ 個と $n-\alpha$ 個の 2 つの集合 $train(c_i)$ と $test(c_i)$ に分割する．データセットの分割に関しては，ランダムに分ける方法やホールドアウト法，交差検定法など，様々な手法がある．

**STEP3**：機械学習による認識モデルの構築

訓練データ $train(c_1), train(c_2), ..., train(c_m)$ を入力として教師あり機械学習アルゴリズム $A$ を適用し，認識モデル $M$ を構築する．ここで，$M$ は入力 $d_{ij}(1 \leq j \leq n)$ に対して，カテゴリ $c_i$ を出力する多値分類器である．$A$ の代表的なものとして，ニューラルネットワーク，SVM（Support Vector Machine），決定木（Decision Tree），各種の深層学習手法等が存在する．

**STEP4**：モデルの評価

テストデータ訓練用 $test(c_1), test(c_2), ..., test(c_m)$ を $M$ に入力して，対応するコンテキスト $c_1, c_2, ..., c_m$ が出力されるかを評価する．十分に高い精度が得られたら次へ，得られなければ以前のステップを見直す．

**STEP5**：モデルの配備と運用

訓練済のモデル $M$ を保存し，対象の環境からオンラインでアクセス可能にする．機械学習プラットフォームには，モデルを Web サービスとしてクラウド上に配備し，API でアクセスできるようにする機能も備わっている．対象の環境ではデータを取得し，$M$ に入力し，得られた出力 $c$ を認識された宅内コンテキストとする．得られたコンテキストは，宅内の状況に応じた情報提供や付加価値サービスのトリガに利用される．

## 4.　提案する宅内コンテキスト認識手法

### 4.1　キーアイディア

3. のフレームワークを，カメラから取得する画像データによって実装することを考える．画像を入力とする高精度な機械学習モデルを構築する最も強力な手法は，深層学習を利用することである．すなわち，STEP3 の $A$ に深層学習アルゴリズムを用いて $M$ を構築する．図 2 にその概要を示す．しかしながら，この方法は非常に多くの訓練データを必要とすることと，モデルの構築に多大な計算資源を必要とすることから，個別の一般家庭で実施するにはコストが高すぎて現実的ではない．それに対し，図 3 に提案手法の流れを示す．

そこで本研究では，コグニティブ API を活用して，訓練データの画像に含まれる情報（タグ集合）を抽出し，これを特徴量として軽量な教師あり機械学習にかけ，宅内コンテキストの認識モデルを構築する．2.4 で述べた通り，タグ集合そのものはコンテキスト認識に使えなかった．これは，API が提供する汎用的な画像認識では，世帯固有のコンテキストを十分に性質づけられなかったことにある．提案手法では，汎用的な画像認識で特徴量を抽出し，これを軽量な機械学習にかけて世帯固有のコンテキスト認識のための識別器を新たに作成する．これによって，深層学習を利用する場合に比べてはるかに少ないエフォートでのモデル構築をねらう．

### 4.2　提案手法の流れ

上記のキーアイデアをフレームワークの STEP3 に適用する．提案手法は，以下の 3 つのサブステップより構成される．

**STEP3.1**：画像認識 API を用いた特徴量抽出

訓練データ $train(c_i)(1 \leq i \leq n)$ に含まれる各画像 $d_{ij}$ を画像認識可能なコグニティブ API に送信し，API が $d_{ij}(1 \leq j \leq \alpha)$ 内に認識する単語の集合 $Tag(d_{ij}) = \{w_1, w_2, w_3, ...\}$ を得る．

**STEP3.2**：タグ集合のベクトル化

STEP3.1 で得られた全タグ集合の和 $\bigcup_{i,j} Tag(d_{ij})$ を文書コーパスとみなして，各 $Tag(d_{ij})$ を文書ベクトル $V_{ij} = [v_1, v_2, ...]$ に変換する．$V_{ij}$ に対して，コンテキスト $c_i$ をラベルとして付ける．文書をベクトル化する手法としては，TF-IDF, Word2Vec, GloVe, FastText, Exponential Family Embeddings 等がある．

**STEP3.3**：認識モデルの構築

STEP3.2 で取得されたベクトル $V_{ij}$ とそのラベル $c_i$ を訓練データとして，通常の教師あり機械学習のアルゴリズムを実行し，認識モデル $M$ を構築する．ここで，$M$ は入力ベクトルを $c_1, c_2, ..., c_n$ に分類する多値分類器である．

## 5.　評 価 実 験

### 5.1　データの準備

本実験では，我々の研究室の一部を定点カメラで撮影し，その画像によって研究室内のコンテキストを認識する．まず，フレームワークの STEP1 に従い，カメラは USB カメラを設置し，画像はプログラムによって 5 秒間隔で自動的に撮影・保存し，2 週間分の研究室内の日常的な状態の画像を蓄積した．STEP2-I では，全体ミーティング（$all\_meeting$），掃除（$cleaning$），食事（$eating$），無人（$no\_people$），個別ミーティング（$personal\_discussion$），遊び（$playing$），仕事（$working$）の 7 種類のコンテキストを定義した．STEP2-II では，7 種類それぞれのコンテキストに対して，そのコンテキストをよく示

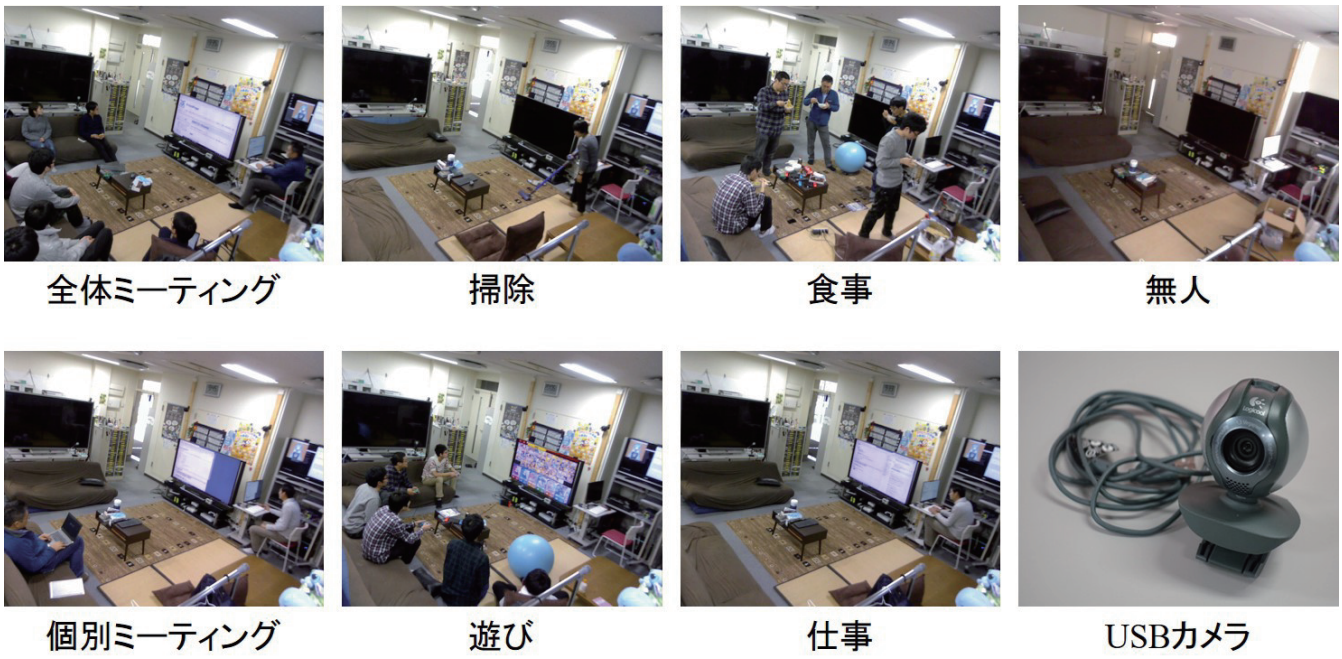


図 4　研究室の各コンテキストの画像例と利用した USB カメラ

表 1　Microsoft Azure Computer Vision API で抽出されたタグの例

| Context Label | Tag Results |
|---|---|
| cleaning | indoor, living, room, table, television, fire, fireplace, man, standing, filled, video, playing, woman, furniture, large, people, wii, dog, game |
| eating | indoor, person, room, table, living, man, sitting, food, filled, luggage, people, standing, suitcase, television, young, large, fire, kitchen |
| playing | indoor, person, room, table, sitting, living, people, man, food, standing, large, group, woman, playing, computer, kitchen, game |

している代表的な画像 100 枚を手動で選別した．STEP2-III では，各コンテキスト 100 枚の画像を半分に分け，訓練データとテストデータに分割した．

### 5.2　提案手法の実行

フレームワークの STEP3 を 4.2 の 3 つのサブステップに従って実行する．はじめに，STEP3.1 の画像認識 API を用いた特徴量の抽出には，Microsoft Azure Computer Vision API [6] を利用した．訓練データの画像を API に送信し，API の認識結果よりタグ集合を特徴量として抽出した．表 1 に Microsoft Azure Computer Vision API で抽出されたタグの例を示す．次に，STEP3.2 のタグ集合のベクトル化に関しては，TF-IDF（Term Frequency - Inverse Document Frequency）の手法を利用した．TF-IDF とは文書中に含まれる単語の出現頻度（TF）と希少性（IDF）を掛け合わせた値を計算する手法である．この手法を通し，文書中に存在する各単語のベクトルを取得できる．最後に，STEP3.3 の認識モデルの構築では，Microsoft Azure Machine Learning Studio の機械学習プラットフォーム上に，Multiclass Neural Network のアルゴリズムを利用して，宅内コンテキスト認識のためのモデルを構築した．

### 5.3　モデルの評価

STEP4 のモデルの評価では，はじめにテストデータの各画像を，STEP3.1, STEP3.2 と同様のやり方で，文書ベクトルに変換する．次にこの文書ベクトルを，STEP3.3 で構築したモデルに入力する．こうして得られた出力と，テストデータにあらかじめ付与されたラベルを比較して，モデルの精度を評価する．

モデルの評価には，以下のメトリクスを利用した．モデルの認識結果に対し，モデルの総体精度（Overall accuracy），平均精度（Average accuracy），各ラベルのテストデータのうちに正しく認識できたデータの精度（Micro-averaged precision），全体のテストデータのうちに正しく認識できたデータの精度（Macro-averaged precision），各ラベルのテストデータのうちに正しく認識できたデータの再現度（Micro-averaged recall），および全体のテストデータのうちに正しく認識できたデータの再現度（Macro-averaged recall）を算出する．また，コンテキスト毎の精度を見るために，混同行列（Confusion Matrix）を用いて評価する．

### 5.4　実 験 結 果

実験の結果を図 5 のメトリクス，図 6 の混同行列に示す．

図 5 の結果から，提案手法はどのメトリクスに対しても 0.9 以上の高い認識精度を達成することが分かった．

また図 6 の混同行列から，各コンテキスト毎の精度を評価する．全体ミーティングを正しく識別した割合は 95.3% であり，残り 4.7% を遊びに誤って認識した．掃除を正しく認識できた割合は 90.9% であり，掃除を食事に誤って認識したのは 5.5% であり，個別ミーティングに誤って認識したのは 3.6% であった．食事を正しく認識した割合は 83.3% であり，掃除に誤って認識したのは 4.2%，遊びに誤って認識したのは 12.5% である．

Metrics

| | |
|---|---|
| Overall accuracy | 0.928571 |
| Average accuracy | 0.979592 |
| Micro-averaged precision | 0.928571 |
| Macro-averaged precision | 0.924293 |
| Micro-averaged recall | 0.928571 |
| Macro-averaged recall | 0.925448 |

図 5　メトリクスによる認識モデルの評価

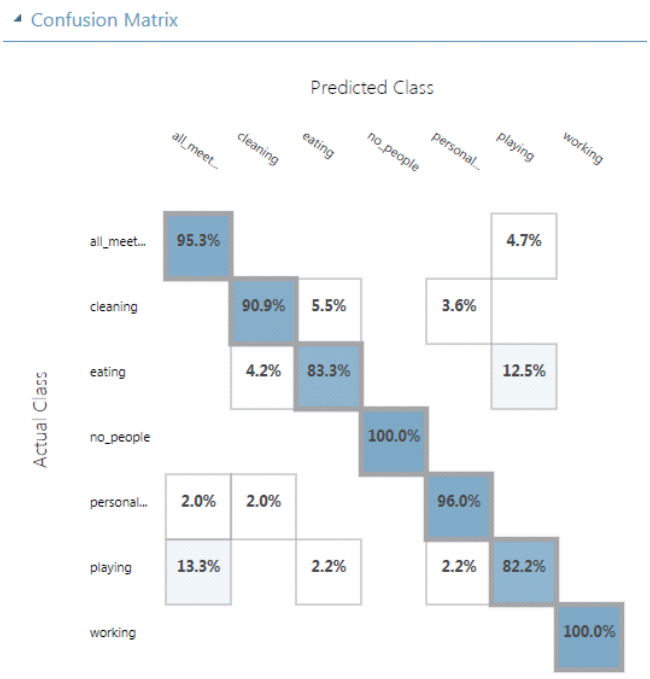


図 6　混 同 行 列

無人を正しく認識できた割合は 100% である．個別ミーティングを正しく認識できた割合は 96.0% であり，全体ミーティングに誤って認識したのは 2.0%，掃除に誤って認識したのは 2.0% であった．遊びを正しく認識できた割合は 82.2% であり，全体ミーティングに誤って認識したのは 13.3%，食事に誤って認識したのは 2.2%，個別ミーティングに誤って認識したのは 2.2% であった．仕事を正しくの認識できた割合は 100% であった．

以上の結果から，無人と仕事の認識できた比率が最も高く，食事と遊びの認識できた比率が最も低いことが明らかになった．そして，食事を遊びと誤るケース，遊びを全体ミーティングと誤るケースが，他と比較して多いことが分かった．

### 5.5 考　　察

本実験で構築したモデルの評価結果において，少人数（特に一人）あるいは無人を映したコンテキストの認識のほうが精度が高く，機械にとってより容易なタスクであると推察される．これに対して，食事や遊び，全体ミーティングといった多人数が参加するコンテキストを区別することはより難しいことがわかる．多人数が参加するコンテキストは，ある程度の期間内における人同士のインタラクションやアクションの系列で特徴づけられることが自然である．しかしながら，画像によるコンテキスト認識においては，瞬間的なスナップショットとしてコンテキストが切り取られるため，人間の目でも判別しにくい場合がある．このことが数字で表れたのではないかと考えている．

## 6. おわりに

本稿では，宅内コンテキスト認識に向けた機械学習フレームワークを提案し，一般家庭で導入可能な画像を活用したコンテキスト認識手法を提案した．提案手法では，コグニティブ API を活用して画像に含まれる情報を抽出し，これを特徴量として軽量な教師あり機械学習にかける．これにより，深層学習によるアプローチに比べてはるかに少ないエフォートで，宅内コンテキスト認識が実現できる．また，評価実験では，Microsoft Azure Computer Vision API で抽出したタグ集合を TF-IDF でベクトル化し，研究室内の 7 種類のコンテキストを認識する実験を行った．その結果，0.92 以上の認識精度を達成するモデルが構築できた．混同行列による分析では，食事を遊びに誤るケースや，遊びを全体ミーティングに誤るケースなど，多人数が参加するコンテキストに誤認識が多いことがわかった．

今後の課題として，異なるデータセットの分割方法や，教師あり学習のアルゴリズムを用いた評価が挙げられる．また，今回構築したモデルを研究室に配備し，オンラインでコンテキスト認識を行って，その性能と有効性を評価したい．最後に，新たなコンテキストの追加や他の環境での実施に伴う，モデルの再構築や再利用法の検討も行いたい．

**謝辞**　この研究の一部は，科学技術研究費（基盤研究 B 16H02908，18H03242, 18H03342，基盤研究 A 17H00731），および，立石科学技術振興財団の研究助成を受けて行われている．

### 文　　献

[1] 上田健揮，玉井森彦，荒川　豊，諏訪博彦，安本慶一，“ユーザ位置情報と家電消費電力に基づいた宅内生活行動認識システム，” 情報処理学会論文誌，vol.416-425，no.57，p.2，Feb. 2016.
[2] 大内一成，土井美和子，“スマートフォンを用いた生活行動認識技術，” 東芝レビュー，vol.68，no.6，pp.40–43，June 2013.
[3] 玉水一柔，榊原誠司，佐伯幸郎，中村匡秀，安田　清，“宅内の環境変化と声掛けに基づく在宅高齢者の日常生活行動センシングシステムの検討 (ライフインテリジェンスとオフィス情報システム)，” 電子情報通信学会技術研究報告: 信学技報，vol.116，no.405，pp.7–12，Jan. 2017.
[4] 陳　思楠，佐伯幸郎，中村匡秀，“画像に基づくコグニティブ api の宅内センシングへの適用可能性，” 電子情報通信学会技術研究報告: 信学技報，no.SC-19，pp.31–36，Aug. 2018.
[5] S. Chen, S. Saiki, and M. Nakamura, “Evaluating feasibility of image-based cognitive apis for home context sensing,” ICSPIS2018, Nov. 2018.
[6] M. Azure, “Computer vision,” https://azure.microsoft.com/ja-jp/services/cognitive-services/computer-vision/. visited on 2019-02-01.
[7] “Tf-idf で文書をベクトル化，” http://ailaby.com/tfidf/, Aug. 2016. visited on 2019-02-01.
[8] M. Azure, “Azure machine learning studio,” https://azure.microsoft.com/ja-jp/services/machine-learning-studio/. visited on 2019-02-01.
[9] Google, “Cloud machine learning engine,” https://cloud.google.com/ml-engine/?hl=ja. visited on 2019-02-01.
[10] Amazon, “Aws での機械学習，” https://aws.amazon.com/jp/machine-learning/. visited on 2019-02-01.