

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

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

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、クラウドサービスで行われる画像解析を通して様々なコンテキスト情報を抽出するコグニティブ API (Application Programming Interface) が提供されている。ある場所の画像とこれらの API を連携して活用することで、人間の活動や周辺環境の変化をリアルタイムにセンシングすることが期待できる。本研究では、宅内の状態を取得する画像に基づく宅内センシングシステムの実現を目的とする。本稿では宅内センシングの実現に向け、まずコグニティブ API の一つである画像タグ付け (ラベリング, タギング) API から、宅内のどのようなコンテキスト情報が得られるかを調査し、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を考察する。

キーワード 宅内センシング, コグニティブ API, 画像, コンテキスト情報

Feasibility study of image-based cognitive API for home sensing

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

[†] Kobe University Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo, 657-8501 Japan

E-mail: [†]chensinan@ws.cs.kobe-u.ac.jp, ^{††}sachio@carp.kobe-u.ac.jp, ^{†††}masa-n@cs.kobe-u.ac.jp

Abstract In recent years, cognitive API (Application Programming Interface) that extracts various context information through image analysis performed in cloud service is provided. By using the image of a certain place and these APIs in cooperation, it can be expected to sense changes in human activities and the surrounding environment in real time. In this research, we aim to realize home sensing system based on an image to acquire the condition in a house. To realize home sensing, we first investigate what kind of context information at home can be obtained from the image tagging (labeling, tagging) API which is one of the cognitive APIs, and check the cognitive API consider the applicability to home sensing.

Key words Home sensing, Cognitive API, Image, Context information

1. はじめに

IOT 技術を活用した様々なセンシングデバイスの創出に伴い、スマートホームと連携した宅内センシングシステムの研究開発が盛んに行われている。宅内センシングとは、環境センサや電力消費量、家電操作ログなど様々なデバイスから得られた情報をもとに、宅内の状況や状態などを取得するものである。これまでの宅内センシングの研究では、センシングデバイスとして温度・湿度・人感などのセンサを利用することが一般的である。例えば、人感センサを利用した高齢者の見守りシステム [1] や環境センサを用いた在宅高齢者の日常生活活動センシングシステム [2] など、様々なセンサ装置を活用し、宅内センシングを実現している。このような宅内センシングにおいて、カメラなどで得られる静止画・動画などの画像データを用いたセンシングは、設置コストが高い、取得したデータは直接にセンシングデータとして利用できず、様々な解析を通したうえで

活用する必要があるため、結果としての利便性が低い、さらに日常生活への侵襲性が高い、プライバシー侵害等の心理的障壁があること等の問題から積極的な活用は進んでいない。しかし、実環境で得られる画像データは、人間の外観や動作、顔表情および周囲の状況など、宅内センシングにおいて有益であると言える情報を含んでいるため、積極的な活用を進めることで、より品質の高い宅内センシングが実現できると考えられる。

近年、クラウドコンピューティング技術の急速な発展に伴い、これまでエッジ側で実行していた様々な処理を、クラウドサービスでの代替が可能となり、画像を用いた宅内センシングへの研究が必要となる。画像からの様々な情報認識も、人工知能の広がり共にクラウドサービスの一つであるコグニティブサービスとして利用環境が整いつつある。コグニティブサービスでは従来コンピュータでの処理が困難であった様々な認知に関する機能を様々な環境から統一的に利用できる多種多様なコグニティブ API (Application Programming Interface)

が提供されている。

コグニティブ API では、視覚、音声、言語などの多くの分野の情報をもとに、あたかもコンピュータが自ら考え、見たり、聞いたり、人間のよりよい活動を行えるようなサポートの提供などが実現している。視覚分野におけるコグニティブ API では、画像解析を通して、画像中に存在している物や人間の活動、顔表情など、様々なコンテキスト情報が得られ、コンピュータがより画像内容を理解できるようになる。そこで、宅内で撮影した画像を活用すれば、コグニティブ API より得られた画像のコンテキスト情報から、宅内センシングとして宅内での人間の活動や周辺環境の変化をリアルタイムにセンシングすることが期待できる。図 1 に、コグニティブ API を利用した画像認識の利用モデルを示す。一方、コグニティブ API による出力は、汎用的な用途が想定されているものであり、その出力を宅内センシングへの適用するには様々な課題がある。

本研究では、宅内に設置した簡易的な撮影システムをもとに、宅内の状態を取得する画像に基づく宅内センシングシステムの実現を目的とする。実現する宅内センシングシステムでは、宅内であるエッジシステムでの画像の撮影、簡易的な前処理、コグニティブ API への送付のみを行い、負荷処理が高い部分や、認識などの処理はすべてクラウドサービスで行う。

本稿では、宅内センシングシステムの実現に向け、まずコグニティブ API の一つである画像タグ付け（ラベリング、タギング）API が、宅内コンテキスト認識に適用可能であるかの調査を行う。具体的には、サービス API が出力したタグ集合が、与えられたコンテキストの認識（分離）が十分可能であるか、また公開されているコグニティブ API によりどのような差が存在するかを確認する。この知見をもとに、異なる条件下においても同様の適用実験を行うことができる。さらに、コグニティブ API の画像解析機能がどの程度で宅内センシングにおいて実用性が備えるかを明らかにすることが期待できる。

主要なアプローチとして、クラウド各社が提供しているコグニティブ API から得られたタグ集合を文書（コーパス）とし、文書間類似度 [3] を用いて、様々な宅内コンテキストの分離が実現可能であるかの評価を行う。これらの評価のため、我々の研究室の日常活動状態を実験対象とし、日常的な状態の撮影を行い、これらの画像を Microsoft Azure, IBM Watson, Google が提供している各コグニティブ API を用いてタグ集合を取得し、得られたタグ集合の記録は **TF-IDF(Term Frequency - Inverse Document Frequency)** [4] を適用した上で文書化し、あらかじめ設定したコンテキストへの分離性能についてコサイン類似度をもとに、内的結合、外的分離の観点から評価する。内的結合とは、同クラス内での類似性であり、外的分離とは異なるクラス間での分離性を意味する。

実験の結果に関しては、内的結合における同クラス内の対象では、それぞれ類似度の大きな違いはないが、Google は Microsoft Azure と IBM Watson と比べ、API の出力したタグ集合の類似度がより高いことが明らかになった。また、外的分離の観点では、Google は Microsoft Azure と IBM Watson より、複数人がコンテキストとして存在するような状態におい

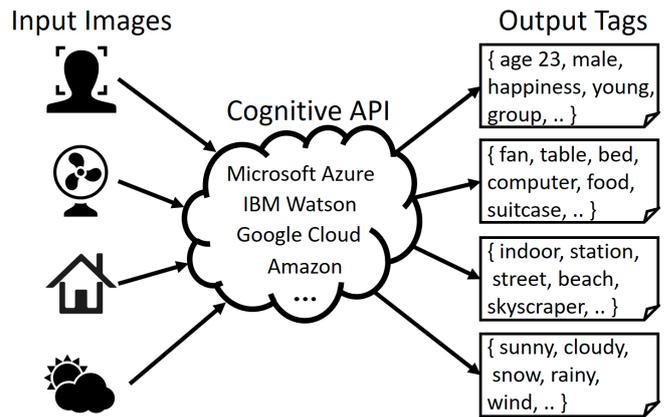


図 1 コグニティブ API を利用した画像認識の利用モデル

て、分離性能が高いことがわかった。また、人が映っていない状態を定常的なバイアス成分として捉え、その認識結果を各コンテキストの出力から除去することで、より分離性能が向上する傾向が明らかになった。

2. 準備

2.1 宅内センシング

宅内センシングとは、環境センサや電力消費量、家電操作ログなど様々なデバイスから得られた情報をもとに、宅内の状況や状態などを取得するものである。センシング結果としては、宅内の温度や湿度、時間による消費電力の変動などの宅内状態を示す時系列数値データや、何時にどのような家電をどう操作したかなどのログ、および食事や勉強などの宅内における人間の活動状況を表すコンテキストなどが含まれる。

2.2 画像に基づくコグニティブ API

クラウドコンピューティング技術の急速な発展に伴い、これまでエッジ側で実行していた様々な処理を、クラウドサービスでの代替が可能となった。このような中、近年爆発的に普及が進んでいる人工知能を活用する様々なサービスもクラウドサービスとして利用が可能となりつつある。人工知能を活用するクラウドサービスとして、視覚・音声・言語・知識などをもとに、そこに含まれている情報をコンピュータに認識させる、コグニティブサービスがある。コグニティブサービスでは従来コンピュータでの処理が困難であった様々な認知に関する機能を実現することができる。画像に基づくコグニティブサービスは、画像から様々な情報を抽出することが可能なサービスであり、Microsoft Azure, IBM Watson, Google, Amazon 等主要なクラウドサービスで広く提供されているサービスである。一般的にクラウドサービスは API という形で外部から利用することが可能であるため、これらのコグニティブ API についても利用者はこれらの API に画像を送信することで、画像内にある様々な情報を受け取ることができる。例えば、人間の顔画像から年齢や性別、顔表情による感情値を推測することができる。そして、家電や使用道具、食べ物、色等、様々な物品に対する画像認識もできる。また、画像背景から場所や天気に関するタグも出力することができる。

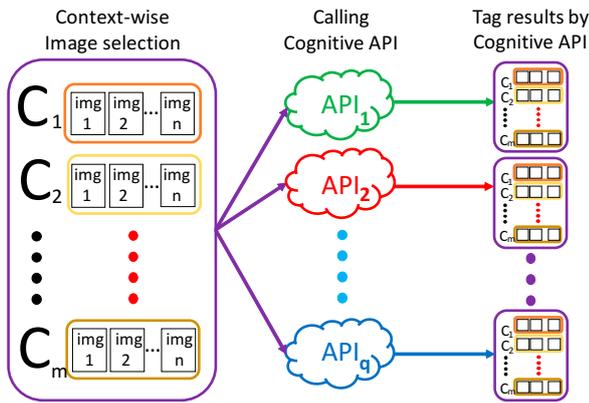


図2 コンテキストのラベル設定から結果の分析までの流れ

2.3 Microsoft Azure Computer Vision API

Computer Vision API [5] は、Microsoft が提供するクラウドサービスである Microsoft Azure における Cognitive Services の視覚分野として提供されている API の一つである。Computer Vision API では、画像の分析や画像内のテキストの読み取り、手書き文字の読み取り等、画像解析を活用した様々な機能を提供する API である。

2.4 IBM Watson Visual Recognition API

Visual Recognition API [6] は、IBM が提供するクラウドサービスである IBM Cloud における Cognitive Computing System である IBM Watson においてディープラーニングを使用し、画像に写った様々なものを分析・認識する画像認識サービスとして提供されている API である。

2.5 Google Cloud Vision API

Vision API [7] は、Google が提供するクラウドサービスである Google Cloud Platform における画像分析 API である。各画像を素早く分類する機能や、画像内の個々の物体や人の顔を検出する機能、画像内に含まれているテキストを検出して読み取る機能を備えている。

3. コグニティブ API の比較手法の提案

3.1 目的

クラウドサービスとして提供されているコグニティブ API は、サービス提供元ごとに得られる情報の種類や傾向が異なる。宅内センシングとして活用する際、それぞれの API はどのような特徴を持つかをあらかじめ評価しておくことが必要である。本節では複数のコグニティブ API における出力が、宅内の各種コンテキストを分離可能であるかの比較手法について提案する。

3.2 全体の流れ

本稿で提案するコグニティブ API の比較手順は 6 ステップから成る。図 2 にコンテキストのラベル設定から結果の分析までの流れを示す。以下に各ステップについて述べる。

Step1: 画像の取得

観測対象となる空間に、USB カメラなど画像を撮影可能なデバイスを設置し、適当な間隔で撮影し、検証用画像の蓄積を行う。

Step2: 認識ラベルの設定

コグニティブ API から取得された情報をもとに判別したい m 個のコンテキスト ($C = \{c_1, c_2, \dots, c_m\}$) を定義する。

Step3: 画像の分類

Step1 で蓄積した画像に対し、Step2 で定義したコンテキストをもとに、それぞれの画像がどのコンテキストの映像であるかを分類し、コンテキストごとに特徴的な n 枚の画像 ($IMG(c)_i = \{img(c)_1, img(c)_2, \dots, img(c)_n\}$) を選択する。

Step4: API の呼び出し

分類した画像を比較対象である q 種のコグニティブ API ($API = \{api_1, api_2, \dots, api_q\}$) にそれぞれ投入する。API の返り値を各画像について記録し、画像ごとに得られたラベルの集合 ($L = \{l_1, l_2, \dots\}$) を作成する。なお L の長さは画像ごとに異なる。以上のステップにより $m \times n \times q$ 個の L が作成される。

Step5: 結果の分析

Step4 で得られたすべての L それぞれを 1 文書とみなし、 api_x ごとに同一コンテキスト内で文書間類似度をもとに、内的結合であるコンテキスト内類似度 $InSim_x(c_i)$ ($1 \leq i \leq m$) を算出する。同様に、 api_x ごとに、外的分離として異なるコンテキスト間類似度 $ExSim_x(c_i, c_j)$ ($1 \leq i, j \leq m$) を算出する。

内的結合の評価に関しては、各コンテキストのクラスタ内部の画像タグの間に類似度 ($InSim_x(c_i)$) が大きければ、同じコンテキスト (c_i) に対し、毎回よく似た特徴が抽出され、その API (api_x) の認識性能が強いことをいえる。外的分離の評価では、各コンテキストのクラスタ間の画像タグ集合の類似度 ($ExSim_x(c_i, c_j)$) が小さければ、その API (api_x) は各コンテキストのクラスタ間 (c_i と c_j) がよく分離できるといえる。そして、各コンテキストのクラスタ間 (c_i と c_j) の分離において、各 API の間で比較することが可能である。例えば、 api_x は api_y より外的結合の類似度が大きければ ($ExSim_x(c_i, c_j) > ExSim_y(c_i, c_j)$)、 api_x は api_y より分離性能が優れているといえる。

4. 適用実験

3. で提案した評価手法に基づき、実際の環境下で撮影した画像データをもとに提案手法の適用実験を行う。

4.1 実験環境

本実験は、我々研究室の一部を実験場所とし、人間の日常活動状態の画像を取得するため、研究室内で USB カメラを設置した。実験における画像の取得は、あらかじめ作成した 5 秒ずつ USB カメラで撮影するプログラムを常に実行し、1 週間の研究室内の日常的な状態の画像を蓄積した。コンテキストは、全体ミーティング、漫画、掃除、食事、ゲーム、無人、個別ミーティング、研究、睡眠、スマートフォン、テレビの 11 種類を設定した。図 3 に本実験で撮影した研究室の各コンテキストの画像例と利用した USB カメラを示す。また、本実験における API へ投入する画像の選択は、撮影した画像ファイル中からそれぞれコンテキストをよく示していると思われる代表的な画像 10 枚をなるべく異なる日時のものから目視で行った。利用したコグニティブ API は、Computer Vision API, Visual Recognition API, Vision API の 3 種類である。ラベルの特徴



図3 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 実験環境

画像収集場所	研究室内
画像蓄積期間	7日
撮影方法	USBカメラ
画像解像度	1280 × 1024
コンテキスト数	11
抽出画像数	10
比較API数	3
特徴量化	TF-IDF
文書間類似度	コサイン類似度

量化には TF-IDF を、文書間類似度としてはコサイン類似度を用いている。これらについては後述する。実験環境を表1に示す。本実験では、設定した全ての認識コンテキストのレベルを利用し、コンテキスト内のタグ集合の文書間類似度の評価と各コンテキストの文書間類似度を評価したもの (Method1)、および状況に関わらず常に出現するタグとみなせるタグをそれぞれの L から除去したもの (Method2) の3通りで行う。

4.1.1 特徴量化

本実験では認識ラベル L の特徴量として **TF-IDF (Term Frequency - Inverse Document Frequency)** を利用している。TF-IDF とは文書中に含まれる単語の重要度に応じそれ

ぞれの単語に重み付けを行い評価する手法である。多くの文書に出現する語は重要度が低いとみなされ重みが小さく、特定の文書にしか出現しない単語は重要度が高いとされ重み付けが大きくなる特性を持つ。

4.1.2 文書間類似度の算出

本実験では、文書間類似度としてコサイン類似度を利用している。コサイン類似度は、文書同士を比較する際に用いられる類似度計算手法であり二つの文書から生成されたベクトル同士の成す角度の大きさを類似度を表現する。1に近ければ類似しており、0に近ければ似ていないことを意味する。

4.2 評価項目

実験は提案手順に従い、あらかじめ分類した画像を3種類の各APIに投入した。各出力結果をもとに以下の3通りを評価している。

4.2.1 内的結合

各コンテキスト内で画像ごとに作成された L の文書間類似度を全ての組み合わせで算出し、その平均値を評価する。

4.2.2 外的分離 (Method1, 単純分類)

Method1では、各コンテキスト内で画像ごとに作成された L を全て結合し、新たに L_{all} の集合を作成する。この時、APIから得られた出力ラベルを全て利用する。 L_{all} を特徴量化し、

表 2 文書間類似度における内的結合の結果

api	all meeting	cartoon	cleaning	eating	game	no people	personal meeting	research	sleeping	smart phone	tv
azure	0.650	0.641	0.581	0.561	0.607	0.807	0.649	0.696	0.819	0.709	0.599
watson	0.643	0.611	0.474	0.512	0.650	0.759	0.555	0.680	0.785	0.818	0.607
google	0.679	0.780	0.727	0.564	0.821	0.738	0.708	0.670	0.784	0.844	0.690

表 3 文書間類似度における単純分類 (Method1) の結果

context	api	all meeting	cartoon	cleaning	eating	game	no people	personal meeting	research	sleeping	smart_phone	tv
all_meeting	Azure	1.000	0.812	0.808	0.767	0.773	0.663	0.874	0.703	0.644	0.732	0.773
	Watson	1.000	0.435	0.454	0.696	0.471	0.432	0.453	0.318	0.391	0.532	0.424
	Google	1.000	0.310	0.351	0.470	0.356	0.238	0.402	0.322	0.260	0.249	0.368
cartoon	Azure	0.812	1.000	0.859	0.697	0.844	0.685	0.895	0.889	0.635	0.720	0.865
	Watson	0.435	1.000	0.887	0.653	0.864	0.703	0.881	0.849	0.577	0.695	0.890
	Google	0.310	1.000	0.922	0.806	0.663	0.851	0.615	0.875	0.892	0.864	0.887
cleaning	Azure	0.808	0.859	1.000	0.666	0.791	0.622	0.908	0.802	0.635	0.690	0.809
	Watson	0.454	0.887	1.000	0.724	0.847	0.587	0.825	0.765	0.501	0.608	0.867
	Google	0.351	0.922	1.000	0.895	0.666	0.732	0.673	0.758	0.794	0.749	0.891
eating	Azure	0.767	0.697	0.666	1.000	0.700	0.595	0.754	0.658	0.554	0.660	0.659
	Watson	0.696	0.653	0.724	1.000	0.657	0.617	0.647	0.617	0.498	0.623	0.679
	Google	0.470	0.806	0.895	1.000	0.750	0.600	0.781	0.761	0.680	0.626	0.871
game	Azure	0.773	0.844	0.791	0.700	1.000	0.827	0.888	0.862	0.763	0.782	0.858
	Watson	0.471	0.864	0.847	0.657	1.000	0.605	0.881	0.817	0.558	0.699	0.901
	Google	0.356	0.663	0.666	0.750	1.000	0.507	0.964	0.839	0.533	0.506	0.837
no_people	Azure	0.663	0.685	0.622	0.595	0.827	1.000	0.740	0.738	0.868	0.813	0.724
	Watson	0.432	0.703	0.587	0.617	0.605	1.000	0.652	0.632	0.711	0.798	0.692
	Google	0.238	0.851	0.732	0.600	0.507	1.000	0.479	0.741	0.942	0.972	0.718
personal_meeting	Azure	0.874	0.895	0.908	0.754	0.888	0.740	1.000	0.846	0.701	0.789	0.843
	Watson	0.453	0.881	0.825	0.647	0.881	0.652	1.000	0.953	0.582	0.749	0.911
	Google	0.402	0.615	0.673	0.781	0.964	0.479	1.000	0.764	0.521	0.489	0.833
research	Azure	0.703	0.889	0.802	0.658	0.862	0.738	0.846	1.000	0.647	0.682	0.874
	Watson	0.318	0.849	0.765	0.617	0.817	0.632	0.953	1.000	0.538	0.670	0.885
	Google	0.322	0.875	0.758	0.761	0.839	0.741	0.764	1.000	0.794	0.763	0.853
sleeping	Azure	0.644	0.635	0.635	0.554	0.763	0.868	0.701	0.647	1.000	0.861	0.645
	Watson	0.391	0.577	0.501	0.498	0.558	0.711	0.582	0.538	1.000	0.803	0.595
	Google	0.260	0.892	0.794	0.680	0.533	0.942	0.521	0.794	1.000	0.970	0.769
smart_phone	Azure	0.732	0.720	0.690	0.660	0.782	0.813	0.789	0.682	0.861	1.000	0.698
	Watson	0.532	0.695	0.608	0.623	0.699	0.798	0.749	0.670	0.803	1.000	0.745
	Google	0.249	0.864	0.749	0.626	0.506	0.972	0.489	0.763	0.970	1.000	0.746
tv	Azure	0.773	0.865	0.809	0.659	0.858	0.724	0.843	0.874	0.645	0.698	1.000
	Watson	0.424	0.890	0.867	0.679	0.901	0.692	0.911	0.885	0.595	0.745	1.000
	Google	0.368	0.887	0.891	0.871	0.837	0.718	0.833	0.853	0.769	0.746	1.000

表 4 各 API の出力したタグ集合の例

context label	api	tag results
All_meeting	Azure	indoor, person, room, living, ceiling, woman, sitting, people, table, watching, playing, cluttered, man, area, computer, dog, filled, office, group, young, standing, television, video, kitchen, large, holding, game
	Watson	classroom, indoors, newsroom, office, building, control room, workroom, control center, test room
	Google	room, technology, office, electronic device, classroom, computer, course, learning, workshop
no_people	Azure	indoor, room, table, living, sitting, computer, desk, kitchen, area, monitor, office, television, filled, large, video, furniture, game, playing, standing, several, man, laptop, white, holding, keyboard, fire, remote, wii, dog, refrigerator
	Watson	living room, indoors, control room, basement, support, supporting structure, office, building, ultramarine color,
	Google	room, electronic device, technology, electronics, furniture, personal computer, flooring, house, interior design, desk

その結果から他のコンテキストで同様に作成された特徴量をもとに文書間類似度を算出する。

4.2.3 外的分離 (Method2, 前処理付き分類)

Method2 では、Method1 に対し、定常画像である無人で得られたタグを各画像のバイアス成分として捉え、これらを除去した L_{all} をもとに特徴量化し、同様にバイアスを除去した他のコンテキストで作成された特徴量をもとに文書間類似度を算出する。表 4 に各 API の出力したタグ集合の例を示す。

4.3 結果

本実験では、評価項目に従い算出した結果を表 2, 表 3, 表 5 に示す。ただし、各表における API 名は全て会社名としている。表 2 に文書間類似度における内的結合の結果を示す。表 2 の結果から、Microsoft Azure と IBM Watson は Google と比べ、API より出力した各コンテキストのクラスタ内部のタグ集合の類似度の結果がより低いことがわかった。表 3 に文書間類似度における単純分類 (Method1) の結果を示す。表 3 の結果から、Microsoft Azure と IBM Watson は Google より、API より出力した全体ミーティングとそれ以外のクラスタのタグ間の類似度の差がより小さいことがわかった。表 5 に文書間類似度における前処理付き分類 (Method2) の結果を示す。表 5 の結果から、Microsoft Azure と IBM Watson は Google と比べ、類似度 0 の結果がより少ないことがわかった。本実験の上述した結果より、少数の宅内画像のコンテキストが認識できる以外

表 5 文書間類似度における前処理付き分類 (Method2) の結果

context	api	all meeting	cartoon	cleaning	eating	game	no people	personal meeting	research	sleeping	smart phone	tv
all_meeting	Azure	1.000	0.528	0.645	0.576	0.412	0.000	0.815	0.188	0.230	0.402	0.336
	Watson	1.000	0.143	0.222	0.616	0.271	0.000	0.222	0.137	0.037	0.281	0.184
	Google	1.000	0.000	0.000	0.369	0.234	0.000	0.368	0.094	0.000	0.232	0.232
cartoon	Azure	0.528	1.000	0.683	0.405	0.687	0.000	0.691	0.725	0.228	0.360	0.703
	Watson	0.143	1.000	0.802	0.272	0.787	0.000	0.821	0.753	0.225	0.000	0.930
	Google	0.000	1.000	0.000	0.000	0.482	0.000	0.000	0.913	0.000	0.000	0.000
cleaning	Azure	0.645	0.683	1.000	0.447	0.579	0.000	0.794	0.642	0.204	0.308	0.593
	Watson	0.222	0.802	1.000	0.471	0.733	0.000	0.747	0.625	0.344	0.154	0.812
	Google	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
eating	Azure	0.576	0.405	0.447	1.000	0.315	0.000	0.536	0.185	0.178	0.394	0.278
	Watson	0.616	0.272	0.471	1.000	0.371	0.000	0.374	0.395	0.145	0.191	0.322
	Google	0.369	0.000	0.000	1.000	0.567	0.000	0.548	0.268	0.000	0.659	0.659
game	Azure	0.412	0.687	0.579	0.315	1.000	0.000	0.686	0.670	0.305	0.383	0.634
	Watson	0.271	0.787	0.733	0.371	1.000	0.000	0.759	0.636	0.235	0.000	0.849
	Google	0.234	0.482	0.000	0.567	1.000	0.000	0.763	0.791	0.000	0.861	0.861
no_people	Azure	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Watson	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Google	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
personal_meeting	Azure	0.815	0.691	0.794	0.536	0.686	0.000	1.000	0.536	0.262	0.461	0.534
	Watson	0.222	0.821	0.747	0.374	0.759	0.000	1.000	0.920	0.195	0.117	0.886
	Google	0.368	0.000	0.000	0.548	0.763	0.000	1.000	0.317	0.000	0.778	0.778
research	Azure	0.188	0.725	0.642	0.185	0.670	0.000	0.536	1.000	0.170	0.131	0.731
	Watson	0.137	0.753	0.625	0.395	0.636	0.000	0.920	1.000	0.207	0.000	0.798
	Google	0.094	0.913	0.000	0.268	0.791	0.000	0.317	1.000	0.000	0.408	0.408
sleeping	Azure	0.230	0.228	0.204	0.178	0.305	0.000	0.262	0.170	1.000	0.394	0.187
	Watson	0.037	0.225	0.344	0.145	0.235	0.000	0.195	0.207	1.000	0.000	0.253
	Google	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
smart_phone	Azure	0.402	0.360	0.308	0.394	0.383	0.000	0.461	0.131	0.394	1.000	0.258
	Watson	0.281	0.000	0.154	0.191	0.000	0.000	0.117	0.000	0.000	1.000	0.158
	Google	0.232	0.000	0.000	0.659	0.861	0.000	0.778	0.408	0.000	1.000	1.000
tv	Azure	0.336	0.703	0.593	0.278	0.634	0.000	0.534	0.731	0.187	0.258	1.000
	Watson	0.184	0.930	0.812	0.322	0.849	0.000	0.886	0.798	0.253	0.158	1.000
	Google	0.232	0.000	0.000	0.659	0.861	0.000	0.778	0.408	0.000	1.000	1.000

に、大部分の宅内画像に対してコグニティブ API を用いた認識コンテキストの効果が理想でないことを明らかにした。

4.4 考察

本実験の分析した結果から、コグニティブ API を用いた宅内画像の認識では、画像中の認識したい場所の不明確になればコンテキスト結果が乱れることになるのではないであろうかと考える。画像中に存在している物品が多ければ、API の出力とした画像の認識結果のコンテキストを、実験で認識したい画像のコンテキストと外れる可能性も有していると思う。一方、画像中で映した宅内環境の明るさも一つの要因と思う。例えば、暗い部屋の画像を認識すれば、人間の目でも判別できない場合があるのであるであろうかと考える。

5. おわりに

本稿では、画像を用いた宅内センシングシステム実現に向けた、コグニティブ API の適用性の評価を行った。実験では Microsoft Azure, IBM Watson, Google の三社のクラウドサービスにおいて提供されている画像に基づくコグニティブ API から得られたタグ集合を文書間類似度に着目し、その分類性能について評価を行った。その結果、Google は Microsoft Azure と IBM Watson と比べ、API より出力した内的結合における類似度の結果が高く、外的分離の Method1 における各クラスタのタグ集合の類似度の差がより大きく、外的分離の Method2 における各クラスタのタグ集合の類似度の 0 結果が多いことがわかった。本実験における各コンテキストの画像に対し、クラウド各社のコグニティブ API の認識したコンテキストの内部

結合と外部分離がそのままでは十分に行えないことが明らかになった。今後の課題として、コグニティブ API を用いた画像解析を行う前に、認識したい部分画像を抽出することなどより分類性の向上に向けた取り組みを行う必要がある。また、コグニティブ API から得られる認識タグ以外の情報を積極的に活用することで、さらなるコンテキスト分類能力の向上を目指す。

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] 津田麻衣, 玉井森彦, 安本慶一, “居室行動のセンシングに基づく独居高齢者を対象とした見守りシステムの提案,” 研究報告モバイルコンピューティングとユビキタス通信 (MBL), vol.2013, no.16, pp.1-5, may 2013.
- [2] 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 (情報通信マネジメント),” 電子情報通信学会技術研究報告 = IEICE technical report : 信学技報, vol.116, no.404, pp.7-12, jan 2017.
- [3] 神島敏弘, “クラスタリング,” <http://www.kamishima.net/archive/clustering.pdf>. visited on 2018-07-23.
- [4] wikipedia, “tf-idf,” <https://ja.wikipedia.org/wiki/Tf-idf>, Oct. 2017. visited on 2018-07-23.
- [5] “Microsoft azure computer vision,” <https://azure.microsoft.com/ja-jp/services/cognitive-services/computer-vision/>. visited on 2018-07-23.
- [6] “Ibm watson visual recognition,” <https://www.ibm.com/watson/jp-ja/developercloud/visual-recognition.html>. visited on 2018-07-23.
- [7] “Google cloud vision api,” <https://cloud.google.com/vision/>. visited on 2018-07-23.