HOKKAIDO UNIVERSITY

Title	異種データを用いた画像共有サービスに投稿される画像の分類に関する研究
Author(s)	斉藤, 直輝
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第13520号
Issue Date	2019-03-25
DOI	https://doi.org/10.14943/doctoral.k13520
Doc URL	https://hdl.handle.net/2115/74221
Type	doctoral thesis
File Information	Naoki_Saito.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

博 士 論 文

異種データを用いた画像共有サービスに投稿される画像の分類に関する研究



北海道大学 大学院情報科学研究科
メディアネットワーク専攻

斉藤　直輝

2019 年 3 月

目 次

第 1 章　序論 **1**

1.1　本研究の背景 1

1.2　本研究の目的 2

1.3　本論文の構成 3

第 2 章　異種データを用いた画像分類に関する関連研究 **6**

2.1　はじめに 6

2.2　関連研究 6

2.2.1　特徴統合による画像分類 7

2.2.2　分類結果統合による画像分類 12

2.3　本論文で解決すべき問題 16

2.4　まとめ 18

第 3 章　分類結果の確信度推定を利用した画像分類 **19**

3.1　はじめに 19

3.2　特徴ベクトルの算出 19

3.3　位置座標を利用した画像分類 21

3.3.1　位置特徴ベクトルを用いた F*K*-NN アルゴリズムによる画像分類 21

3.3.2　確信度推定 22

3.3.3　分類候補のカテゴリ数の削減方法 24

3.4　画像およびテキストを利用した再分類 25

3.4.1　SVM を用いた分類 25

3.4.2 各カテゴリへの帰属確率からの特徴ベクトル算出 27
3.4.3 Multiple Annotator Logistic Regression モデル 27
3.5 実験 32
3.5.1 データセット 32
3.5.2 実験条件 33
3.5.3 実験結果および考察 35
3.6 まとめ 38

第 4 章 機械学習による確信度が高い分類結果の選択による画像分類 48
4.1 はじめに 48
4.2 分類結果の取得 48
4.2.1 撮影地点の位置座標を用いた分類 50
4.2.2 画像およびテキストタグを利用した分類 50
4.3 機械学習による確信度が高い分類結果の選択 51
4.4 実験 51
4.5 まとめ 54

第 5 章 帰属確率を入力とする分類器を導入した画像分類の高精度化 57
5.1 はじめに 57
5.2 分類手法 58
5.3 確信度が高い分類結果の存在の判定 59
5.3.1 確信度の算出 60
5.3.2 確信度が高い分類結果の存在判定 61
5.4 実験 62
5.5 まとめ 68

第 6 章 結論 69
6.1 本論文の成果 69
6.2 今後の課題 70

謝辞 **71**

参考文献 **72**

著者の研究業績 **79**

図 目 次

1.1 本研究における課題とその解決策および章構成の関係. 5

2.1 関連研究と本論文における提案手法の概要. 16

3.1 分類結果の確信度推定および統合処理を導入した画像分類手法の概要. 20
3.2 確信度 C に関する予備実験結果. 24
3.3 データセット A において提案手法により分類された画像例. . . . 38
3.4 位置特徴ベクトルを正しいカテゴリ毎にプロットした図. 46
3.5 画像およびテキストを利用した分類によって正確な分類を可能とした観光画像例：(a) および (b) の正しいカテゴリは Performances である. 46

4.1 機械学習により推定された確信度が高い分類結果の選択による画像分類手法の概要. 49
4.2 確からしい分類結果の選択処理の概要. 52
4.3 実験で用いる画像共有サービスに投稿された観光画像の一例. . . 54

5.1 確信度が高い分類結果の存在を考慮した画像分類手法の概要. . . 58
5.2 実験で利用するデータセットに含まれる観光画像の例. 64

表 目 次

3.1 データセット A における分類する観光カテゴリとデータセットに含まれる各カテゴリへ属する画像数. 33
3.2 データセット B における分類する観光カテゴリとデータセットに含まれる各カテゴリへ属する画像数. 33
3.3 本実験で用いる比較手法. 36
3.4 分類結果の確信度推定を導入した画像分類に関する実験における提案手法および比較手法による分類結果の再現率 (データセット A). 40
3.5 分類結果の確信度推定を導入した画像分類に関する実験における提案手法および比較手法による分類結果の適合率 (データセット A). 41
3.6 分類結果の確信度推定を導入した画像分類に関する実験における提案手法および比較手法による分類結果の F 値 (データセット A). 42
3.7 分類結果の確信度推定を導入した画像分類に関する実験における提案手法および比較手法による分類結果の再現率 (データセット B). 43
3.8 分類結果の確信度推定を導入した画像分類に関する実験における提案手法および比較手法による分類結果の適合率 (データセット B). 44
3.9 分類結果の確信度推定を導入した画像分類に関する実験における提案手法および比較手法による分類結果の F 値 (データセット B). 45
3.10 画像およびテキストを利用した分類による結果の F 値. 47

4.1 実験で用いる比較手法. 55
4.2 確信度が高い分類結果の選択による画像分類手法に関する実験における提案手法および比較手法による分類結果の再現率. 55
4.3 確信度が高い分類結果の選択による画像分類手法に関する実験における提案手法および比較手法による分類結果の適合率. 56
4.4 確信度が高い分類結果の選択による画像分類手法に関する実験における提案手法および比較手法による分類結果のF値. 56

5.1 各カテゴリの観光画像の枚数. 63
5.2 各カテゴリへの帰属確率を入力とする分類器を導入した画像分類手法に関する実験における提案手法および比較手法により算出された分類結果のF値. 67

第1章　序論

1.1　本研究の背景

情報通信技術の発達およびモバイル端末の普及により，Flickr[1]等の画像共有サービスの利用者数が増加している．サービスの利用者は，自らが撮影した画像を投稿するだけでなく，他の利用者により投稿された画像 (以降，投稿画像) を閲覧し，興味のある画像を共有することができる．しかしながら，投稿画像は多様かつ大量であり，利用者が興味のある画像を見つけ出すことは容易ではない．この問題の解決策として，投稿画像をカテゴリに分類することが挙げられる．カテゴリを手掛かりにして，利用者はあらかじめ分類された画像から興味のある画像を効率的に見つけ出すことができる．

ところで，近年の画像分類に関する研究では，画像に加えて，テキストなどのデータを同時に利用する手法が提案されている．画像分類において，異なる種類のデータ (以降，異種データ) を利用することで，単一のデータのみでは捉えることができない情報を画像分類に用いることが可能となるため，分類精度の向上が期待でき，異種データを利用した様々な画像分類手法が提案されている [1–3]．画像共有サービスにおいても，異種データの利用が単一のデータを用いた場合よりも分類精度の向上に効果的であることが報告されており，分類結果統合による分類手法が提案されている [4–6]．画像共有サービスには，画像とともにその内容を表すテキストが投稿され，画像データには撮影地点の位置座標も含まれていることから，これらの異種データに対する分類結果統合により，画像をカテゴリに分類する手法が提案されている．しかしながら，従来の統合手法では，各データから得られる分類の結果のみに注目して統合を行っており，各々の分類結果の

[1] https://www.flickr.com/

確からしさを検討するに至っておらず，データによっては誤った結果を出力していたとしても，他のデータと同様に扱うことから，統合して得られる最終的な分類結果の精度向上には限界があった．

1.2　本研究の目的

そこで，本論文では，従来手法の精度限界を打破するため，各データから得られる分類結果が最終結果の精度向上に貢献する度合い，すなわち，分類結果の確からしさを評価する指標 (以降，確信度) を導入した画像のカテゴリ分類手法を提案する．画像を複数のカテゴリの中から一つの最適なカテゴリへ分類する問題において，分類器から出力される各カテゴリへの帰属確率は，分類結果の確からしさとして用いることが可能であることが報告されている．そこで，提案手法では，各データを用いて構築された分類器から分類結果とともに出力される各カテゴリへの帰属確率を利用して確信度を算出し，この指標が最も高い結果を最終的な分類結果とする．これにより，画像共有サービスに投稿される画像毎に，分類に用いるデータの選択が可能となり，従来の手法と比較して高精度な画像分類が実現可能となる．

さらに本論文では，全てのデータによる分類結果の整合性とその確信度に注目し，確信度が高い分類結果を選択する手法では正確な分類が困難であると判断された場合の最終分類結果の取得手法を導入する．具体的に，提案手法では全ての分類結果の確信度が低い場合，もしくは，確信度が高い結果が複数存在しているにも関わらず，これらの結果として得られたカテゴリが一致していないため整合性が低い場合に，確信度が高い分類結果の選択による高精度な分類の実現が困難であると判断する．さらに，提案手法では，全てのデータから得られた各カテゴリへの帰属確率と正しいカテゴリとの間には関連性が存在することに注目し，これらの帰属確率を入力とする分類器により，投稿画像を最終的なカテゴリへ分類する．これにより，確信度が高い結果を選択する手法では高精度な分類が困難である場合においても，正確に最終的な分類結果を取得することが可能となる．以上に示すように，本論文では，画像共有サービスへ投稿される画像の高精度なカ

テゴリ分類を可能とする．上記の本研究における課題とその解決策および章構成の関係を図 1.1 に示す．

1.3　本論文の構成

本論文は 6 章からなり，以下に各章の概要を述べる．第 1 章では，本論文の研究背景および目的について述べた．

第 2 章では，関連研究として異種データを利用する画像分類手法を紹介し，本論文で解決すべき課題を明らかにする．

第 3 章では，分類を行う際に得られる各カテゴリへの帰属確率から導出される確信度を利用した画像分類手法について説明を行う．本手法では，分類の際に得られる各カテゴリへの帰属確率の大きさに注目して分類結果に対する確信度を推定する．さらに，推定された確信度があらかじめ設定したしきい値より大きい場合は，その結果の確からしさが高いと判断し，これを最終的な分類結果とする．一方で，確信度が低い場合は，他のデータを利用した分類手法による再分類を行う．これにより，分類精度向上に貢献するデータを利用した画像分類を実現する．また，実験により実際に画像共有サービスへ投稿された観光画像のカテゴリ分類を行うことで，確信度を利用する有効性を確認する．

第 4 章では，確信度が高い分類結果の選択を機械学習により可能とした画像のカテゴリ分類手法について説明を行う．あるカテゴリに属するサンプルが誤分類される場合，正解カテゴリと誤って分類されたカテゴリ間には，関連性が存在する [7]．そこで，本手法では，上記の誤分類に関する関連性を学習した識別器を利用することで，確信度を高精度に推定する方法を導入した画像分類手法を提案する．具体的には，各データから分類結果を取得する際に算出されるカテゴリへの帰属確率を要素とした特徴を算出し，Support Vector Machine による識別器を構築する．この識別器から出力された「分類結果が正しい」と識別される確率値を確信度として推定する．これにより，確信度推定に機械学習を導入することで，より高精度な画像分類の実現が期待される．さらに，本手法により，分類精度の向上が実際に可能となることを実験により確認する．

第 5 章では，確信度が高い結果を選択する手法では正確な分類が困難である場合を想定し，各カテゴリへの帰属確率を入力とする分類器を導入した画像分類手法について説明を行う．本手法では，すべての選択対象の分類結果の確信度が低い，すなわち確からしい結果が存在しない場合には，確信度が高い結果を選択する手法では，正確な分類が困難であると判断する．具体的に，上記の場合には，各カテゴリへの帰属確率を入力とする分類器を導入した画像分類手法により最終的な分類結果を取得する．さらに，本手法より，画像共有サービスへ投稿される画像の高精度なカテゴリ分類が実現可能であることを実際に画像共有サービスに投稿された画像を分類する実験により確認する．

最後に第 6 章で，結論として本論文の内容をまとめ，今後の課題について述べる．

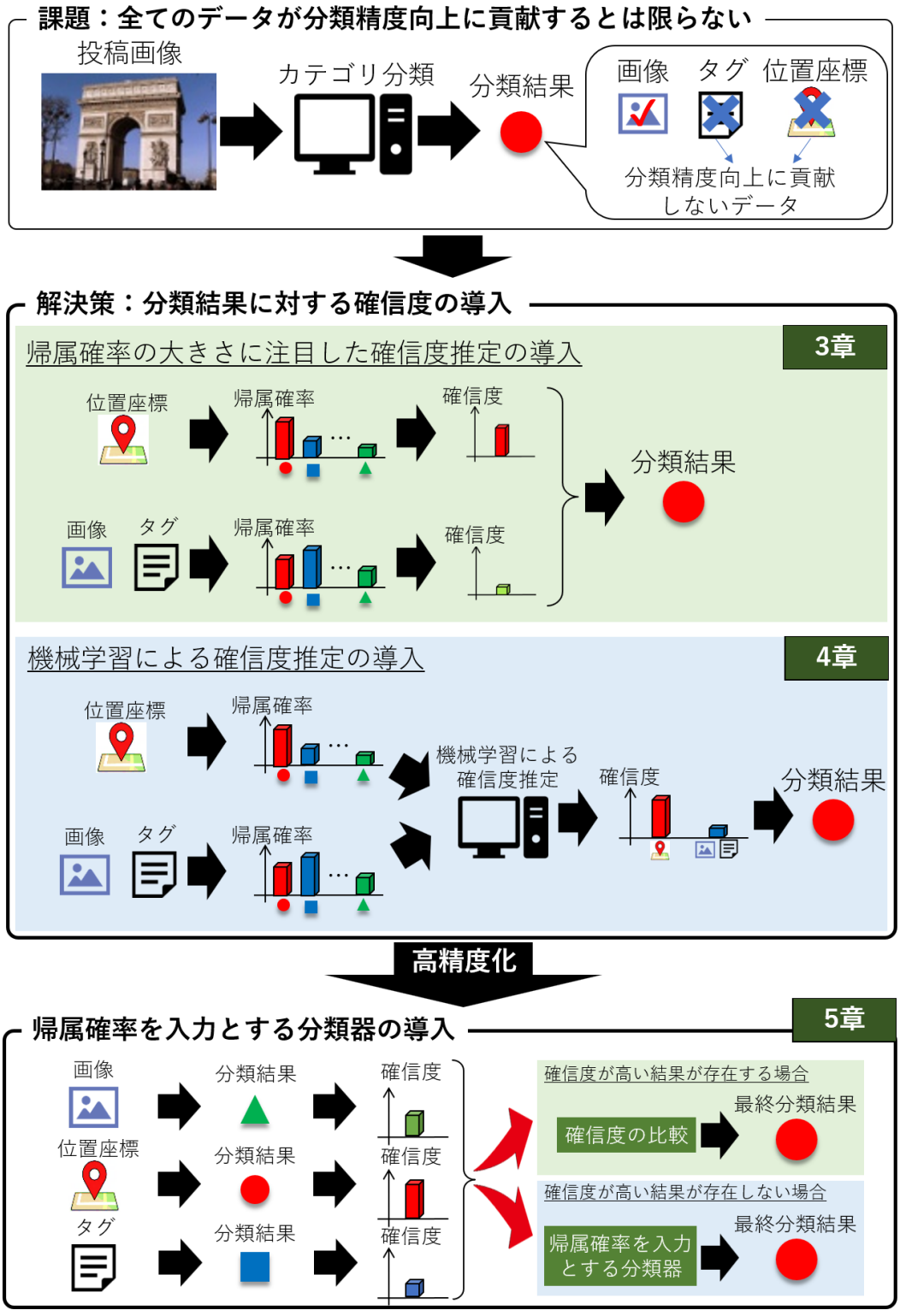


図 1.1: 本研究における課題とその解決策および章構成の関係.

第2章　異種データを用いた画像分類に関する関連研究

2.1　はじめに

本章では，本論文と関連する研究について説明する．まず，**2.2** 節において，異種データを用いて画像共有サービスに投稿される画像の分類に関する関連研究をまとめる．さらに，**2.3** 節において，本論文で解決すべき課題を明確にする．

2.2　関連研究

本論文では，異種データを利用して画像共有サービスに投稿された画像を分類する手法を提案している．異種データを利用した画像分類手法は，以下の 2 つに大別される．

特徴統合による分類
: 異種データから算出された特徴をひとつの特徴に統合し，これを分類器に入力することで最終的な分類結果を取得する方法．

分類結果統合による分類
: 異種データから算出された特徴を用いることで得られた複数の分類結果をひとつの最終分類結果に統合する手法．

そこで，本節では，特徴統合による画像分類手法 (**2.2.1**) および分類結果統合による画像分類手法 (**2.2.2**) に関する関連研究をまとめる．なお，画像共有サービスに投稿された画像の分類に応用可能な研究について整理するため，以降で説明

する関連研究が対象とする画像は，画像共有サービスに投稿されたものに限らない．図 2.1 に，以下で説明される関連研究と提案手法の概要を示す．

2.2.1　特徴統合による画像分類

本節では，異種データを利用した画像分類に特徴統合を採用している関連研究について述べる．

文献 [1] (2008 年)

文献 [1] では，撮影された画像に関連しているイベントを撮影地点の位置座標および画像特徴を用いて分類する手法が提案されている．具体的にはまず，分類対象の画像から位置座標および画像に関する特徴を算出する．これらの特徴をマルチクラス AdaBoost により統合することで，分類結果を取得する．以上の分類処理により，単一の特徴を用いて分類を行う場合よりも，高精度な分類を実現している．

文献 [2] (2009 年)

文献 [2] では，画像共有サービスから取得した画像で構成される大規模なデータセットに対して，画像に撮影された観光名所を分類する手法が提案されている．この手法では，データセット内の画像から画像とテキストに関する特徴を算出し，これらの特徴を Multiple Kernel Learning (MKL) により統合しながら分類を行うことで，複数のデータを同時に用いて分類結果を取得する．

文献 [8] (2010 年)

文献 [8] では，自動的にタグが付与された画像から撮像されている観光名所やイベントを検出する手法が提案されている．この手法では，タグと画像に関する特徴の類似度を算出し，得られた類似度をそれぞれ単独で用いてグラフを構築する．さらに，これらのグラフを統合することで得られたグラフを用いてクラスタリングを行う．最後に観光名所のクラスタ群に対

して分類を行うことで，各クラスタに対応する観光名所を特定し，各画像に対する検出結果を取得する．

文献 [9] (2012 年)

文献 [9] では， Deep Boltzman Machine [10] を利用して異種データを統合することで統一的な特徴を取得する手法が提案されている．この手法では，異種データの空間上の確率密度を学習することによって，特徴を統合する．また，いくつかのデータが存在しない場合でも，特徴の統合を行うことが可能である．この文献では，自己符号化器などの他の深層学習による特徴統合手法と比較して，顕著に高精度な分類を実現可能であることを実験により示している．

文献 [3] (2015 年)

文献 [3] では，観光名所が撮影された画像をその観光名所毎に高精度に分類するための特徴量の算出手法が提案されている．この手法では，画像から抽出した色や形状，テクスチャに関する複数の特徴量をひとつに統合し，特徴選択を施すことで，分類に有効な特徴量を算出している．また実験により，実際にその特徴量を用いて，観光名所が撮影された画像の高精度な分類が実現可能であることを示している．

文献 [11] (2015 年)

文献 [11] では，位置座標と画像特徴を用いて，Flickr などの画像共有サービスに投稿された GPS 情報付きの画像を高精度に分類する手法が提案されている．この手法では，まず位置座標から算出される特徴量と画像特徴量を別々に算出する．これらの特徴量を用いて，近年注目されている画像分類手法である Convolutional Neural Network (CNN) により分類を行っており，高精度な分類を実現している．

文献 [12] (2015 年)

文献 [12] では，高精度な画像分類を実現するためのタグ特徴の算出方法が提案されている．具体的には，画像から算出される画像特徴とその撮影地

点の位置座標において，それぞれの特徴の近傍に存在する他の画像を取得する．このとき取得した画像に付与されているタグを用いて新たなタグ特徴を算出する．以上より，この手法では，画像特徴および撮影地点の位置座標に基づいて，高精度な分類を可能とするタグ特徴を算出する．

文献 [13] (2015 年)

文献 [13] では，Flickr に投稿された画像を用いて，公園やジム，病院など，その画像を撮影した土地の利用方法を自動で検出する手法が提案されている．この手法では，Flickr に投稿された画像から取得可能である画像特徴およびその画像が撮影された位置座標を用いて CNN により分類を行っている．さらに，土地の利用方法を分類する際，まず，投稿された画像を “indoor” もしくは “outdoor” に分類することでフィルタリングを施した後，その利用方法に関する分類を行う．上記の前処理により，この手法では，高精度な分類を実現可能としている．

文献 [14] (2015 年)

文献 [14] では，複数のデータを同時に用いて分類を行う手法のひとつである重み付きスコアレベル統合において，その重みを自動かつ適応的に決定する Adaptive Weighted Fusion Approach (AWFA) が提案されている．具体的には，訓練サンプルとテストサンプルから算出した各データの特徴ベクトル間の距離を用いて，各データにおける重みを決定する．この手法では，互いの利点を補完しながら，分類に用いるデータを効果的に統合することを可能とし，高精度な分類を実現している．

文献 [15] (2016 年)

文献 [15] では，辞書学習に基づいて異種データを統合することで，ひとつの特徴表現を取得し，これを用いて画像を分類する手法が提案されている．具体的には，各データに対して学習を行うことで得られた辞書を用いて，各々のデータに対応する潜在特徴を算出する．これらの潜在特徴をベクトル結合によりひとつの特徴へ統合し，これを分類に利用する．この手法で

算出する辞書は非常にコンパクトであるが，最終的な分類精度は，非常に高い．また，この文献では，マルチモーダル顔認識，マルチビュー顔認識，マルチビュー行動認識，およびマルチモーダルバイオメトリクス認識の 4 つのタスクへの適用を想定した実験を行い，提案した手法を上記のタスクへ応用可能であることを示している．

文献 [16] (2017 年)

文献 [16] では，画像，音声，および言語の 3 種類のデータを統合するための深層学習の学習方法が提案されている．具体的には，各データについて構築されたネットワーク間の Loss が最小となるように学習することで，異種データが協調的に統合された深層表現を得ることができる．この手法は，クロスモーダル検索や異種データ間の分類器の転移などのタスクに有効であることが，実験により示されている．

文献 [17] (2018 年)

文献 [17] では，自然画像の分類精度の向上を目的として，画像とテキストを同時に利用した分類手法が提案されている．この手法では，自然画像に付与されているテキストには，オブジェクトやシーンと関連性が強いセマンティックな情報が含まれていることに注目しており，単語表現と画像特徴を組み合わせて，CNN により分類結果を取得する．この文献では，提案した手法の有効性を検証するために 2 種類のデータセットについて実験を行った結果より，画像特徴のみによる分類よりも分類性能が優れていることが示されている．

文献 [18] (2018 年)

文献 [18] では，複数のセンサー情報を利用した土地被覆の分類手法が提案されている．具体的には，multimodal Latent Dirichlet Allocation (mmLDA) [19] を用いて SAR とマルチスペクトル衛星画像に共通する潜在モデルを学習し，得られた特徴を用いて土地被覆分類を行う．実験の結果より，単一の

データを利用して得られた分類結果と比較して，高精度な分類が実現可能であることが確認されている．

文献 [20] (2018 年)

文献 [20] では，様々なマルチモーダル情報を単一の出力に統合することを目的とした手法が提案されている．この手法では，非侵襲的診断や画像誘導手術などの医療画像の臨床応用に向けて，医療画像の統合を行っている．具体的には，医療画像融合に用いる係数の選択をパターン認識タスクとしてモデル化している．これにより，異なる情報を有する複数の医療画像を一枚の画像に集約し，これを用いて分類を行う．この文献では，実際の医療画像のデータセットを用いた実験を行い，最先端の手法と比較して，統合された画像を利用した分類が高精度であることを確認している．

文献 [21] (2018 年)

文献 [21] では，画像とテキストの統合モデルと分類モデルを同時に学習可能なネットワークを提案している．この文献で提案されている Multimodal Matching and Classification Network (MMC-Net) は，統合用のネットワークと分類用のネットワークから構成されている．このネットワークではまず，統合用のネットワークで画像とテキストの統合特徴を取得する．これにより得られた特徴を分類用のネットワークに入力することで画像分類結果を得る．この手法では，統合および分類の損失関数を同時に最小化することによって，高精度なマルチモーダル分類モデルの構築を実現している．

文献 [22] (2018 年)

文献 [22] では，高速かつ高精度なマルチモーダル画像分類を実現するために Extreme Learning Machine (ELM) [23] を利用した特徴統合および画像分類手法を提案している．この手法では，まず，ELM に基づく Sparse Autoencoder により，各データから画像分類に適した特徴を算出する．さらに，これらの特徴を統合するためのモデルを学習する．最後に統合された特徴を利用して，分類結果を取得する．この手法は，勾配ベースのマルチモーダル深

層学習よりも学習効率が優れている．また，MNIST，XRMB，および NUS の 3 種類のデータセットを利用した実験により，この手法が従来のマルチモーダル深層学習による分類モデルと比較して，収束が速く，分類性能の向上を実現していることが確認されている．

文献 [24] (2018 年)

文献 [24] では，分類性能の向上を実現するために，画像およびテキストを統合する新たなマルチモーダル画像分類手法が提案されている．この手法では，エンコーダにより符号化されたテキストを画像に埋め込むことで，分類精度向上に貢献する情報を有する新たな画像を構築する．この画像を利用して CNN により分類を行うことで，最終的な分類結果を取得する．この文献では，大規模なマルチモーダル分類用のデータセットを利用して，従来の特徴統合および分類結果統合によるマルチモーダル画像分類手法との分類性能の比較を行う実験の結果より，提案した手法の有効性を示している．

以上のように，異種データを用いて特徴統合により画像を分類する手法がこれまでに多く提案されている．特に近年では，深層学習が様々なタスクに対して，高精度な分類を実現していることから，深層学習を導入した特徴統合手法が多く提案されている．

2.2.2　分類結果統合による画像分類

本節では，異種データより得られた複数の分類結果を統合することで，画像分類を行う関連研究について述べる．

文献 [25] (2008 年)

文献 [25] では，教師なしの方法で観光名所の画像を検出する手法が提案されている．この手法では，画像共有サービスに投稿された画像から取得可能な画像，テキストおよび，撮影地点の位置座標を用いて撮影されている観光名所およびイベントに基づいて自動的にクラスタリングを行っている．具体的にはまず，位置座標付きの画像を画像共有サービスから取得した後，

類似した観光名所やイベント毎に画像をクラスタリングする．さらに，各クラスタを関連している観光名所やイベント毎に分類し，これらのクラスタを Wikipedia と教師なしでリンクさせることで，画像共有サービスに投稿された観光画像を検出し，それに関連する観光名所やイベントを高精度に取得することを可能としている．

文献 [4] (2008 年)

文献 [4] では，画像共有サービスに投稿された画像を関連するイベント毎に分類する手法が提案されている．具体的には，画像共有サービスに投稿された画像と，その画像が撮影された位置座標を用いて Google Earth により取得した衛星写真からそれぞれ画像特徴を算出した後，それぞれ単独で分類を行うことで得られた分類結果を相乗的に統合することで最終的な分類結果を取得する．投稿された画像が撮影された地点の衛星画像から得られた画像特徴は，イベントの分類に有効である場合があると考えられるため，この手法では，投稿された画像と衛星画像を同時に用いて分類を行うことで高精度なイベントの分類を可能としている．

文献 [5] (2009 年)

文献 [5] では，画像特徴およびメタデータに含まれる位置座標とテキストを同時に用いて，観光名所を撮影し，Flickr に投稿された画像をその観光名所毎に分類する手法が提案されている．この手法では，まず画像が撮影された地点の位置座標を用いて，分類対象の画像を解析した後，テキスト情報と画像特徴を用いて詳細な画像分類を行っている．

文献 [26] (2010 年)

文献 [26] では，複数の画像特徴をそれぞれ単独で用いて得られた分類結果を統合することで画像分類を行う手法が提案されている．具体的には，複数の分類結果を重み付き多数決により統合することで，最終的な分類結果を取得する．ここで，重み付き多数決における重みを各画像特徴による識別器の正解率とすることで，効果的な画像分類を実現している．

文献 [27] (2017 年)

文献 [27] では，ソーシャルネットワーク上に投稿された画像によって誘起される感情を分類するために，画像およびテキストを利用したマルチモーダル分類手法を提案している．この手法では，5 種類の分類器を組み合わせるベイジアンモデル平均化法に基づくアンサンブル学習手法を利用する．この文献では，ユニモーダル分類器の分類結果統合に基づくアンサンブル法が最も高精度な感情分類を実現可能であることを示している．また，画像特徴として深層学習に基づいて算出される特徴を利用することで，より高精度な分類が実現可能となることが示されている．

文献 [6] (2018 年)

文献 [6] では，画像とテキストの 2 種類のデータを用いた分類手法が提案されている．この手法では，画像とテキストの二重ニューラルネットワークを構築することで，訓練サンプルが少量しか存在しない場合でも，高精度な分類の実現を可能としている．具体的に，二重ニューラルネットワークは，画像およびテキストをそれぞれ利用した分類モデル (以降，サブモデル) で構成されている．これらのサブモデルを別々に訓練した後，これらのサブモデルから出力される分類結果を単純に統合するのではなく，協調的にこれらの結果を融合させることで，最終的な分類結果を取得する．また，この文献では，LabelMe と UIUC-Sports の 2 種類のデータセットを利用した実験を行うことで，提案した手法により高精度な分類が可能であることを示している．

文献 [28] (2018 年)

文献 [28] では，画像およびテキストから算出される特徴を用いて，Social Networking Service に投稿された画像を分類する手法が提案されている．この文献では，複数の特徴を統合するマルチモーダル学習アルゴリズムに基づいて，画像分類を行う．具体的には，Kernel Ridge Regression (KRR) [29]

に基づいて，画像およびテキストをそれぞれ用いて分類を行うモデルを構築する．これらのモデルを統合することで，共同モデルを構築する．この手法では画像特徴を用いた分類をテキスト特徴を用いた分類モデルによって補強することで高精度な分類を実現する．テスト画像は，上記の 2 種類の分類モデルから得られる分類結果を組み合わせることによって分類結果を取得する．この文献では，実世界のデータセットに対して広範な実験を行うことで，提案した手法の優位性を示している．

文献 [30] (2018 年)

文献 [30] では，細胞が撮像された画像の分類を局所特徴と深層学習に基づいて算出された特徴を組み合わせて分類する手法を提案している．具体的には，局所特徴として，Rotation Invariant Co-occurrence among Local Binary Patterns および Joint Motif Labels を用いる．また，深層学習により算出される特徴量として，CNN モデルの一つである VGG19 の中間層の出力値を用いる．最後に，これらの特徴を利用したランダムフォレストによる分類結果統合手法により分類結果を取得する．

文献 [31] (2018 年)

文献 [31] では，CNN の中間層から取得可能な深層特徴を統合することで，より高精度な画像分類を実現するための分類器を構築する手法を提案している．具体的にはまず，一般画像で事前に学習された CNN モデルを用いて中間層から深層特徴を抽出する．さらに，教師あり分類モデルを利用して，深層特徴毎に分類器を構築し，複数の分類結果を得る．この手法では，これらの分類結果を統合することで，最終的な分類結果を得る．この文献では，実験により，複数の深層特徴により得られた結果を統合することで，分類精度が向上することを示している．

以上のように，複数のデータを同時に用いて画像を分類する手法が種々提案されている．また，特徴統合による分類手法と同様に，近年は深層学習を導入した統

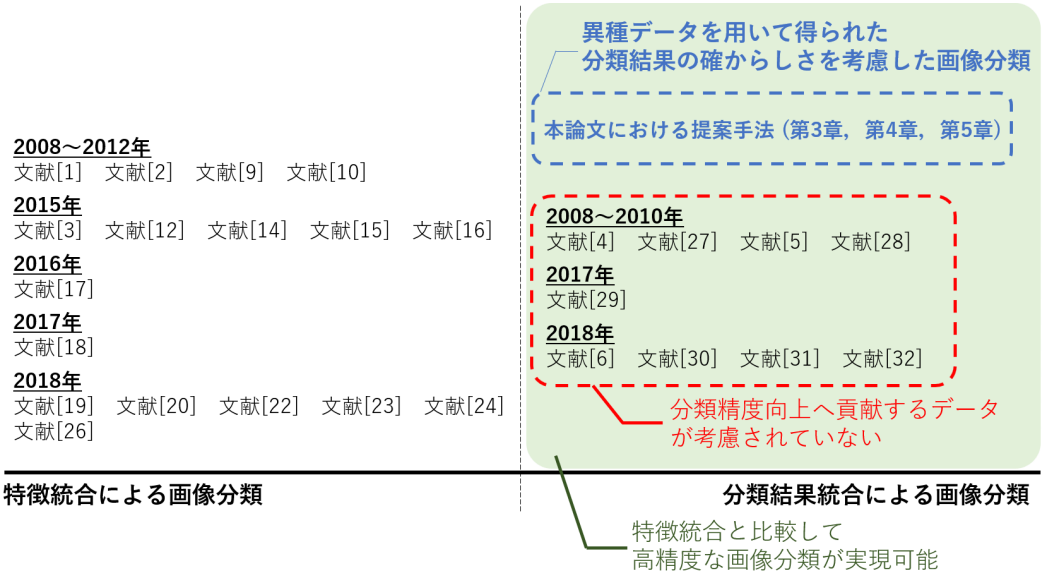


図 2.1: 関連研究と本論文における提案手法の概要.

合手法が多く提案されている.

2.3　本論文で解決すべき問題

本節では，異種データを用いて画像共有サービスに投稿される画像の高精度な分類を実現するために，本論文で解決すべき問題を明確にする．前節までに示したように，特徴統合および分類結果統合により，異種データを用いて画像を分類する従来手法が種々提案されている．特徴統合では，データ毎に取得した特徴をそれぞれの関連性を考慮しながら統合し，その分類を行う．しかしながら，特徴間において求められる関連性には，分類精度の向上に貢献しないものも含まれるため，統合後の特徴が分類に適していない場合がある．これにより，分類対象の画像によっては，正確な分類が困難な場合が存在する．一方，分類結果統合は，データ毎に構築された分類器から取得された結果より最終的な分類結果を選択するものであり，その精度は特徴統合よりも高いことが報告されている [32]．そこで，本論文では分類結果統合による異種データを利用した画像分類手法について

議論する．

画像共有サービスには，画像とともにその内容を表すテキストが投稿され，画像データには撮影地点の位置座標も含まれていることから，これらの異種データを画像分類に利用することが可能である．しかしながら，従来の分類結果統合では，各データから得られる分類の結果のみに注目して統合を行っており，各々の分類結果の確からしさを検討するに至っておらず，データによっては誤った結果を出力していたとしても，他のデータと同様に扱うことから，統合して得られる最終的な分類結果の精度向上には限界があった．例えば，観光画像をカテゴリに分類する場合において，撮影地点の位置座標を利用して得られた分類結果は，観光名所で撮影された画像とその名所の近隣の飲食店で提供された料理が撮影された画像を同じカテゴリへ分類してしまう．従来の分類結果統合による分類では，分類対象の画像によって，分類精度向上に貢献するデータが異なるものの，全てのデータより得られた結果を上記に示した性質を考慮せずに統合する．

そこで，本論文では，上記の問題を解決するため，各データから得られる分類結果が最終結果の精度向上に貢献する度合い，すなわち，分類結果の確からしさを評価する指標 (以降，確信度) を導入した画像のカテゴリ分類手法を提案する．画像を複数のカテゴリの中から一つの最適なカテゴリへ分類する問題において，分類器から出力される各カテゴリへの帰属確率は，分類結果の確からしさとして用いることが可能であることが報告されている．そこで，提案手法では，各データを用いて構築された分類器から分類結果とともに出力される各カテゴリへの帰属確率を利用して確信度を算出し，この指標が最も高い結果を最終的な分類結果とする．これにより，画像共有サービスに投稿される画像毎に，分類に用いるデータの選択が可能となり，従来の手法と比較して高精度な画像分類が実現可能となる．さらに，本論文では，全てのデータによる分類結果の整合性とその確信度に注目し，確信度が高い分類結果を選択する手法では正確な分類が困難であると判断された場合の最終分類結果の取得手法を導入する．具体的に，提案手法では全ての分類結果の確信度が低い場合，もしくは，確信度が高い結果が複数存在しているにも関わらず，これらの結果として得られたカテゴリが一致していない

ため整合性が低い場合に，確信度が高い分類結果の選択による高精度な分類の実現が困難であると判断する．さらに，提案手法では，全てのデータから得られた各カテゴリへの帰属確率と正しいカテゴリとの間には関連性が存在することに注目し，これらの帰属確率を入力とする分類器により，投稿画像を最終的なカテゴリへ分類する．これにより，確信度が高い結果を選択する手法では高精度な分類が困難である場合においても，正確に最終的な分類結果を取得することが可能となる．以上に示すように，本論文では，分類結果統合による異種データを利用した画像分類手法に存在する課題を解決し，画像共有サービスへ投稿される画像の高精度な分類を可能とする．

2.4　まとめ

本章では，本論文と関連する研究について説明した．また，関連研究における問題点について述べ，画像共有サービスに投稿される画像の高精度な分類を実現するために，本論文で解決すべき課題を明確にした．

第3章　分類結果の確信度推定を利用した画像分類

3.1　はじめに

本章では，異種データから得られた分類結果の確信度推定を利用した画像分類手法について提案する．提案手法の概要を図 3.1 に示す．提案手法ではまず，位置座標を用いた F*K*-NN アルゴリズムにより分類を行い，得られた結果の確からしさを確信度として推定する．ここで，確信度が予め設定したしきい値よりも高い場合，得られた結果が正しい確率が高いと判断し，この分類結果を最終的な分類結果として出力する．一方，確信度がしきい値よりも低い場合，誤分類が発生している可能性があると判断し，テキストおよび撮影地点の位置座標を用いた分類を行うことで，最終的な分類結果を取得する．以上の処理により，提案手法では，異種データが分類精度向上に貢献する度合いを考慮することで，高精度な画像のカテゴリ分類を実現可能とする．以降，**3.2** 節で，提案手法で分類を行うために利用する特徴ベクトルの算出方法を示す．さらに，**3.3** 節で，位置座標を利用した分類および確信度推定について説明し，**3.4** 節で画像およびテキストを利用した分類手法について説明する．また，**3.5** 節で提案手法の有効性を検証するために行った，実際に画像共有サービスへ投稿された観光画像をカテゴリへ分類する実験の結果を示し，最後に **3.6** で本章をまとめる．

3.2　特徴ベクトルの算出

本節では，提案手法において，データ毎に分類結果を取得するために利用する特徴ベクトルの算出方法について説明する．提案手法では，投稿画像とそのメタ

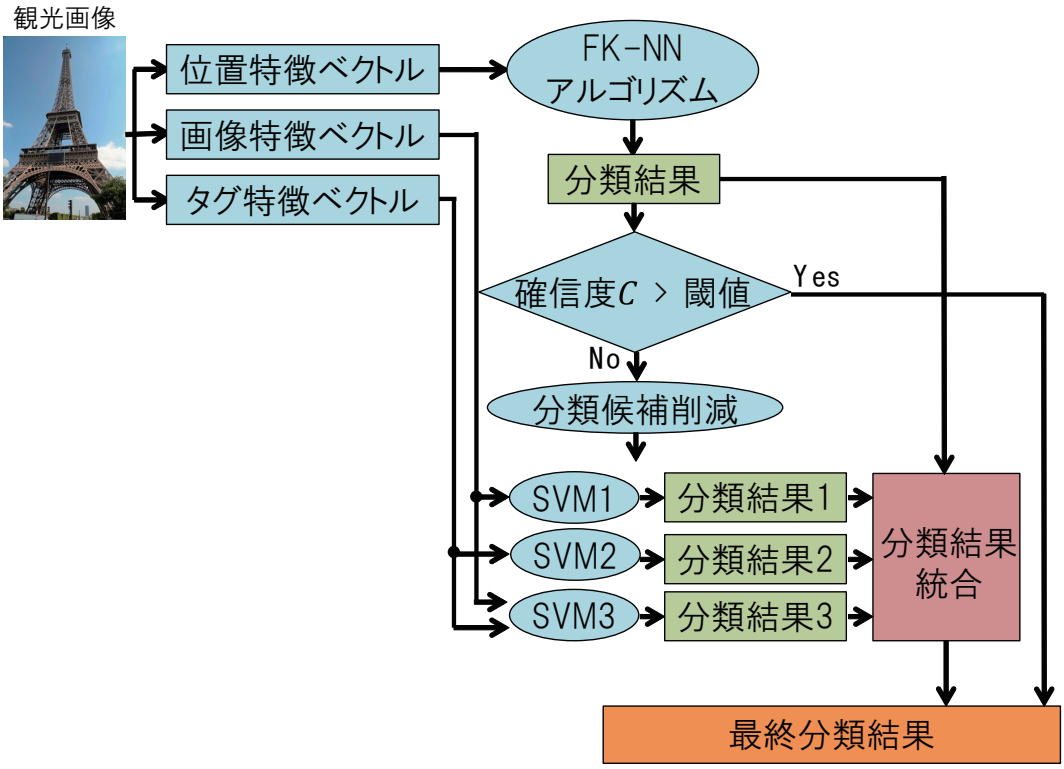


図 3.1: 分類結果の確信度推定および統合処理を導入した画像分類手法の概要.

データである撮影地点の位置座標およびテキストから，画像特徴ベクトル，位置特徴ベクトル，テキスト特徴ベクトル，および上記の 3 種類の特徴ベクトルを結合した統合特徴ベクトルを算出する．まず，画像特徴ベクトルは，投稿画像より Speed Up Robust Feature (SURF) 特徴 [33] を算出し，得られる複数の特徴量から Bag-of-Features ベクトル [34] を求めることで算出する．さらに位置特徴ベクトルは，撮影地点の位置座標から取得できる緯度と経度の 2 つの値を要素に持つベクトルとして算出する．テキスト特徴ベクトルは，各々の単語が各要素に対応するベクトルを用意し，投稿画像に付与されているテキストに存在している単語に対応する要素については 1，それ以外の要素を 0 とすることで取得する．最後に統合特徴ベクトルは，画像特徴ベクトルおよびテキスト特徴ベクトルを一つのベクトルに結合し，このベクトルに対して minimum-Redundancy-Maximum-Relevance (mRMR) アルゴリズム [35] による特徴選択を適用することで選択され

た特徴により構成される．提案手法では，以上の 4 種類の特徴ベクトルを用いて，画像分類を行う．

3.3　位置座標を利用した画像分類

本節では，提案手法の位置座標を利用した画像分類手法について説明する．提案手法では，位置座標を用いた FK-NN アルゴリズムによる分類を行い，得られた結果の確からしさを確信度として推定する．ここで，確信度が予め設定したしきい値よりも高い場合，得られた結果が正しい確率が高いと判断し，この分類結果を最終的な分類結果として出力する．一方，確信度がしきい値よりも低い場合，誤分類が発生している可能性があると判断し，再分類を行う．このとき，推定した確信度に基づいて分類候補のクラス数を削減することで，再分類を簡単化し，より高精度な分類の実現が可能となる．以降，**3.3.1** で，位置特徴ベクトルを用いた FK-NN アルゴリズムによる画像分類を示し，**3.3.2** で分類結果の確信度の推定方法について示す．また，**3.3.3** で分類候補のカテゴリ数削減の方法について示す．

3.3.1　位置特徴ベクトルを用いた FK-NN アルゴリズムによる画像分類

提案手法ではまず，位置特徴ベクトルを用いて FK-NN アルゴリズムにより画像共有サービスに投稿された画像の分類を行う．FK-NN アルゴリズムは，特徴空間において，テストサンプルの近傍に存在する訓練サンプルが属するカテゴリに基づいて分類を行う手法である [36]．分類結果を取得する際に算出される各カテゴリへの帰属確率を確信度推定に利用可能であるため，提案手法では，FK-NN アルゴリズムを撮影地点の位置座標を利用した分類手法として採用した．

提案手法では，i ($= 1, 2, \ldots, N$;N は訓練サンプル数) 番目の訓練サンプルにおける分類候補のカテゴリ j($\in \{1, 2, ..., J\}$;J は分類候補のカテゴリ数) への帰属確率

$u_{ij}(i = 1, 2, \ldots, N; j = 1, 2, \ldots, J)$ を以下の通りに算出する．

$$u_{ij} = \begin{cases} 0.51 + (n_{ij}/K) \cdot 0.49 & \text{if } j = g_i \\ (n_{ij}/K) \cdot 0.49 & \text{otherwise} \end{cases} \tag{3.1}$$

ただし，K は近傍数を表し，$n_{ij}(\leq K)$ は i 番目の訓練サンプルの近傍に存在する K 個の訓練サンプルのうち，カテゴリ j に属するサンプル数を表す．また，g_i は i 番目の訓練サンプルが属するカテゴリを表す.

以上で算出された各カテゴリへの帰属確率 u_{ij} およびテストサンプルの特徴ベクトル $\hat{\boldsymbol{x}}$ を用いて，以下の通りにテストサンプルの帰属確率 $u_j(\hat{\boldsymbol{x}})$ を得る．

$$u_j(\hat{\boldsymbol{x}}) = \frac{\sum_{k=1}^{K} \bar{u}_{kj}(1/\|\hat{\boldsymbol{x}} - \boldsymbol{x}_k\|^2)}{\sum_{k'=1}^{K}(1/\|\hat{\boldsymbol{x}} - \boldsymbol{x}_{k'}\|^2)} \ (j = 1, 2, ..., J) \tag{3.2}$$

ただし，$\bar{u}_{kj}$ は，テストサンプルの K 近傍に存在する訓練サンプルが属するカテゴリ j への帰属確率を表す．また，$\boldsymbol{x}_k$ は，K 近傍の訓練サンプルの特徴ベクトルである．以上の FK-NN アルゴリズムにより，提案手法では，帰属確率が最も高いカテゴリを分類結果として得る．

3.3.2　確信度推定

提案手法では，**3.3.1** の FK-NN アルゴリズムにより得られた分類結果の確信度を推定する．具体的には，FK-NN アルゴリズムにより得られた分類結果の確信度を次式の通りに推定する．

$$C = \max_j u_j(\hat{\boldsymbol{x}}) \tag{3.3}$$

以上の確信度推定により，提案手法では，撮影地点の位置座標を利用した分類結果の確からしさを定量的に評価することが可能となる．

以降では，分類結果の確信度の有効性を検証するために行った予備実験の結果について説明する．本実験では，Flickr[1] へ投稿された観光画像 49760 枚に対して，撮影地点の位置座標を用いた FK-NN アルゴリズムによる分類を行う．このとき，分類対象のカテゴリは，観光名所に関するカテゴリ (観光カテゴリ) とした．本予備実験では，以上で得られた分類結果の確信度を算出する．ただし，本実験で用いるデータセットの詳細は **3.5** に詳しい．また，FK-NN アルゴリズムによる分類で利用する近傍数 K として，5，50，および 100 の 3 種類を利用した．さらに，分類結果の確信度が予め決定したしきい値よりも大きい場合，分類結果は正しいと判定することで，分類結果の正誤を判定した．予備実験では，しきい値 T を変化させながら分類結果の正誤を判定し，以下に示す判定結果の正解率を評価した．

$$\text{正解率} = \frac{\text{正しく判定された画像数}}{\text{テスト画像数}} \tag{3.4}$$

予備実験により得られた，確信度による正誤判定のためのしきい値を変化させた際の正解率の変動の様子を図 3.2 に示す．以降では，図 3.2 に示した結果における，しきい値 T の変化による正解率への影響について，以下の 3 つの場合に分けて説明する．

(i) 確信度のしきい値が適切な値よりも低い場合,「誤分類された結果」を「正しく分類された結果」と判定してしまう場合が多くなるため，正誤判定の正解率は向上しない．

(ii) 確信度のしきい値を適切に設定することで，高精度な分類結果の正誤判定が実現可能となる．

(iii) 確信度のしきい値が適切な値よりも高い場合,「正しく分類された結果」であるにも関わらず,「誤分類された結果」として判定されてしまう場合が多くなるため，正誤判定の正解率が低下する．

図 3.2 より，近傍数 $K = 5$ の FK-NN アルゴリズムによる結果に注目すると，確

[1] https://www.flickr.com/

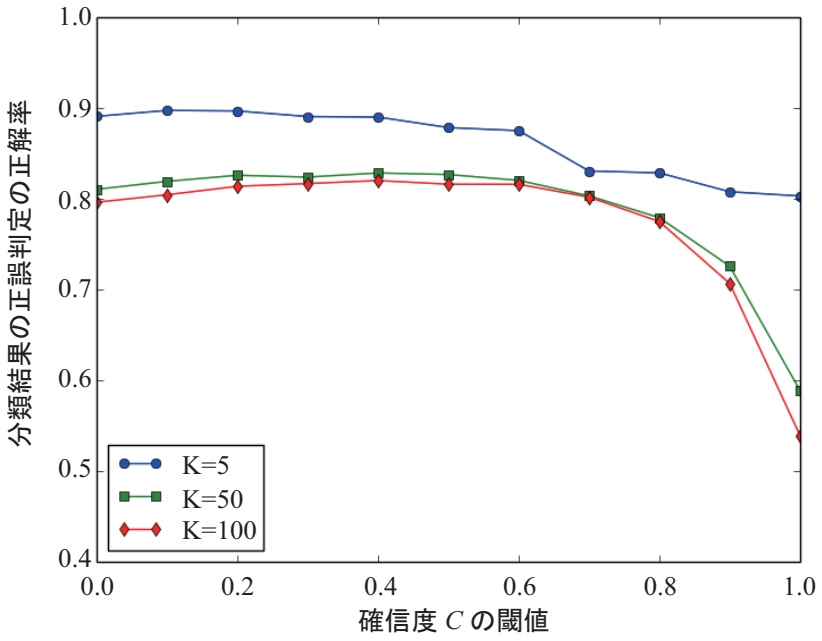


図 3.2: 確信度 C に関する予備実験結果.

信度 C のしきい値 T を適切に設定することで，分類結果の正誤判定の正解率が 0.903 になることが確認できる．以上より，提案手法では確信度のしきい値を適切に設定することで，分類結果の確からしさを正確に評価することが可能となる．

提案手法では，推定された確信度 C が予め設定したしきい値より高い場合，位置座標を用いた分類により得られた結果を最終分類結果とする．一方，確信度がしきい値より低い場合，推定された確信度を用いて分類候補のカテゴリ数を削減することで，分類問題を簡単化した後に，画像およびテキストを用いた分類により最終結果を得る．

3.3.3　分類候補のカテゴリ数の削減方法

提案手法では，帰属確率が低いカテゴリは分類結果として取得する正しいカテゴリの候補ではないと考え，帰属確率が予め設定したしきい値よりも低いカテゴリには分類しないように，分類候補から除外する．さらに提案手法では，テスト

サンプルと各カテゴリに属する訓練サンプル間の距離に注目し，この距離の最小値がしきい値より大きいカテゴリを分類候補から除外する．具体的には，以下に示す通り，新たな分類候補カテゴリの集合 Q を定義する．

$$Q = \{j | u_j(\hat{\boldsymbol{x}}) > h\} \tag{3.5}$$

ただし，h は，分類候補カテゴリ数の削減方法において，帰属確率について予め設定するしきい値である．以上により，提案手法では分類問題を簡単化し，画像およびテキストを利用した画像分類を行うことで，最終的な分類結果を取得する．

3.4　画像およびテキストを利用した再分類

本節では，提案手法における画像およびテキストを利用した再分類について示す．提案手法では，画像特徴ベクトル，テキスト特徴ベクトル，統合特徴ベクトルをそれぞれ用いて Support Vector Machine (SVM) による分類結果から，これらの分類結果の精度を考慮することで，最終分類結果を算出する．以降，**3.4.1** で，SVM による分類手法を示し，**3.4.2** で画像およびテキストを利用した分類に用いる特徴ベクトルについて説明する．また，**3.4.3** で Multiple Annotator Logistic Regression (MALR) モデル [37] による分類手法について示す．

3.4.1　SVM を用いた分類

本節では，画像特徴ベクトル，タグ特徴ベクトル，および統合特徴ベクトルを用いた分類手法について説明する．提案手法では，上記に示した 3 種類の特徴ベクトルをそれぞれ単独で用いて Support Vector Machine (SVM) [38] により分類を行うことで，3 種類の分類結果を取得する．

SVM は，訓練データとなる特徴ベクトル $\boldsymbol{x}_i$ (i=1, 2, ... , N; N は訓練データ数) とこれに割り当てられたクラスラベル y_i から，テストデータの特徴ベクトル $\boldsymbol{x}$ のクラスを識別するためのアルゴリズムである．このとき，識別超平面をマー

ジンが最大となるように算出することで，高い汎化性を持つことが一般的に知られている．

SVM では，次式に示される識別超平面における $\boldsymbol{\zeta}$ と b をマージンが最大となるように設定する．

$$f^{\mathrm{svm}}(\boldsymbol{x}) = \boldsymbol{\zeta} \cdot \boldsymbol{x} + b \tag{3.6}$$

ここで，識別超平面のマージンは次式で算出される．

$$\mathrm{Margin} = \frac{y_i(\boldsymbol{\zeta}^{\mathrm{T}} \cdot \boldsymbol{x}_i + b)}{\|\boldsymbol{\zeta}\|} \tag{3.7}$$

したがって，上記のマージンを最大とする $\boldsymbol{\zeta}$，b を求める問題は，以下の制約付き最適化問題を解くことに帰着される．

$$\begin{aligned} &\arg\min_{\boldsymbol{\zeta},b} \frac{1}{2}\|\boldsymbol{\zeta}\|^2 \\ &s.t.\ \ y_i(\boldsymbol{\zeta} \cdot \boldsymbol{x} + b) \geq, i = 1, 2, ..., N \end{aligned} \tag{3.8}$$

ここで，ラグランジュ乗数 $\boldsymbol{a}$ $(=[a_1, a_2, ..., a_N]^{\mathrm{T}})$ を定義しラグランジュの未定乗数法を用いると式 (3.8) は以下の目的関数を $\boldsymbol{a}$ について最大化する問題に帰着される．

$$L(\tilde{\boldsymbol{a}}) = \sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i',j=1}^{N} a_{i'} a_j y_{i'} y_j \boldsymbol{x}_{i'}^{\mathrm{T}} \boldsymbol{x}_j \tag{3.9}$$

ただし，$\boldsymbol{a}$ は以下の制約条件を満たす．

$$\begin{aligned} &\sum_{i=1}^{N} a_i y_i = 0 \\ &0 \leq a_i, i = 1, 2, ..., N \end{aligned} \tag{3.10}$$

以上より識別超平面 $f^{\mathrm{SVM}}(\boldsymbol{x})$ は次式で表される．

$$f^{\mathrm{SVM}}(\boldsymbol{x}) = \sum_{i=1}^{N} a_i y_i \boldsymbol{x}_i \cdot \boldsymbol{x} + b \tag{3.11}$$

ここで，特徴ベクトル $\boldsymbol{x}_i$ を非線形写像 $\phi(\boldsymbol{x}_i)$ により変換し，その空間上で SVM を適用することで次式の通りに非線形分離が可能となる．

$$f^{\mathrm{SVM}}(\boldsymbol{x}) = \sum_{i=1}^{N} a_i y_i K^{\mathrm{SVM}}(\boldsymbol{x}_i, \boldsymbol{x}) + b \tag{3.12}$$

ただし，カーネル関数 $K^{\mathrm{SVM}}(\boldsymbol{x}_i, \boldsymbol{x}) = \phi(\boldsymbol{x}_i) \cdot \phi(\boldsymbol{x})$ で表される．

3.4.2　各カテゴリへの帰属確率からの特徴ベクトル算出

提案手法では，分類器により分類結果を取得する際に算出可能な各カテゴリへの帰属確率から特徴ベクトル $\boldsymbol{z}$ を算出する．まず，文献 [39] に基づいて，前節で算出した SVM により得られる各カテゴリへの帰属確率を算出する．次に，これらの帰属確率 ($3J$ 次元) および近傍数 $K = \{5, 10, 15, 20, 25, 50, 100\}$ の FK-NN アルゴリズムにより取得された帰属確率 ($7J$ 次元) を並べたベクトル ($10J$ 次元) を取得する．このベクトルに対して mRMR アルゴリズムによる特徴選択を適用することで特徴ベクトル $\boldsymbol{z}$ を得る．

3.4.3　Multiple Annotator Logistic Regression モデル

投稿画像およびテキストを利用した分類では，MALR モデルに基づいて最終的な分類結果を取得する．MALR モデルでは，以下の多クラスロジスティック回帰を利用して各カテゴリへの帰属確率 $P(j|\boldsymbol{z}, \boldsymbol{W})$ を算出し，最も帰属確率が高いカテゴリを最終分類結果とする．

$$P(j|\boldsymbol{z}, \boldsymbol{W}) = \frac{\exp(\boldsymbol{w}_j^{\mathrm{T}} \boldsymbol{z})}{\sum_{j'=1}^{J} \exp(\boldsymbol{w}_{j'}^{\mathrm{T}} \boldsymbol{z})} \tag{3.13}$$

ただし，$\boldsymbol{W}(=[\boldsymbol{w}_j]_{j=1}^{J};\boldsymbol{w}_j$ は重みベクトル) は，重み行列を示す．分類器 $r(\in\{V,T,F\};V,T,F$ はそれぞれ画像特徴ベクトルによる分類器，テキスト特徴ベクトルによる分類器，および統合特徴ベクトルによる分類器) において，$y_i^{(r)}$ を i 番目の訓練サンプルを分類した際に得られた分類結果とする．さらに，$s_i^{(r)}$ を $y_i^{(r)}$ が正しいカテゴリと一致すれば 1，そうでなければ 0 となる値とするとき，訓練データセット $\mathcal{D}$ $(=[y_i^{(V)},y_i^{(T)},y_i^{(F)},s_i^{(V)},s_i^{(T)},s_i^{(F)},\boldsymbol{z}_i]_{i=1}^{N})$ が与えられたときの尤度 $P(\mathcal{D}|\boldsymbol{W})$ を最大化することで重み行列 $\boldsymbol{W}$ を求める．具体的に，尤度 $P(\mathcal{D}|\boldsymbol{W})$ は次式により定義される．

$$P(\mathcal{D}|\boldsymbol{W})=\prod_{i=1}^{N}P(\mathcal{L}_i|\boldsymbol{z}_i,\boldsymbol{W}) \tag{3.14}$$

ただし，$\mathcal{L}_i=[y_i^{(V)},y_i^{(T)},y_i^{(F)},s_i^{(V)},s_i^{(T)},s_i^{(F)}]$ とする．ここで，π_r を分類器 r を用いた訓練サンプルの交差検定を行うことで得られる正解率とするとき，尤度 $P(\mathcal{D}|\boldsymbol{W})$ は以下のように書き換えられる．

$$P(\mathcal{D}|\boldsymbol{W})=\prod_{i=1}^{N}\prod_{r\in\{V,T,F\}}\left[P(y_i^{(r)}|\boldsymbol{z}_i,s_i^{(r)},\boldsymbol{W})\quad\cdot P(s_i^{(r)}|\pi_r)\right] \tag{3.15}$$

ただし，各分類器は互いに独立であると仮定し，$P(y_i^{(r)}|\boldsymbol{z}_i,s_i^{(r)},\boldsymbol{W})$ および $P(s_i^{(r)}|\pi_r)$ を以下のように定義する．

$$P(y_i^{(r)}|\boldsymbol{z}_i,s_i^{(r)},\boldsymbol{W})=\left[P(y_i^{(r)}|\boldsymbol{z},\boldsymbol{W})^{s_i^{(r)}}\cdot\left(\frac{1}{J}\right)^{1-s_i^{(r)}}\right] \tag{3.16}$$

$$P(s_i^{(r)}|\pi_r)=(\pi_r)^{s_i^{(r)}}\cdot(1-\pi_r)^{1-s_i^{(r)}} \tag{3.17}$$

式 (3.16) および (3.17) より，式 (3.15) は次式のように書き換えられる．

$$P(\mathcal{D}|\boldsymbol{W}) = \prod_{i=1}^{N} \prod_{r\in\{V,T,F\}} \{\pi_r P(y_i^{(r)}|\boldsymbol{z},\boldsymbol{W})\}^{s_i^{(r)}} \cdot \{\frac{1}{J}(1-\pi_r)\}^{1-s_i^{(r)}} \tag{3.18}$$

したがって，訓練サンプルの対数尤度は，以下の通りとなる．

$$\ln P(\mathcal{D}|\boldsymbol{W}) = \sum_{i=1}^{N} \sum_{r\in\{V,T,F\}} \left[s_i^{(r)} \ln\{\pi_r P(y_i^{(r)}|\boldsymbol{z},\boldsymbol{W})\} + (1-s_i^{(r)}) \ln\{\frac{1}{J}(1-\pi_r)\} \right] \tag{3.19}$$

ただし，式 (3.19) の第 2 項は省略可能であるため，次式の通りに書き換えられる．

$$\ln P(\mathcal{D}|\boldsymbol{W}) = \sum_{i=1}^{N} \sum_{r\in\{V,T,F\}} \left[\gamma_i^{(r)} \ln\{\pi_r P(y_i^{(r)}|\boldsymbol{z}),\boldsymbol{W})\} \right] \tag{3.20}$$

提案手法では，重み行列 $\boldsymbol{W}_{\mathrm{ML}}$ を以下の式を解くことで求める．

$$\boldsymbol{W}_{\mathrm{ML}} = \arg\max_{\boldsymbol{W}} \left[\ln P(\mathcal{D}|\boldsymbol{W}) - \lambda \|\boldsymbol{W}\|_{\mathrm{F}}^2 \right] \tag{3.21}$$

ただし，第 2 項は，正則化項を表し，$\lambda\ (\geq 0)$ は正則化パラメータ，$\|\cdot\|_{\mathrm{F}}$ はフロベニウスノルムを表す．

さらに提案手法では，文献 [40] に基づいて，訓練サンプル数の不均衡性を考慮するため，行列 $\boldsymbol{W}_{ML}$ を以下の最適化問題を解くことで求める．

$$\boldsymbol{W}_{\mathrm{ML}} = \arg\max_{\boldsymbol{W}} \left[\ln P(\mathcal{D}|\boldsymbol{W}) + N \cdot \zeta \sum_{k=1}^{K} E_k - \lambda \|\boldsymbol{W}\|_F^2 \right] \tag{3.22}$$

ただし，$\zeta\ (\in [0,1])$ は不均衡を考慮する項の重みを示す．さらに，L_k をカテゴリ k に属する訓練サンプルの特徴ベクトル $\boldsymbol{z}_i$ の集合とし，$|L_k|$ を集合 L_k に含まれる

要素数とするとき，E_k は次式で定義される．

$$E_k = \frac{1}{|L_k|} \sum_{\boldsymbol{z}_i \in L_k} P(k|\phi(\boldsymbol{z}_i), \boldsymbol{W}) \tag{3.23}$$

ここで，式 (3.22) は直接解くことが困難であるため，Limited memory-Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) 法 [41] に基づく準ニュートン法により最適化を行う．具体的には，次式の通り，τ ステップ目の更新により得られる重み行列 $\boldsymbol{W}_\tau$ (= $[\boldsymbol{w}_1^{(\tau)}, \boldsymbol{w}_2^{(\tau)}, ..., \boldsymbol{w}_J^{(\tau)}] \in \mathbb{R}^{d_{\boldsymbol{z}} \times J}; \boldsymbol{w}_j^{(\tau)} \in \mathbb{R}^{d_{\boldsymbol{z}}}; d_{\boldsymbol{z}}$ は特徴ベクトル $\boldsymbol{z}$ の次元数) を算出する．

$$\boldsymbol{W}_\tau \leftarrow \boldsymbol{W}_{\tau-1} + \eta \boldsymbol{H}_{\tau-1}^{-1} \boldsymbol{G}_{\tau-1} \tag{3.24}$$

ただし，η はステップ幅を表す．また，$\boldsymbol{G}_\tau$ (=$[\boldsymbol{g}_1^{(\tau)}, \boldsymbol{g}_2^{(\tau)}, \cdots, \boldsymbol{g}_J^{(\tau)}] \in \mathbb{R}^{d_{\boldsymbol{z}} \times J}; \boldsymbol{g}_j^{(\tau)}$ は τ ステップ目の更新で得られるカテゴリ j に対応する勾配ベクトル) は τ ステップ目の更新により得られる勾配行列を示し，$\boldsymbol{g}_j^{(\tau)}$ ($\in \mathbb{R}^{d_{\boldsymbol{z}}}$) は，以下の通りに定義される．

$$\boldsymbol{g}_j^{(\tau)} = \sum_{i=1}^{N} \sum_{r=\{V,T,F\}} \left[\gamma_i^{(r)} \left\{ a_{ij}^{(r)} - P(j|\boldsymbol{z}_i, \boldsymbol{W}_\tau) \right\} \boldsymbol{z}_i \right] \cdot P(l|\boldsymbol{z}_i, \boldsymbol{W}_\tau) \boldsymbol{z}_i \Big\} - 2\lambda \boldsymbol{w}_j^{(\tau)} \Bigg] \tag{3.25}$$

ただし，$a_{ij}^{(r)}$ は，次式で定義される．

$$a_{ij}^{(r)} = \begin{cases} 1 & \text{if } y_i^{(r)} = j \\ 0 & \text{otherwise} \end{cases} \tag{3.26}$$

さらに，$\boldsymbol{H}_\tau^{-1}$ ($\in \mathbb{R}^{d_{\boldsymbol{z}} \times d_{\boldsymbol{z}}}$) は τ ステップ目の更新により得られるヘッセ行列の逆行列であり，以下の通りに近似される．

$$\boldsymbol{H}_{\tau+1}^{-1} \leftarrow \boldsymbol{L}_0^{\mathrm{T}} \boldsymbol{H}_0^{-1} \boldsymbol{L}_0 + \sum_{t=0}^{\tau-1} [\rho_t \boldsymbol{L}_{t+1}^{\mathrm{T}} \boldsymbol{\Xi}_t \boldsymbol{\Xi}_t^{\mathrm{T}} \boldsymbol{L}_{t+1}] + \rho_\tau \boldsymbol{\Xi}_\tau \boldsymbol{\Xi}_\tau^{\mathrm{T}} \tag{3.27}$$

$$\boldsymbol{L}_t = \boldsymbol{\Lambda}_{\tau-\nu+t} \cdot \boldsymbol{\Lambda}_{\tau-\nu+t+1} \cdot \ \ldots \ \cdot \boldsymbol{\Lambda}_{\tau} \tag{3.28}$$

ただし，$\boldsymbol{H}_0^{-1}$ $(\in \mathbb{R}^{d_{\boldsymbol{z}} \times d_{\boldsymbol{z}}})$ は，ヘッセ行列の逆行列の初期値を示し，ν は，ヘッセ行列の逆行列 $\boldsymbol{H}_{\tau}^{-1}$ を近似するために重み行列と勾配行列を保持しておく過去のステップ数を示す．また，$\boldsymbol{\Xi}_{\tau}$ $(\in \mathbb{R}^{d_{\boldsymbol{z}} \times J})$，$\rho_{\tau}$ および $\boldsymbol{\Lambda}_{\tau}$ $(\in \mathbb{R}^{d_{\boldsymbol{z}} \times d_{\boldsymbol{z}}})$ は以下の通りに定義される．

$$\boldsymbol{\Xi}_{\tau} \leftarrow \boldsymbol{W}_{\tau} - \boldsymbol{W}_{\tau-1} \tag{3.29}$$

$$\rho_{\tau} \leftarrow \frac{1}{(\tilde{\boldsymbol{g}}_{\tau} - \tilde{\boldsymbol{g}}_{\tau-1})^{\mathrm{T}} \tilde{\boldsymbol{\xi}}_{\tau}} \tag{3.30}$$

$$\boldsymbol{\Lambda}_{\tau} \leftarrow \boldsymbol{I} - \rho_{\tau} (\boldsymbol{G}_{\tau} - \boldsymbol{G}_{\tau-1}) \boldsymbol{\Xi}^{\mathrm{T}} \tag{3.31}$$

ただし，$\tilde{\boldsymbol{g}}_{\tau} = [\boldsymbol{g}_1^{(\tau)\mathrm{T}}, \boldsymbol{g}_2^{(\tau)\mathrm{T}}, \cdots, \boldsymbol{g}_J^{(\tau)\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{(d_{\boldsymbol{z}} \times J)}$，$\tilde{\boldsymbol{\xi}}_{\tau} = [(\boldsymbol{w}_1^{(\tau)} - \boldsymbol{w}_1^{(\tau-1)})^{\mathrm{T}}, (\boldsymbol{w}_2^{(\tau)} - \boldsymbol{w}_2^{(\tau-1)})^{\mathrm{T}}, \cdots, (\boldsymbol{w}_J^{(\tau)} - \boldsymbol{w}_J^{(\tau-1)})^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{(d_{\boldsymbol{z}} \times J)}$ であり，$\boldsymbol{I}$ $(\in \mathbb{R}^{d_{\boldsymbol{z}} \times d_{\boldsymbol{z}}})$ は単位行列を表す．L-BFGS 法では，τ ステップ目の更新による探索方向行列 $\boldsymbol{H}_{\tau}^{-1} \boldsymbol{G}_{\tau}$ を近似しながら，最適な重み行列 $\boldsymbol{W}_{\mathrm{ML}}$ を算出する．

以上で算出した重み行列 $\boldsymbol{W}_{\mathrm{ML}}$ を用いて，以下の通りに，最終的な分類結果 $\hat{q}$ を取得する．

$$\hat{q} = \arg\max_{q \in \boldsymbol{Q}} P(q | \hat{\boldsymbol{z}}, \boldsymbol{W}_{\mathrm{ML}}) \tag{3.32}$$

ただし，$\boldsymbol{Q}$ は，**3.3.2** で得られた新たな分類候補のカテゴリの集合を表す．提案手法で用いられる分類モデルは，重み行列を算出する際に，各分類結果の正解率を考慮することで，分類精度の向上に貢献するデータを考慮して，最終的な分類結果を取得可能とする．

3.5　実験

本節では，提案手法の有効性を確認するために行った実験の結果を示す．なお本実験では，実際に画像共有サービスに投稿された観光画像をカテゴリへ分類することで，提案手法の性能を評価する．以降，**3.5.1** で本実験で用いるデータセットについて説明し，**3.5.2** で実験の条件を示す．また，**3.5.3** で実験の結果を示し，提案手法の有効性について考察を行う．

3.5.1　データセット

本実験では，実際に画像共有サービスに投稿された観光画像をカテゴリへ分類することで，提案手法の有効性の確認を行う．本実験では，2 種類のデータセット (データセット A およびデータセット B) を用いる．データセット A では，Flickr から取得したパリの観光名所に関する観光画像 49760 枚 (N=49760) を用いる．具体的には，2014 年 9 月 11 日の時点で，trip advisor[2]において紹介されている観光名所のカテゴリ Cultural，Museums，Performances，Landmarks，Amusement，Outdoors，Sports，および Zoo & Aquariums の 8 種類のカテゴリ (J=8) の観光名所ランキングにおいて，上位 5 件の観光名所の名称をタグにもつ画像を取得した．ただし，データセット収集に用いたタグを複数持った画像は収集対象から削除した．さらに，データセット B では，Flickr から取得したロンドンの観光名所に関する観光画像 32545 枚 (N=32545) を用いる．具体的には，2015 年 4 月 10 日の時点で，trip advisor において紹介されている観光名所のカテゴリ Sights，Museums，Theater，Nature，Shopping，および Zoo の 6 種類のカテゴリ (J=6) の観光名所ランキングにおいて，上位 5 件の観光名所の名称をタグにもつ画像を取得した．ただし，データセット収集に用いたタグを複数持った画像は収集対象から削除した．表 3.1 および 3.2 にデータセット A およびデータセット B における各クラスの画像数をそれぞれ示す．また，実験に用いる各データセット内のすべての投稿画像において，Flickr から画像を取得する際に用いた観光名所の名称に関するタグは

[2]http://www.tripadvisor.jp

表 3.1: データセット A における分類する観光カテゴリとデータセットに含まれる各カテゴリへ属する画像数.

観光カテゴリ	画像数 (枚)
Cultural	7,742
Museums	7,094
Performances	2,028
Landmarks	14,241
Amusement	630
Outdoors	10,835
Sports	4,060
Zoos & Aquariums	3,130

表 3.2: データセット B における分類する観光カテゴリとデータセットに含まれる各カテゴリへ属する画像数.

観光カテゴリ	画像数 (枚)
Sight	9,682
Museums	6,525
Theater	2,087
Nature	9,998
Shopping	3,287
Zoo	966

すべて削除し，実験を行った．以上のデータセットを用いて，本実験では提案手法の有効性の確認を行う．

3.5.2　実験条件

本実験では，画像特徴ベクトル算出の際，Bag of Features ベクトルの次元数を 2,000 とした．また，テキスト特徴ベクトル算出の際，データセットにおいて頻出タグ上位 1,000 種類のタグを用いた．なお，統合特徴ベクトルの算出では，上

記 2 つの特徴ベクトルを並べた 3,000 次元のベクトルに，mRMR アルゴリズムを施すことで選択された 200 次元の特徴ベクトルを用いる．また，ロンドンの観光画像のデータセットにおいて，画像特徴ベクトルから選択された特徴数は 59，タグ特徴ベクトルから選択された特徴数は，141 であった．提案手法における分類結果の確信度推定および統合処理に必要な各パラメータは，パリの観光画像のデータセットでは，確信度のしきい値 T を 0.59，分類候補のクラス数削減のためのしきい値 h_1 を 0.01，および h_2 を 10^{-6} と実験的に設定した．さらに，特徴ベクトル $\boldsymbol{z}$ を求める際には，近傍数 $K \in \{5, 10, 15, 20, 25, 50, 100\}$ の 7 つ $(E = 7)$ の FK-NN アルゴリズムにより算出された各カテゴリへの帰属確率を利用した．また，特徴ベクトル $\boldsymbol{z}$ を求める際，mRMR アルゴリズムにより選択する特徴数は 75 とした．なお，確信度算出には，先に述べた近傍数の中で，最も分類結果の正誤判定の正解率が高かった $K = 5$ の FK-NN アルゴリズムによるメンバーシップを用いた．また，分類候補削減には $K = 100$ の FK-NN アルゴリズムにより得られた帰属確率を用いた．ロンドンの観光画像のデータセットを用いた実験では，確信度の閾値 T を 0.81，分類候補のクラス数削減のための閾値 h_1 を 0.06，$h2$ を 10^{-6} に設定した．また，特徴ベクトル $\boldsymbol{z}$ を求める際に必要となるメンバーシップは，近傍数 $K \in \{5, 10, 15, 20, 25, 50, 100\}$ の 7 つ $(E = 7)$ の FK-NN アルゴリズムにより算出した．ただし，特徴ベクトル $\boldsymbol{z}$ を求める際，mRMR アルゴリズムにより選択する特徴数は 37 とした．確信度算出には $K = 5$，分類候補のクラス数の削減には，$K = 100$ の FK-NN アルゴリズムにより得られる帰属確率を用いた．なお，確信度算出には $K = 5$ の FK-NN アルゴリズムによる帰属確率，分類候補削減には $K = 100$ の FK-NN アルゴリズムによる帰属確率を用いた．また，SVM により構築する識別器のカーネル関数はすべてガウシアンカーネルとし，各パラメータは grid search [42] により決定した．

本実験では，表 3.3 に示す 7 種類の比較手法を用いて提案手法の有効性を検証する．具体的には，比較 1 により，異種データを用いた従来の分類手法と比較することで，提案手法の有効性を確認する．また，比較 2，比較 3，および比較 4 により，提案手法が複数の特徴を統合的に用いることの有効性を確認する．さらに，

比較 5 および比較 6 により，提案手法において，位置座標を利用した分類と画像およびテキストを利用した分類を別々に行う有効性を確認する．また，比較 8 により，分類候補のカテゴリ数を削減する有効性を確認する．

実験では，8 クラス分類を行うため，提案手法および比較手法で用いる SVM を one-versus-all [43] により多クラス分類に拡張した．さらに，提案手法および比較手法の評価のために，5 分割交差検定を行い，評価指標として以下に示す再現率，適合率，および F 値を用いた．具体的には，カテゴリ $j\,(=1,2,...,8)$ における評価指標をそれぞれ 再現率$_j$，適合率$_j$ および F 値$_j$ とするとき，以下の通りに定義される．

$$\text{再現率}_j = \frac{\text{正しくカテゴリ } j \text{ に分類された画像数}}{\text{カテゴリ } j \text{ に属する画像数}} \tag{3.33}$$

$$\text{適合率}_j = \frac{\text{正しくカテゴリ } j \text{ に分類された画像数}}{\text{カテゴリ } j \text{ に分類された画像数}} \tag{3.34}$$

$$\text{F 値}_j = \frac{2 \times \text{再現率}_j \times \text{適合率}_j}{\text{再現率}_j + \text{適合率}_j} \tag{3.35}$$

3.5.3　実験結果および考察

データセット A における実験結果を表 3.4 から 3.6 に示し，データセット B における実験結果を表 3.7 から 3.9 に示す．また，データセット A において，提案手法により分類された画像の例を図 3.3 に示す．実験結果より提案手法の評価指標がすべての比較手法を上回っていることが確認できる．

まず，比較 1 よりも提案手法の評価指標の平均が上回っていることから，従来手法と比較して提案手法の有効性が確認できる．比較 1 の従来手法では，訓練サンプルとテストサンプルの特徴ベクトル間の距離に基づいて，データの重要度を

表 3.3: 本実験で用いる比較手法.

比較手法	概要
比較 1	Adaptive Weighted Fusion Approach (AWFA) [14] を用いて分類を行う手法
比較 2	位置特徴ベクトルを用いて SVM により分類を行う手法
比較 3	画像特徴ベクトルを用いて SVM により分類を行う手法
比較 4	テキスト特徴ベクトルを用いて SVM により分類を行う手法
比較 5	位置特徴ベクトルを用いた FK-NN アルゴリズムにより分類を行う手法
比較 6	MALR モデルにより，画像特徴ベクトル，タグ特徴ベクトルおよびそれらを統合した特徴ベクトルによる分類結果を統合する手法
比較 7	提案手法の MALR モデルにおいて，分類候補カテゴリ数の削減を行わずに分類を行う手法

表すスコアを各データから算出し，このスコアを利用して異種データを統合している．しかしながら，比較 1 では，異種データに対する分類精度向上への貢献の度合いを考慮しながら統合することまでを想定していないため，本実験では高精度な分類が困難であったと考えられる．また，比較 2，比較 3 および比較 4 より，異種データを効果的に統合した画像分類の実現が可能であることが確認できる．ここで，位置特徴ベクトルを用いて SVM により分類を行う比較 2 に注目すると，その評価指標の値が非常に低い値になっている．図 3.4 に本実験で用いた観光画像から算出された位置特徴ベクトルを正しい観光カテゴリ毎にプロットした図を示す．パリ周辺の位置特徴空間上では，異なる観光カテゴリに属するサンプルが散在している．したがって，位置特徴ベクトルでは，SVM のマージン最大化によって算出された識別超平面を用いて，正確な分類を行うことが困難であると考えられる．さらに，比較 5 および比較 6 との比較から，提案手法では分類の誤りを確信度により検出し，再分類を行うことで高精度な画像分類を実現可能であることが確認できる．図 3.4 より，位置座標の空間では，異なるカテゴリに属するサンプルが散在しているものの，その近傍には，同一のカテゴリに属するサンプルが多数存在していることが確認できる．したがって，特徴空間上でテストサン

プルの近傍に存在する訓練サンプルに基づいて分類結果を取得する FK-NN アルゴリズムは，位置特徴ベクトルを用いた分類に適しているため，比較 5 の分類精度が比較 2 よりも非常に高くなったと考えられる．ここで，データセット A における提案手法と比較 5 の精度を比較すると，カテゴリ Cultural に対する提案手法の精度が，比較 5 を下回っていることが確認できる．この原因として，提案手法で用いた MALR モデルに導入したサンプル数の不均衡性を考慮したモデルが考えられる．具体的には，提案手法の MALR モデルに不均衡性を考慮したモデルを導入していない手法では，Cultural のみ提案手法の精度を上回っていることから，不均衡性を考慮したモデルの導入により Cultural の分類精度が低下したと考えられる．

図 3.5 にデータセット A において，比較 5 の手法で誤分類が発生したものの，提案手法によって正確な分類を実現した画像を示す．図 3.5 に示した画像の正しいカテゴリは，Performances であるものの，これらの画像が撮影された地点の近傍では，Landmarks のカテゴリに属する画像が多く撮影されているため，撮影地点の位置座標を用いた分類では，正確な分類が困難である．一方，提案手法では，確信度の推定を用いて誤分類を検出することで最終的な分類結果を取得するため，撮影地点の位置座標を用いた分類が困難な画像に対しても，正確な分類を可能とする．提案手法は比較 7 における評価指標の平均も上回っていることから，分類候補のカテゴリ数を削減する有効性が確認できる．

ここで，提案手法では訓練サンプル数の不均衡性を考慮した拡張を行っている．以下では，データセット A を用いて，不均衡性を考慮した拡張の有効性を検証する．表 3.10 に，提案手法における画像およびテキストを用いた分類を訓練サンプル数の不均衡性を考慮していない MALR モデル [37] により行った結果との比較を示す．表 3.10 より，訓練サンプル数の不均衡性を考慮したモデルに拡張した手法が，拡張前の手法よりも高精度な分類を実現していることが確認できる．特に，訓練データ数が少ない Amusement に注目すると，他のカテゴリと比較して，F 値が大きく上回っていることが確認できる．以上の実験結果より，提案手法では，分類結果の確信度推定に基づいた画像分類により，高精度な分類を実現



図 3.3: データセット A において提案手法により分類された画像例.

可能であることが確認できる.

3.6　まとめ

本章では，分類結果の確信度推定および統合処理の導入による画像共有サービスに投稿された観光画像分類の高精度化手法について提案した．提案手法は，位置座標に基づいて高精度な分類を行い，誤分類が発生している確率が高い場合，画像およびタグを用いた分類手法を用いて最終的な分類結果を取得する．具体的には，まず位置座標を用いた FK-NN アルゴリズムにより分類を行い分類結果を取得する．さらに，得られた分類結果の確信度を算出し，予め設定したしきい値

よりも高い場合，この分類結果を最終的な分類結果として出力する．一方で，確信度がしきい値よりも低い場合，算出した確信度に基づいて分類候補のクラス数を削減した後，画像およびタグを用いた分類結果統合手法によって最終的な分類結果を取得する．このとき，提案手法では，統合する分類結果の精度を考慮して分類結果の統合を行うことで高精度な分類を実現可能とする．また実験により，提案手法の有効性を定量的に確認した．

表 3.4: 分類結果の確信度推定を導入した画像分類に関する実験における提案手法および比較手法による分類結果の再現率 (データセット A).

	提案手法	比較 1	比較 2	比較 3	比較 4	比較 5	比較 6	比較 7
Cultural	**0.841**	0.207	0.051	0.310	0.718	0.822	0.672	0.782
Museums	**0.941**	0.277	0.000	0.322	0.692	0.930	0.700	0.924
Performances	**0.900**	0.009	0.000	0.188	0.550	0.924	0.574	0.889
Landmarks	**0.907**	0.359	0.070	0.530	0.548	0.892	0.661	0.901
Amusement	0.969	0.037	0.982	0.032	0.304	**0.972**	0.855	0.958
Outdoors	**0.792**	0.319	0.000	0.332	0.518	0.772	0.567	0.783
Sports	**0.986**	0.651	0.014	0.958	0.712	0.982	0.958	0.983
Zoos & Aquariums	**0.969**	0.142	0.062	0.164	0.628	0.964	0.739	0.963
平均値	**0.913**	0.261	0.147	0.324	0.615	0.907	0.716	0.898

表 3.5: 分類結果の確信度推定を導入した画像分類に関する実験における提案手法および比較手法による分類結果の適合率 (データセット A).

	提案手法	比較 1	比較 2	比較 3	比較 4	比較 5	比較 6	比較 7
Cultural	0.847	0.288	0.866	0.368	0.776	0.800	**0.857**	0.838
Museums	**0.941**	0.212	0.400	0.336	0.672	0.922	0.678	0.930
Performances	**0.958**	0.134	0.000	0.286	0.872	0.940	0.826	0.955
Landmarks	**0.854**	0.371	0.734	0.428	0.752	0.802	0.679	0.846
Amusement	0.948	0.056	0.010	0.088	0.394	**0.972**	0.411	0.771
Outdoors	0.855	0.292	0.000	0.356	0.780	0.822	0.783	**0.860**
Sports	**0.985**	0.659	0.970	0.588	0.986	0.978	0.815	0.960
Zoos & Aquariums	0.921	0.130	0.594	0.224	0.506	**0.964**	0.617	0.867
平均値	**0.914**	0.268	0.447	0.334	0.717	0.900	0.708	0.879

表 3.6: 分類結果の確信度推定を導入した画像分類に関する実験における提案手法および比較手法による分類結果の F 値 (データセット A).

	提案手法	比較 1	比較 2	比較 3	比較 4	比較 5	比較 6	比較 7
Cultural	**0.844**	0.241	0.092	0.336	0.736	0.801	0.753	0.809
Museums	**0.941**	0.240	0.000	0.328	0.626	0.925	0.665	0.927
Performances	0.928	0.111	0.000	0.230	0.672	**0.931**	0.674	0.921
Landmarks	**0.880**	0.365	0.128	0.476	0.612	0.874	0.654	0.872
Amusement	0.958	0.0438	0.024	0.046	0.204	**0.972**	0.508	0.842
Outdoors	**0.822**	0.305	0.000	0.344	0.622	0.796	0.656	0.820
Sports	**0.985**	0.655	0.026	0.642	0.972	**0.980**	0.875	0.971
Zoos & Aquariums	0.942	0.136	0.114	0.188	0.488	**0.964**	0.623	0.905
平均値	**0.913**	0.262	0.048	0.324	0.617	0.905	0.676	0.883

表 3.7: 分類結果の確信度推定を導入した画像分類に関する実験における提案手法および比較手法による分類結果の再現率 (データセット B).

	提案手法	比較 1	比較 2	比較 3	比較 4	比較 5	比較 6	比較 7
Sights	**0.933**	0.346	0.757	0.717	0.460	0.913	0.884	0.911
Museums	0.953	0.352	0.454	0.429	0.494	0.953	0.943	**0.954**
Theater	0.865	0.120	0.0704	0.0296	0.534	0.874	0.845	**0.877**
Nature	0.926	0.422	0.218	0.544	0.586	0.925	0.902	**0.928**
Shopping	**0.935**	0.152	0.0736	0.236	0.628	0.932	0.920	0.932
Zoo	0.900	0.106	0.000	0.0391	0.580	0.905	0.891	**0.906**
平均値	**0.919**	0.249	0.262	0.332	0.547	0.917	0.898	0.918

表 3.8: 分類結果の確信度推定を導入した画像分類に関する実験における提案手法および比較手法による分類結果の適合率 (データセット B).

	提案手法	比較 1	比較 2	比較 3	比較 4	比較 5	比較 6	比較 7
Sights	0.899	0.443	0.291	0.537	0.665	0.920	0.903	**0.927**
Museums	**0.955**	0.283	0.544	0.455	0.727	0.942	0.920	0.940
Theater	**0.900**	0.132	0.535	0.176	0.722	0.870	0.855	0.881
Nature	**0.942**	0.387	0.352	0.507	0.728	0.928	0.906	0.924
Shopping	**0.958**	0.194	0.662	0.347	0.624	0.930	0.898	0.924
Zoo	**0.933**	0.0672	0.000	0.208	0.717	0.900	0.865	0.906
平均	**0.931**	0.251	0.397	0.372	0.697	0.915	0.891	0.917

表 3.9: 分類結果の確信度推定を導入した画像分類に関する実験における提案手法および比較手法による分類結果の F 値 (データセット B).

	提案手法	比較 1	比較 2	比較 3	比較 4	比較 5	比較 6	比較 7
Sights	0.915	0.389	0.420	0.614	0.539	0.916	0.894	**0.919**
Museums	**0.954**	0.313	0.491	0.440	0.533	0.948	0.932	0.947
Theater	**0.882**	0.126	0.120	0.0506	0.599	0.872	0.850	0.879
Nature	**0.934**	0.404	0.155	0.525	0.579	0.926	0.904	0.926
Shopping	**0.947**	0.170	0.132	0.281	0.500	0.931	0.909	0.928
Zoo	**0.916**	0.0817	0.000	0.0655	0.621	0.902	0.878	0.905
平均	**0.925**	0.247	0.220	0.329	0.562	0.916	0.894	0.917

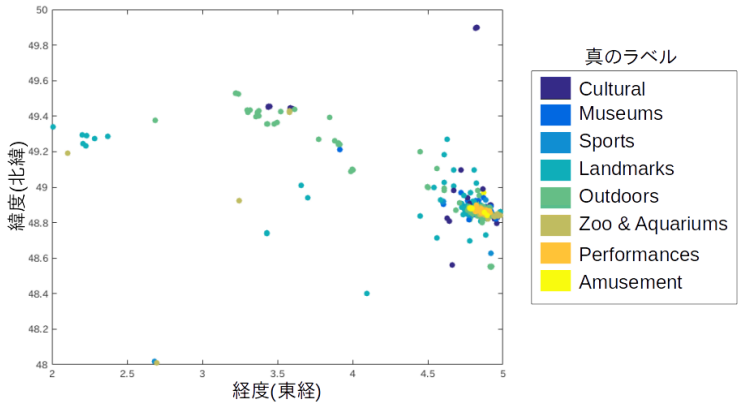


図 3.4: 位置特徴ベクトルを正しいカテゴリ毎にプロットした図.



(a) オペラ座の外観を撮影した画像



(b) オペラ座の内部を撮影した画像

図 3.5: 画像およびテキストを利用した分類によって正確な分類を可能とした観光画像例：(a) および (b) の正しいカテゴリは Performances である.

表 3.10: 画像およびテキストを利用した分類による結果の F 値.

	訓練サンプル数の不均衡性を考慮していないモデル	訓練サンプル数の不均衡性を考慮したモデル (=比較 6)
Cultural	**0.752**	0.751
Museums	**0.902**	0.898
Performances	0.888	**0.903**
Landmarks	0.840	**0.844**
Amusement	0.846	**0.956**
Outdoors	0.770	**0.772**
Sports	0.975	**0.976**
Zoos & Aquariums	0.941	**0.946**
平均	0.864	**0.881**

第4章　機械学習による確信度が高い分類結果の選択による画像分類

4.1　はじめに

本章では，機械学習による確信度が高い分類結果の選択による画像分類手法を提案する．提案手法の概要を図 4.1 に示す．提案手法ではまず，画像共有サービスへ投稿された画像，およびそのメタデータから得られた複数のデータからそれぞれ分類結果を取得する．これらの複数の分類結果に対して，機械学習を利用して確信度を推定する．提案手法では，上記の分類結果の中から最も確信度が高い結果を確からしい結果として選択し，これを最終分類結果とする．確信度を機械学習に基づいて推定することで，より高精度な確からしい結果の選択の実現が期待され，高精度な画像分類の実現が可能となる．以降では，**4.2** で異種データからの分類結果取得方法を説明し，**4.3** で確信度が高い分類結果の選択方法を説明する．**4.4** で実際に画像共有サービスに投稿された観光画像のカテゴリを分類する実験の結果を示し，提案手法の有効性について検証する．最後に **4.5** で本章の結論を述べる．

4.2　分類結果の取得

本節では，画像共有サービスへ投稿された画像，およびそのメタデータから分類結果を取得する方法について説明する．提案手法では，撮影地点の位置座標，投稿された画像，およびテキストタグの 3 種類のデータを利用して 2 つの分類結果を取得する．以降では，各分類結果の取得方法について説明する．

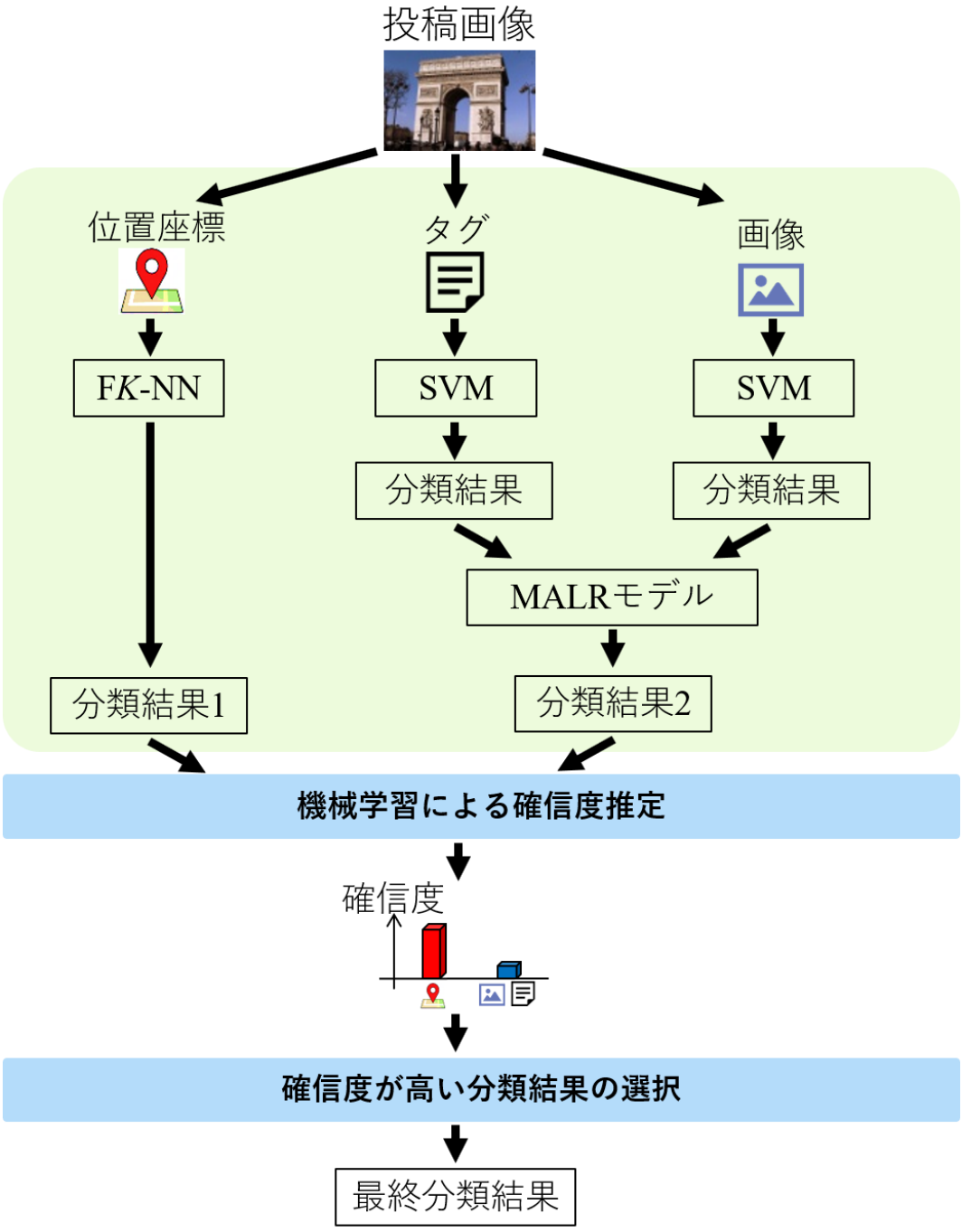


図 4.1: 機械学習により推定された確信度が高い分類結果の選択による画像分類手法の概要.

4.2.1　撮影地点の位置座標を用いた分類

提案手法では，撮影地点の位置座標を利用して一つ目の分類結果を取得する．具体的に，まず，投稿画像に付与されている GPS 情報から画像の撮影地点の緯度と経度を取得し，それら 2 つの値を要素に持つベクトルを位置特徴ベクトルとする．この特徴ベクトルを用いて FK-NN アルゴリズムにより分類を行うことで分類結果を取得する．なお提案手法では，分類結果とともに FK-NN アルゴリズムにより分類を行うことで得られる各分類候補カテゴリへの帰属確率を取得する．この帰属確率は，次節で説明する確信度が高い分類結果の選択において，確信度が高い結果を判定するための分類器への入力ベクトルを算出する際に利用する．以上の手法により，提案手法では撮影地点の位置座標を利用して分類結果を取得する．

4.2.2　画像およびテキストタグを利用した分類

提案手法では，画像およびテキストタグを利用して二つ目の分類結果を取得する．まず，画像およびテキストタグをそれぞれ用いて分類を行う．

画像を用いた分類では，まず深層学習を利用して特徴ベクトルを算出する．具体的には，深層学習のオープンソースソフトウェアである Caffe [44] を用いて，文献 [45] と同様に 4,096 次元の Deep convolutional activation feature (Decaf) を算出し，これを画像特徴ベクトルとする．提案手法では，上記で算出された特徴ベクトルを用いて SVM [38] により分類を行うことで，画像を利用した分類結果を取得する．

一方で，テキストタグを用いた分類では，まず各々のテキストタグが各要素に対応するベクトルを用意し，画像に付与されているタグについては 1，それ以外の値を 0 とすることで，タグ特徴ベクトルを算出する．このベクトルを用いて画像と同様に SVM により分類を行うことで，テキストタグを利用した分類結果を得る．

以上で算出された 2 種類の分類結果を統合することで，画像とテキストタグを

利用した分類結果を取得する．具体的には，分類結果を統合する際に，各々の分類精度を考慮可能な MALR モデル [37] を用いることで，分類結果を取得する．

4.3　機械学習による確信度が高い分類結果の選択

本節では，機械学習により推定された確信度が高い分類結果の選択方法について説明する．確信度が高い分類結果の選択方法の概要を図 4.2 に示す．提案手法では，前節で得られたすべての分類結果が一致している場合は，その結果を最終的な分類結果とする．一方，得られた分類結果が同一ではない場合，確信度が高い分類結果の選択を行い，これにより選択された結果を最終的な分類結果として取得する．以降では，確信度が高い分類結果の選択の処理について説明する．

提案手法では 2 種類の分類結果を取得する際に得られた各分類候補のカテゴリへの帰属確率を一列に並べたベクトル ($2J$ 次元) を用いて SVM により FK-NN アルゴリズムおよび分類結果統合による分類結果の中から最も正確な分類結果を推定し，これを最終分類結果とする．ここで，最も正確な分類結果を出力している分類手法を SVM を用いて推定するための識別器は以下のように構築する．まず，訓練データに対して交差検定を行うことで，各々の分類手法による分類結果とそれらの正誤を取得する．以上で取得した訓練データに対する分類結果とその正誤を用いて，正しい分類結果を出力した分類方法をラベルとして SVM の学習を行うことで，識別器を構築する．ただし，提案手法では上記の識別器を構築する際に，FK-NN アルゴリズムおよび分類結果統合による分類結果が同一である訓練データと，これらの分類結果が共に誤っている訓練データは除外して学習を行う．提案手法では，以上の手法により確信度が高い分類結果の選択を実現する．

4.4　実験

本節では，実際に画像共有サービスに投稿された観光画像を観光名所に関するカテゴリーへ分類する実験の結果を示し，提案手法の有効性について検証を行う．本実験では，データセットとして Flickr から取得したパリの観光名所に関する観光

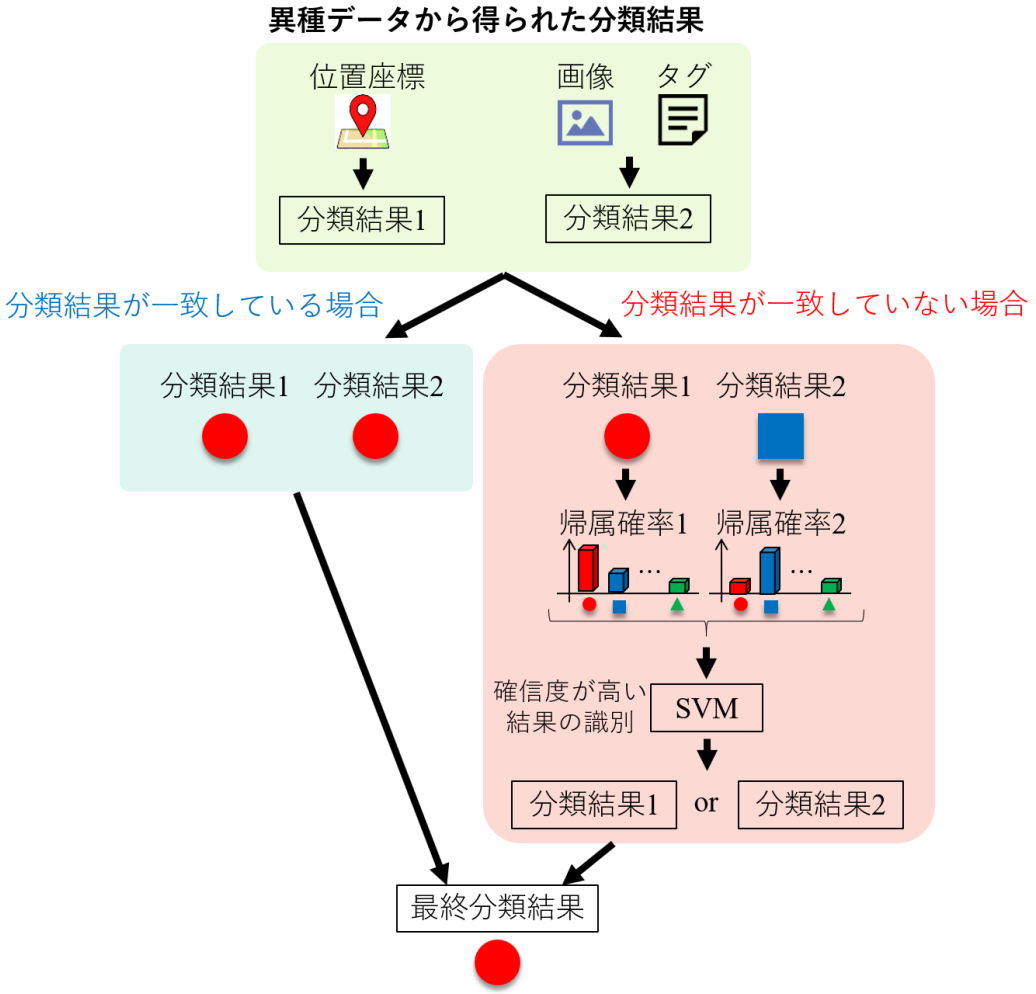


図 4.2: 確からしい分類結果の選択処理の概要.

画像 49760 枚 (N=49760) を用いる．実験に用いる観光画像の一例を図 4.3 に示す．具体的には，2014 年 9 月 11 日の時点で，trip advisor[1]において紹介されている観光名所のカテゴリ Cultural，Museums，Performances，Landmarks，Amusement，Outdoors，Sports，および Zoo & Aquariums の 8 種類のカテゴリ (J=8) の観光名所ランキングにおいて，上位 5 件の観光名所の名称をタグにもつ画像を取得した．ただし，データセット収集に用いたタグを複数持った画像は収集対象から削除した．また，実験に用いる各データセット内のすべての投稿画像において，Flickr

[1]http://www.tripadvisor.jp

から画像を取得する際に用いた観光名所の名称に関するタグはすべて削除し，実験を行った．また，SVM により構築する識別器のカーネル関数はすべてガウシアンカーネルとし，各パラメータは grid search [42] により決定した．

本実験では，表 4.1 に示す 3 種類の比較手法を用いて提案手法の有効性を検証する．具体的には，比較 1 および比較 2 により，提案手法により複数の特徴を協調的に用いることの有効性を確認する．さらに，比較 3 により，提案手法によって，前章で提案した手法の問題点を解決可能か確認する．

実験では，8 クラス分類を行うため，提案手法および比較手法で用いる SVM を one versus all 法 [43] により多クラス分類に拡張した．さらに，提案手法および比較手法の評価のために，10 分割交差検定を行い，評価指標として以下に示す再現率，適合率，および F 値を用いた．

実験結果を表 4.2 から 4.4 に示す．実験結果より提案手法の評価指標がすべての比較手法を上回っていることが確認できる．以降，提案手法と各比較手法との比較を行い，提案手法の有効性について考察を行っていく．

まず，比較 1 および比較 2 よりも提案手法の各評価指標が上回っていることから，提案手法によって異種データを協調利用することで高精度な画像分類が実現可能であることが確認できる．また，提案手法の各評価指標が比較 3 を上回っていることから，第 3 章で提案した手法における問題点を解決し，高精度な画像分類を実現したことが確認できる．以上より，提案手法によって異種データを協調利用することにより，正確な画像分類が実現可能であることが確認された．一方で，提案手法による誤分類の原因の多くは，異種データにより得られた選択対象となるすべての分類結果が誤っているとき，提案手法では必ず誤分類結果が最終的な分類結果として出力されてしまうという問題が存在しているためであると考えられる．上記の問題を解決する処理を導入し，より正確な画像分類を実現することが今後の課題としてあげられる．



図 4.3: 実験で用いる画像共有サービスに投稿された観光画像の一例.

4.5　まとめ

本章では，機械学習による確信度が高い分類結果の選択による画像分類手法を提案した．また，実際に画像共有サービスへ投稿された観光画像を対象として，そのカテゴリの分類に本手法を適用することによって，本手法の有効性を確認した．具体的に提案手法は，異種データを用いて得られた 2 種類の分類結果の中でどちらの方が確信度が高いかを機械学習に基づく方法で識別する．このとき，確

表 4.1: 実験で用いる比較手法.

	概要
比較手法 1	撮影地点の位置座標を用いて F*K*-NN アルゴリズムにより分類を行う手法
比較手法 2	画像およびテキストタグを用いて MALR モデルにより分類を行う手法
比較手法 3	第 3 章で提案した画像分類手法

表 4.2: 確信度が高い分類結果の選択による画像分類手法に関する実験における提案手法および比較手法による分類結果の再現率.

	提案手法	比較手法 1	比較手法 2	比較手法 3
Cultural	**0.818**	0.781	0.786	0.777
Museums	**0.938**	0.926	0.926	0.922
Performances	0.924	0.920	**0.934**	0.925
Landmarks	**0.893**	0.884	0.871	0.876
Amusement	0.959	0.970	0.953	**0.971**
Outdoors	**0.820**	0.785	0.800	0.769
Sports	0.983	0.981	**0.991**	0.981
Zoos & Aquariums	0.964	**0.969**	0.951	0.966
平均値	**0.912**	0.902	0.902	0.898

信度が高い分類結果として判定された結果を最終的な分類結果として取得した．選択候補の分類結果を取得する際，正しく分類された場合と誤分類が発生した場合とでは，分類候補のカテゴリへの帰属確率などの出力値の傾向が異なると考えられる．提案手法では，確信度が高い分類結果の選択に機械学習を利用することで，上記の分類結果の正誤による出力値の傾向を捉えることが可能となる．

ここで，前章および本章で提案した確信度が高い分類結果の選択では，選択対象の分類結果がすべて誤っている場合，最終分類結果も必ず誤ってしまう．この問題を解決するためには，上記のように選択処理では正しい結果を導出できない場合を検出し，別の方法により再分類を行う手法を提案する必要がある．そこで

表 4.3: 確信度が高い分類結果の選択による画像分類手法に関する実験における提案手法および比較手法による分類結果の適合率.

	提案手法	比較手法 1	比較手法 2	比較手法 3
Cultural	**0.829**	0.792	0.799	0.817
Museums	**0.940**	0.926	0.927	0.929
Performances	**0.967**	0.948	0.955	0.954
Landmarks	0.874	0.850	0.857	**0.875**
Amusement	0.985	**0.987**	0.978	0.984
Outdoors	0.821	0.813	0.798	**0.834**
Sports	0.987	0.987	**0.988**	0.986
Zoos & Aquariums	0.977	0.964	**0.978**	0.975
平均値	**0.923**	0.908	0.910	0.919

表 4.4: 確信度が高い分類結果の選択による画像分類手法に関する実験における提案手法および比較手法による分類結果の F 値.

	提案手法	比較手法 1	比較手法 2	比較手法 3
Cultural	**0.823**	0.786	0.792	0.796
Museums	**0.939**	0.926	0.927	0.926
Performances	**0.945**	0.933	0.942	0.939
Landmarks	**0.883**	0.867	0.864	0.875
Amusement	0.972	**0.978**	0.965	0.977
Outdoors	**0.820**	0.799	0.799	0.800
Sports	0.985	0.984	**0.989**	0.983
Zoos & Aquariums	**0.971**	0.966	0.964	0.970
平均値	**0.917**	0.905	0.905	0.908

次章では，上記のように，再分類の必要性を判定するために，選択候補の分類結果の中に確信度が高い結果が存在しているかを判定する方法を導入した画像分類手法を提案する.

第5章　帰属確率を入力とする分類器を導入した画像分類の高精度化

5.1　はじめに

本章では，分類精度の向上に向けて提案した帰属確率を入力とする分類器を導入した画像分類の高精度化について説明する．提案手法の概要を図 5.1 に示す．具体的にまず，画像共有サービスへ投稿された画像とそのメタデータである撮影地点の位置座標およびテキストタグの 3 種類のデータをそれぞれ利用して，3 種類の分類結果を取得する．提案手法では，これらの分類結果の中に確信度が高い結果が存在するかを判定する．確信度が高い分類結果が存在していると判定された場合，提案手法では，3 種類の分類結果の中で最も確信度が高い結果を最終的な分類結果とする．一方で，確信度が高い分類結果が存在しないと判定された場合，上記の異種データを単独で利用して得られた分類結果の中から確からしい結果を選択する手法では正確な分類が困難であると判断し，各分類器から出力される各カテゴリへの帰属確率を入力とする分類手法に基づいて，最終分類結果を取得する．以上の処理により，画像共有サービスに投稿された画像の高精度な分類を実現可能とする．

以降では，**5.2** で画像，位置座標およびテキストタグをそれぞれ単独で利用して分類結果を取得する方法について説明し，**5.3** で確信度が高い分類結果が存在しているか否かを判定する手法について説明する．さらに，最終分類結果を算出する方法について説明し，**5.4** で実際に画像共有サービスに投稿された観光画像のカテゴリーを分類する実験の結果を示すことで提案手法の有効性について検証

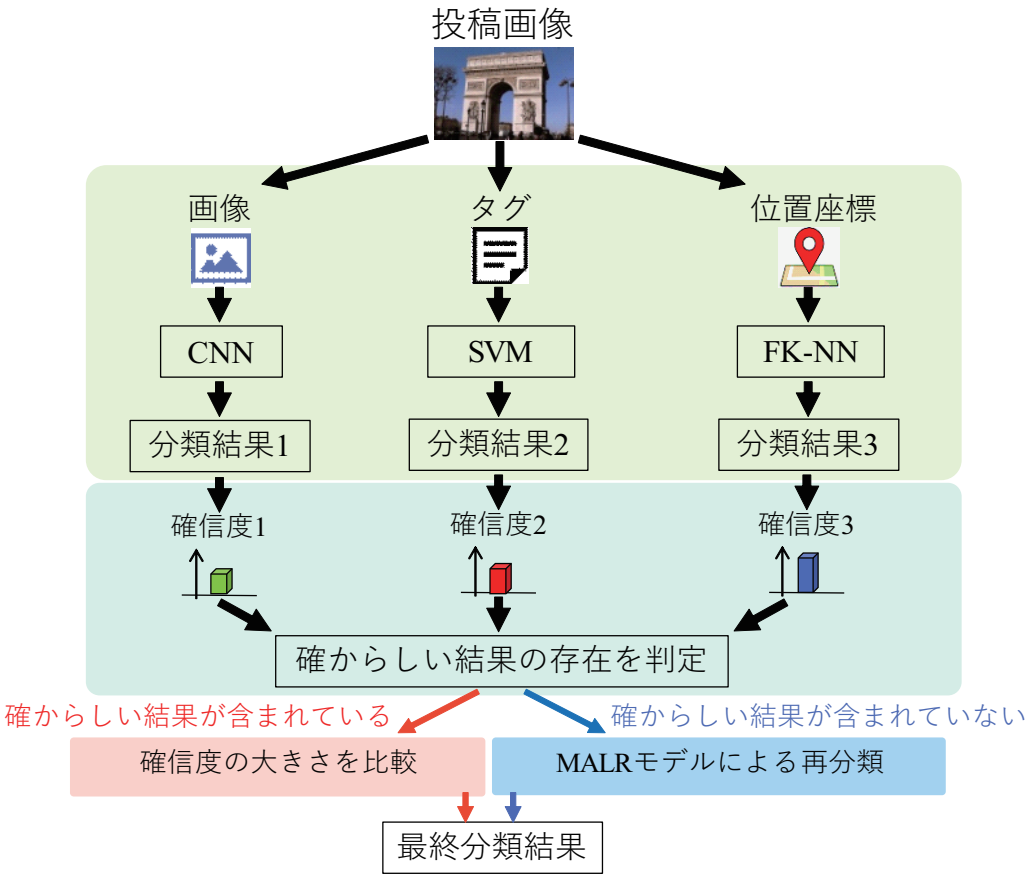


図 5.1: 確信度が高い分類結果の存在を考慮した画像分類手法の概要.

する．最後に **5.5** で本章をまとめる．

5.2　分類手法

本節では，画像共有サービスへ投稿された画像，撮影地点の位置座標，およびテキストタグをそれぞれ利用して分類結果を取得する方法について説明する．まず，投稿画像を利用した分類について説明する．提案手法では，画像を利用した分類に CNN [46] を利用する．具体的には，ILSVRC2012[1]のデータセットで事前学習された CNN モデルである VGG16 [47] に対して，対象とする画像の分類が

[1]http://www.image-net.org/challenges/LSVRC/2012/

可能となるように Fine tuning を行い，最終畳み込み層および全結合層の重みを求め直す．提案手法では，以上で得られた CNN モデルを用いて画像分類を行う．

画像の撮影地点の位置座標を用いた分類では，画像が撮影された地点の緯度および経度で構成される 2 次元のベクトルを利用した FK-NN アルゴリズムにより分類する．なお，次章以降で説明する確信度が高い分類結果の存在の判定，および最終分類結果の取得に各カテゴリへの帰属確率を利用する．そこで，提案手法では位置座標を利用した分類手法として，上記の帰属確率を算出可能な FK-NN アルゴリズムを採用した．

最後に，テキストタグを利用した分類手法について説明する．提案手法ではまず，テキストタグから Global Vectors for word representation (GloVe) [48] により特徴ベクトルを算出する．ここで，GloVe によりテキストタグから特徴ベクトルを算出する際には，すべてのテキストタグが同一の言語で付与されている必要がある．しかしながら，画像共有サービス上に存在する画像は世界各国から投稿されており，これらの画像には，様々な言語でテキストタグが付与されている．そこで提案手法では，上記の問題点を解決し，GloVe により特徴ベクトルを算出するため，英語以外の言語で付与されているテキストタグを Google Translation API[2] により英語へ翻訳したテキストタグを利用して，特徴ベクトルを算出する．提案手法では，上記で算出された特徴ベクトルを用いて Support Vector Machine (SVM) [38] により分類を行う．以上の手法により，提案手法では画像，位置座標，およびテキストタグをそれぞれ利用して，3 種類の分類結果を得る．

5.3　確信度が高い分類結果の存在の判定

本節では，前節で取得した 3 種類の分類結果の中に確信度が高い分類結果が含まれているかを判定する処理について説明する．提案手法ではまず，文献 [49] に基づいて，前節で得られた 3 種類の分類結果の確信度をそれぞれ算出する．これらの確信度を利用して前節で得られた 3 種類の分類結果の中に確信度が高い分類

[2]https://cloud.google.com/translate/

結果が存在しているかを判定する．以降では，**5.3.1** で確信度の算出方法について説明し，**5.3.2** で確信度が高い分類結果の存在を判定する手法について説明する．

5.3.1　確信度の算出

確信度の算出では，分類結果の正誤を識別するための識別器を構築し，この識別器によって出力される「分類結果が正しい」と識別される確率値をその分類結果の確信度として取得する．具体的にはまず，予めカテゴリ情報が既知の訓練サンプルの交差検定を行うことで，訓練サンプルに対する各データを利用した際の分類結果を取得する．さらに，上記の分類を行う際に得られる各カテゴリへの帰属確率を並べたベクトルを取得する．ただし,「各カテゴリへの帰属確率」は画像を利用した分類の場合は，CNN における最終層の出力値，位置座標を利用した分類の場合は，F*K*-NN アルゴリズムにより得られる各カテゴリへの帰属確率を表す．また，テキストタグを用いた分類の場合は，SVM により出力される各カテゴリへの帰属確率を表す．上記の交差検定により得られた分類結果と実際のカテゴリ情報を比較し，正しい訓練サンプルと誤っている訓練サンプルに分割し，これらのサンプルと各カテゴリへの帰属確率並べたベクトルを利用して，先に述べた分類結果の正誤を識別するための識別器を構築する．以上で，得られた識別器に対して，テストサンプルの分類により得られた帰属確率のベクトルを入力することで，分類結果の正誤の識別を行う．これにより得られた「分類結果が正しい」と識別される確率値をテストサンプルの確信度として取得する．

ここで提案手法で算出された確信度は，分類結果が正しいと識別される確率値として算出され，その値域が [0, 1] となるように求められたものである．提案手法における上記の SVM による分類結果の正誤の識別では,「正しく分類された結果」として識別される確率をテストサンプルの特徴と SVM の識別超平面からの符号付き距離をシグモイド関数に入力することで算出する．さらに，画像，位置座標，およびテキストを用いて得られた分類結果の確信度は，異なる識別器の出力として算出されるものの，その分布は同一であると考えることができる．したがって，提案手法で異種データから算出される確信度は，互いに比較可能となる

ため，加減算を行うことができる．以上で説明した手法により，提案手法ではテストサンプルに対して取得された 3 種類の分類結果の確信度をそれぞれ算出する．

5.3.2　確信度が高い分類結果の存在判定

提案手法では，one-versus-one 法に基づいた手法により確信度が高い分類結果の存在を判定する．今，l^{v}，l^{g}，および l^{t} ($\in \{1, 2, \ldots, M\}$; M は分類対象のカテゴリ数) をそれぞれ画像，位置座標，およびテキストタグを利用して分類結果として得られたカテゴリとする．また，C^{v}，C^{g}，および C^{t} をそれぞれ画像，位置座標，およびテキストタグにより得られた分類結果の確信度とするとき，提案手法では，2 つの特徴により得られた分類結果の確信度を比較することで，確信度が高い分類結果の存在を判定する．具体的に，特徴 f_1 および f_2 ($f_1, f_2 \in \{\mathrm{v}, \mathrm{g}, \mathrm{t}\}$) によりそれぞれ得られた結果 l^{f_1} および l^{f_2} が同じカテゴリであるとき，提案手法では仮の分類結果 $r_{f_1 f_2}$ を次式の通りに算出する．

$$r_{f_1 f_2} = \begin{cases} l^{\mathrm{common}} & \text{if } \frac{C^{f_1} + C^{f_2}}{2} \geq T_1^{f_1 f_2} \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

ただし，$l^{\mathrm{common}} = l^{f_1} = l^{f_2}$ であり，$T_1^{f_1 f_2}$ は特徴 f_1 および f_2 により得られた分類結果の確信度を比較するために事前に決定されたしきい値を表す．なお，仮の分類結果において，カテゴリ ‘0’ は，確信度が高い分類結果が選択候補の結果の中に含まれていないことを表し，カテゴリ ‘1’ は，確信度が高い分類結果が含まれていることを表す．一方で，l^{f_1} と l^{f_2} が異なるカテゴリの分類結果として取得されたとき，提案手法では，以下の通りに仮の分類結果 $r_{f_1 f_2}$ を取得する．

$$r_{f_1 f_2} = \begin{cases} l^{f_1} & \text{if } \{|C^{f_1} - C^{f_2}| \geq T_2^{f_1 f_2}\} \wedge \{C^{f_1} > C^{f_2}\} \\ l^{f_2} & \text{if } \{|C^{f_1} - C^{f_2}| \geq T_2^{f_1 f_2}\} \wedge \{C^{f_1} < C^{f_2}\} \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

ただし，$T_2^{f_1 f_2}$ は，分類結果が異なる際に特徴 f_1 および f_2 により得られた結果の確信度を比較するために事前に決定されたしきい値を表す．提案手法では，以上の確信度の比較を投稿画像，位置座標，タグのすべての組み合わせで行うことで，3 種類の仮の分類結果 r_{vg}，r_{vt}，および r_{gt} を取得する．これらの結果について多数決を行うことで仮の最終結果を得る．この多数決による仮の最終結果の取得では，取得対象の結果は，最終結果を出力するために採用する手法である．さらに，この多数決では，確信度を考慮している．したがって，提案手法で利用する多数決は，確信度を考慮し，分類結果ではなく分類手法を決定するために行うため，最終分類結果を得るために簡便な多数決を利用している従来の分類結果統合で生じていた，分類精度の低下は発生しないと考えられる．仮の最終結果がカテゴリ ‘1’ となった場合，提案手法では，選択対象の分類結果の中に確信度が高い分類結果が存在していると判断し，確信度が最も高い分類結果を最終分類結果として選択する．一方で，仮の最終結果がカテゴリ ‘0’ である場合，提案手法では，各カテゴリへの帰属確率を入力とする分類器により最終分類結果を算出する．以上の処理により，提案手法では選択対象の分類結果の中に確信度が高い結果が含まれているかを判断することで，最終分類結果の取得方法を自動で選択する．これにより，分類精度向上に有効な特徴と分類手法に基づいて，最終分類結果を取得することが可能になるため，正確な画像分類の実現が期待できる．

5.4　実験

本節では，提案手法の有効性を確認するために行った実験の結果を示す，本実験では，実際に画像共有サービスに投稿された観光画像を対象として，撮影されている観光名所に関するカテゴリを分類する．本実験では，100,449 枚の観光画像を 2017 年 12 月 3 日の時点で，trip advisor[3]において紹介されている観光名所のカテゴリのうち，Sights&Landmarks，Museums，Nature&Parks，Concerts&Shows，Shopping，Food&Drink，Zoo&Aquariums，および Event の 8 種類のカテゴリ[4]へ

[3]http://www.tripadvisor.jp
[4]本実験では，観光画像が 1,000 枚以上収集可能なカテゴリのみを分類対象とした．

表 5.1: 各カテゴリの観光画像の枚数.

観光カテゴリ	画像数
Sights&Landmarks	46,672 枚
Museums	11,891 枚
Nature&Parks	19,592 枚
Concerts&Shows	4,783 枚
Shopping	1,262 枚
Food&Drink	7,988 枚
Zoo&Aquariums	1,942 枚
Event	6,319 枚

分類する．ここで，本実験で用いる観光画像は，trip advisor において紹介されている各カテゴリ別の観光名所ランキングにおいて，上位 10 件の観光名所の名称をタグに持つ画像を Flickr から取得した．なお本実験では，複数のカテゴリを含む観光画像を分類対象としていないため，観光画像収集に用いたタグを複数有する画像については，収集対象から除外した．表 5.1 に各カテゴリの観光画像数を示し，図 5.2 にデータセットに含まれる観光画像の例を示す.

本実験では比較手法として以下の 5 種類の比較手法を用いる.

比較手法 1

観光画像を利用した CNN [46] により分類する手法.

比較手法 2

撮影地点の位置情報を利用した FK-NN アルゴリズム [36] により分類する手法.

比較手法 3

テキストから算出された特徴 [48] を用いて SVM により分類する手法.

比較手法 4

提案手法において，最終分類結果を取得するための分類手法の選択を行わ

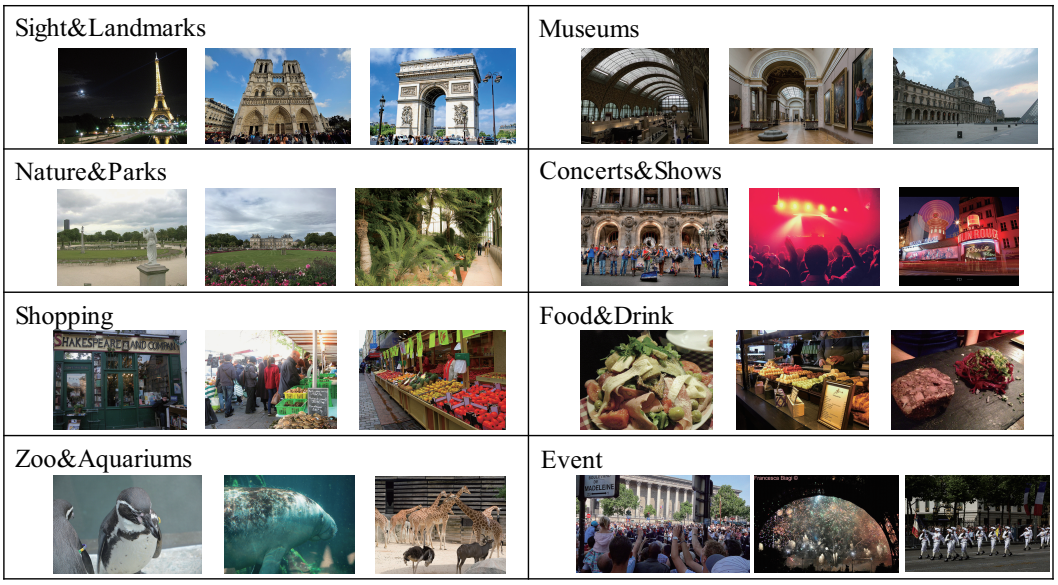


図 5.2: 実験で利用するデータセットに含まれる観光画像の例.

ずに，MALR モデル [37] に基づく各カテゴリへの帰属確率を入力とする分類器により最終分類結果を取得する手法.

比較手法 5

提案手法において，最終分類結果を取得するための分類手法の選択を行わずに，観光画像，撮影地点の位置座標，およびテキストをそれぞれ用いて複数の分類結果の中から確信が高い結果を選択する手法.

上記の比較手法との比較を行うことで，提案手法における複数の情報の同時利用，および分類結果の確信度算出と再分類の有効性を検証する．さらに，本実験では，提案手法および比較手法の分類精度を評価するために，データセットを訓練画像 80,356 枚，およびテスト画像 20,093 枚に分割し，評価指標として F 値を用いた．本実験では，CNN の fine tuning を行う際には，深層学習のオープンソースソフトウェアである TensorFlow[5] を利用した．提案手法において最終分類結果の取得手法を選択するために予め設定するしきい値 T_1^{vg}，T_2^{vg}，T_1^{vt}，T_2^{vt}，T_1^{gt}，および T_2^{gt}

[5] https://www.tensorflow.org/

をそれぞれ 0.4，0.2，0.6，0.9，0.5，および 0.5 に設定した．また，SVM のカーネル関数はガウシアンカーネルとし，そのパラメータはグリッドサーチ [50] により決定した．さらに，MALR モデルにおいて各カテゴリの訓練データ数の不均衡性を考慮するためのパラメータ ζ および正則化パラメータ λ は，それぞれ 0.5 および 1.0 とした．以上の実験方法により，提案手法の有効性を検証する．

表 5.2 に提案手法および比較手法により得られた分類結果の F 値を示す．実験結果より，提案手法の分類精度が他の比較手法を上回っていることが確認できる．以降では，提案手法と各比較手法の分類精度について比較を行い，提案手法の有効性について考察を行っていく．

まず，比較手法 4 の分類精度が比較手法 1，比較手法 2，および比較手法 3 を上回っていることから，提案手法における各カテゴリへの帰属確率を入力とする分類手法により，高精度な画像分類が実現可能であることが確認できる．また，比較手法 5 が比較手法 1，比較手法 2，および比較手法 3 よりも高精度な分類を実現していることから，提案手法における確信度が高い分類結果の選択による最終分類結果の取得が有効であることが確認できる．さらに，提案手法の分類精度が比較手法 1，比較手法 2，および比較手法 3 を上回っていることから，画像共有サービスに投稿される観光画像とそのメタデータである撮影地点の位置座標およびテキストを同時に利用して画像分類を行う有効性が確認できる．また，比較手法 4 および比較手法 5 と提案手法の分類精度を比較すると，比較手法 4 および比較手法 5 には，提案手法における新規点が含まれているため，その差はわずかであるものの，高精度な分類を実現するために適した手法を選択することで，分類精度が向上することが確認された．以上より，提案手法によって，画像共有サービスに投稿された観光画像を高精度に観光カテゴリへ分類する有効性が確認された．

ここで，観光客に対する観光情報の推薦システムへの提案手法の応用について議論する．提案手法を応用して，上記の推薦モデルを構築するためには，観光客が撮影した画像にそのカテゴリに関するタグなどの情報が付与されたデータを収集する必要がある．従来の SNS に採用されている方法を利用すると，画像が撮影

された地点の位置座標に基づいて，そのカテゴリを付与する方法が考えられる．しかしながら実験結果より，位置座標を用いて，上記のカテゴリを自動で付与した場合，誤ったカテゴリが付与される画像が存在するため，これらの画像に対しては手動でカテゴリを付与する必要がある．提案手法および比較手法 2 の位置座標のみを用いた FK-NN アルゴリズムによる分類結果の精度差より，位置座標のみを用いた場合と比較して，提案手法によって，手動による付与が必要が画像の約 10%のに対して，自動付与が可能となることを表している．画像共有サービスに投稿される画像数は膨大であるため，上記の自動付与により，観光情報推薦の実現への貢献が期待される．

さらに，提案手法では，分類結果の確信度に基づいて，分類結果を取得しており，最終分類結果の精度も確信度推定の精度に依存している．ここで，提案手法における確信度は，分類により得られた各カテゴリへの帰属確率を用いて分類結果の正誤を識別することで推定されるものであり，その推定精度は分類対象のカテゴリの細かさに依存しない．したがって，分類対象のカテゴリが本実験で用いたものよりも詳細になったとしても，確信度を正確に推定することが可能であるため，提案手法は，他の比較手法よりも高精度な分類を実現できると考えられる．

表 5.2: 各カテゴリへの帰属確率を入力とする分類器を導入した画像分類手法に関する実験における提案手法および比較手法により算出された分類結果の F 値.

	提案手法	比較手法 1	比較手法 2	比較手法 3	比較手法 4	比較手法 5
Sights&Landmarks	**0.927**	0.841	0.904	0.779	0.922	0.909
Museums	**0.920**	0.736	0.891	0.649	0.909	0.897
Nature&Parks	**0.841**	0.661	0.829	0.546	0.830	0.830
Concerts&Shows	**0.814**	0.556	0.651	0.616	0.788	0.745
Shopping	**0.801**	0.381	0.781	0.596	0.790	0.737
Food&Drink	**0.918**	0.854	0.811	0.567	0.914	0.832
Zoo&Aquariums	**0.986**	0.596	0.983	0.868	0.983	0.966
Event	0.896	0.762	0.788	0.864	**0.898**	0.848
平均	**0.888**	0.673	0.830	0.686	0.879	0.846

5.5　まとめ

本章では，分類精度の向上を目標として，各カテゴリへの帰属確率を入力とする分類器を導入した画像分類手法を提案した．また，実際に画像共有サービスに投稿された観光画像を対象とした分類に本手法を適用することで，その有効性を定量的に確認した．具体的に提案手法では，投稿画像，撮影地点の位置座標，およびテキストタグをそれぞれ用いて 3 種類の分類結果を取得する．これらの分類結果の中に確信度が高い結果が存在するかを判定する．確信度が高い分類結果が存在していると判定された場合，上記の 3 種類の分類結果の中で，最も確信度が高い結果を最終分類結果として取得する．一方で，確信度が高い分類結果が存在しないと判定された場合，選択処理による分類は適切ではないと判断し，各カテゴリへの帰属確率を入力とする分類器を用いた再分類を行い，この結果を最終的な分類結果とする．以上により，提案手法は，高精度な分類の実現のために適したデータと分類方法を特定し，分類に利用することを可能とした．

第6章　結論

6.1　本論文の成果

本論文では，画像共有サービスに投稿される画像の正確な分類を目的として，異種データを用いた新たな画像分類手法を提案した．以下に本論文における各章の概要について述べる．

第2章では，本研究の関連研究として，異種データを利用した画像分類手法を紹介し，本論において解決すべき問題を明らかにした．第3章では，異種データより得られた分類結果の確信度を利用した画像分類手法を提案した．まず，撮影地点の位置座標を用いた画像分類を行い，この結果の確信度を算出する．上記の確信度が予め決定した閾値よりも十分に高ければ，この結果を信頼性の高い結果として選択する．一方で，確信度が低い場合は，画像およびテキストタグを利用して得られた結果の信頼性が高いと判断し，これを最終分類結果とする．また，実際に画像共有サービスに投稿された観光画像を対象とした実験を行い，その有効性を定量的に確認した．これにより，分類器から出力された結果の確信を表す指標である確信度を利用して，信頼性の高い結果を選択することで，分類精度向上に貢献するデータを利用した分類を実現した．第4章では，第3章で利用した確信度では捉えることが困難な誤分類の傾向を考慮するために，機械学習により推定された確信度を用いた確からしい結果の選択処理による画像分類手法を提案した．まず，画像，撮影地点の位置座標，およびテキストタグから2種類の分類結果を取得する．これらの分類結果の中で，最も信頼性の高い分類結果を判定するため，各分類結果を取得する際に得られる分類候補クラスへの帰属確率を入力とする識別器を機械学習に基づいて構築する．上記の識別器により得られた誤分類の識別結果に基づいて確信度を推定し，この確信度が最も高い結果を最終分類

結果として出力する．第 5 章では，分類精度の向上に向けて帰属確率を入力とする分類器を導入し，最終分類結果を取得するために適した分類手法を自動で選択する処理による画像分類手法を提案した．この手法では，異種データを単独で利用して得られたすべての分類結果とそれらの確信度から，確からしい分類結果の選択処理，もしくは，各カテゴリへの帰属確率を入力とする分類手法により，最終分類結果を取得する．これにより，分類精度向上に貢献するデータおよび分類手法を利用した画像分類を実現した．また，実際に画像共有サービスに投稿された観光画像のカテゴリ分類を行った実験の結果より，本手法を適用することで，正確な分類が実現可能であることを示した．

6.2　今後の課題

本論文では，画像共有サービスに投稿される異種データを利用して，画像を分類する手法を提案した．これらの手法では，異種データを効率的に利用して画像分類を行うため，各データの重要度を考慮可能とした．しかしながら，本論文における実験では，上記の手法を最大でも 3 種類のデータにしか適用していない．画像共有サービスに投稿された画像からは，画像，テキスト，および位置座標に加えて，撮影時刻等のデータを取得することができる．したがって，実用性の検証のため，分類に利用する異種データの種類を増やして実験を行い，本論文で提案した手法の有効性を検証する必要がある．また，投稿画像のメタデータはすべての画像には，付与されておらず，位置座標等が欠損している画像が存在する．本論文の手法では，これらのデータの欠損がある画像は対象としていないものの，その数は膨大であるため，この問題を無視することは出来ない．したがって，今後の課題として，異種データの欠損を考慮した分類手法の構築が挙げられる．さらに，本分類手法を画像検索や推薦へ応用するための検証や手法の構築も不十分であった．これらを今後の課題として挙げ，異種データ解析および画像分類技術の更なる発展のために研究に取り組んでいく所存である．

謝辞

本研究は，著者が北海道大学に在学中に 2014 年 4 月から現在に至る約 5 年間にわたって行ったものである．この 5 年間の間に多くの方々の御支援をいただいた．

本研究を行うにあたり，研究の遂行のみならず，終始にわたり丁寧な御指導，御鞭撻を頂きました長谷山美紀教授に深く感謝致します．長谷山教授の御指導がなくては，本論文の完成はありませんでした．約 5 年間にわたり御指導頂いたことを生かし，今後も研究者として，自身の能力の向上に向けて精進して参ります．

本論文をまとめるにあたり，御助言いただき，副査をお引き受けいただいた北海道大学大学院情報科学研究科 情報メディア環境学研究室 山本強特任教授，北海道大学大学院情報科学研究科 言語メディア学研究室 荒木健治教授，北海道大学大学院情報科学研究科 メディア創生学研究室 坂本雄児教授，ならびに北海道大学大学院情報科学研究科 メディアダイナミクス研究室 小川貴弘准教授に深く感謝致します．

本研究を遂行するにあたり，公私にわたり多大なる御助力を頂いた北海道大学大学院工学研究院 先端モビリティ工学研究室 高橋翔准教授，北海道大学 数理・データサイエンス教育研究センター 湧田雄基特任准教授，ならびに北海道大学大学院情報科学研究科 メディアダイナミクス研究室 原川良介特任助教に深く感謝致します．また，本研究期間のみならず，著者が釧路工業高等専門学校在学時から御指導および御助言頂いた釧路工業高等専門学校 浅水仁教授に深く感謝致します．

最後に，著者の研究期間中，御協力頂いた北海道大学大学院情報科学研究科メディアネットワーク専攻メディアダイナミクス研究室の諸先輩，同期ならびに後輩学生の皆様に感謝の意を表し謝辞といたします．

参考文献

[1] J. Yuan, J. Luo, H. Kautz, and Y. Wu, “Mining gps traces and visual words for event classification,” in *Proceedings of the ACM international conference on multimedia information retrieval*, 2008, pp. 2–9.

[2] Y. Li, D. J. Crandall, and D. P. Huttenlocher, “Landmark classification in large-scale image collections,” in *Proceedings of IEEE International Conference on Computer Vision*, 2009, pp. 1957–1964, IEEE.

[3] L. Zhu, J. Shen, H. Jin, L. Xie, and R. Zheng, “Landmark classification with hierarchical multi-modal exemplar feature,” *IEEE Transactions on Multimedia*, vol. 17, no. 7, pp. 981–993, 2015.

[4] J. Luo, J. Yu, D. Joshi, and W. Hao, “Event recognition: viewing the world with a third eye,” in *Proceeding of the ACM International Conference on Multimedia*, 2008, pp. 1071–1080.

[5] D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg, “Mapping the world’s photos,” in *Proceedings of the 18th international conference on World wide web*, 2009, pp. 761–770.

[6] F. Zhu, Z. Ma, X. Li, G. Chen, J.-T. Chien, J.-H. Xue, and J. Guo, “Image-text dual neural network with decision strategy for small-sample image classification,” *Neurocomputing*, 2018.

[7] J. Fu and S. Lee, “A multi-class svm classification system based on learning methods from indistinguishable chinese official documents,” *Expert Systems with Applications*, vol. 39, no. 3, pp. 3127–3134, 2012.

[8] S. Papadopoulos, C. Zigkolis, Y. Kompatsiaris, and A. Vakali, “Cluster-based landmark and event detection for tagged photo collections,” *IEEE MultiMedia*, no. 1, pp. 52–63, 2010.

[9] N. Srivastava and R. R. Salakhutdinov, “Multimodal learning with deep boltzmann machines,” in *Advances in neural information processing systems*, 2012, pp. 2222–2230.

[10] R. Salakhutdinov and H. Larochelle, “Efficient learning of deep boltzmann machines,” in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 693–700.

[11] K. Tang, M. Paluri, L. Fei-Fei, R. Fergus, and L. Bourdev, “Improving image classification with location context,” *arXiv preprint arXiv:1505.03873*, 2015.

[12] S. Liao, X. Li, H. T. Shen, Y. Yang, and X. Du, “Tag features for geo-aware image classification,” *IEEE Trans. Multimedia*, vol. 17, no. 7, pp. 1058–1067, 2015.

[13] Y. Zhu and S. Newsam, “Land use classification using convolutional neural networks applied to ground-level images,” in *Proc. 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2015, p. 61.

[14] Y. Xu and Y. Lu, “Adaptive weighted fusion: A novel fusion approach for image classification,” *Neurocomputing*, vol. 168, pp. 566–574, 2015.

[15] S. Bahrampour, N. M. Nasrabadi, A. Ray, and W. K. Jenkins, “Multimodal task-driven dictionary learning for image classification,” *IEEE transactions on Image Processing*, vol. 25, no. 1, pp. 24–38, 2016.

[16] Y. Aytar, C. Vondrick, and A. Torralba, “See, hear, and read: Deep aligned representations,” *arXiv preprint arXiv:1706.00932*, 2017.

[17] X. Bai, M. Yang, P. Lyu, Y. Xu, and J. Luo, “Integrating scene text and visual appearance for fine-grained image classification,” *IEEE Access*, vol. 6, pp. 66322–66335, 2018.

[18] R. Bahmanyar, D. Espinoza-Molina, and M. Datcu, “Multisensor earth observation image classification based on a multimodal latent dirichlet allocation model,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 3, pp. 459–463, 2018.

[19] P. Das, C. Xu, R. F. Doell, and J. J. Corso, “A thousand frames in just a few words: Lingual description of videos through latent topics and sparse object stitching,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2634–2641.

[20] H. R. Shahdoosti and A. Mehrabi, “Multimodal image fusion using sparse representation classification in tetrolet domain,” *Digital Signal Processing*, 2018.

[21] Y. Liu, L. Liu, Y. Guo, and M. S. Lew, “Learning visual and textual representations for multimodal matching and classification,” *Pattern Recognition*, vol. 84, pp. 51–67, 2018.

[22] F. Du, J. Zhang, N. Ji, G. Shi, and C. Zhang, “An effective hierarchical extreme learning machine based multimodal fusion framework,” *Neurocomputing*, vol. 322, pp. 141–150, 2018.

[23] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: theory and applications,” *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.

[24] I. Gallo, A. Calefati, S. Nawaz, and M. K. Janjua, “Image and encoded text fusion for multi-modal classification,” *arXiv preprint arXiv:1810.02001*, 2018.

[25] T. Quack, B. Leibe, and L. Van Gool, “World-scale mining of objects and events from community photo collections,” in *Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval*, 2008ACM, pp. 47–56.

[26] T. Fagni, F. Falchi, and F. Sebastiani, “Image classification via adaptive ensembles of descriptor-specific classifiers,” *Pattern Recognition and Image Analysis*, vol. 20, no. 1, pp. 21–28, 2010.

[27] S. Corchs, E. Fersini, and F. Gasparini, “Ensemble learning on visual and textual data for social image emotion classification,” *International Journal of Machine Learning and Cybernetics*, pp. 1–14, 2017.

[28] X. Zhang, W. Chao, Z. Li, C. Liu, and R. Li, “Multi-modal kernel ridge regression for social image classification,” *Applied Soft Computing*, vol. 67, pp. 117–125, 2018.

[29] C. Saunders, A. Gammerman, and V. Vovk, “Ridge regression learning algorithm in dual variables,” in *the 15th International Conference on Machine Learning (ICML98)*, 1998, pp. 515–521.

[30] Z. A. Oraibi, H. Yousif, A. Hafiane, G. Seetharaman, and K. Palaniappan, “Learning local and deep features for efficient cell image classification using random forests,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 2446–2450.

[31] X. Liu, R. Zhang, Z. Meng, R. Hong, and G. Liu, “On fusing the latent deep cnn feature for image classification,” *World Wide Web*, pp. 1–14, 2018.

[32] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, “Multimodal fusion for multimedia analysis: a survey,” *Multimedia systems*, vol. 16, no. 6, pp. 345–379, 2010.

[33] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (surf),” *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[34] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints,” in *Proceedings of European Conference on Computer Vision In Workshop on Statistical Learning in Computer Vision*, 2004, vol.1, pp. 1–2.

[35] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[36] J. M. Keller, M. R. Gray, and J. A. Givens, “A fuzzy K-nearest neighbor algorithm,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 15, no. 4, pp. 580–585, 1985.

[37] F. Rodrigues, F. Pereira, and B. Ribeiro, “Learning from multiple annotators: Distinguishing good from random labelers,” *Pattern Recognition Letters*, vol. 34, no. 12, pp. 1428–1436, 2013.

[38] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[39] J. Platt, “Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods,” *Advances in large margin classifiers*, vol. 10, no. 3, pp. 61–74, 1999.

[40] Y. Dong, H. Guo, W. Zhi, and M. Fan, “Class imbalance oriented logistic regression,” in *Proceedings of IEEE International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, 2014, pp. 187–192.

[41] D. C. Liu and J. Nocedal, “On the limited memory bfgs method for large scale optimization,” *Mathematical Programming*, vol. 45, no. 1-3, pp. 503–528, 1989.

[42] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, “A practical guide to support vector classification,” Technical Report, Department of Computer Science, 2003.

[43] R. Rifkin and A. Klautau, “In defense of one-vs-all classification,” *The Journal of Machine Learning Research*, vol. 5, pp. 101–141, 2004.

[44] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in *Proceedings of international conference on multimedia*, 2014, pp. 675–678.

[45] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, “Decaf: A deep convolutional activation feature for generic visual recognition,” in *Proceedings of international conference on machine learning*, 2014, pp. 647–655.

[46] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[47] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[48] J. Pennington, R. Socher, and C. D. Manning, “Glove: Global vectors for word representation,” in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2014, pp. 1532–1543.

[49] 山崎俊彦, 大島辰之輔, 相澤清晴，“各クラスへの中間出力値を用いた多クラス認識のための確信度処理,” 映像情報メディア学会誌，vol. 69, no. 8, pp. J257–J260，2015.

[50] C.-W. Hsu, C.-C. Chang, C.-J. Lin, et al., “A practical guide to support vector classification,” *Technical report, Development of Computer Science*, pp. 1–16, 2003.

著者の研究業績

論 文

(A) 学会誌

[**A-1**] 斉藤 直輝, 小川 貴弘, 浅水 仁, 長谷山 美紀,“画像共有サービスに投稿される画像の観光名所に関するカテゴリ分類,” 電子情報通信学会論文誌 (D), vol. J99-D, no. 9, pp. 848-860, 2016.

[**A-2**] Y. Hirai, N. Okuda, N. Saito, T. Ogawa, R. Machida, S. Nomura, M. Ôhara, M. Haseyama, M. Shimomura, “The Friction Properties of Firebrat Scales,” Biomimetics, vol. 4, no. 1, p. 2, 2019.

(B) 国際学会

[**B-1**] N. Saito, T. Ogawa, Y. Hirai, and M. Haseyama, “Estimation of Salient Region Based on Support Vector Machine for Scanning Electron Microphotographs,” in *Proceedings of the Engineering Neo-Biomimetics VI and Satellite Workshop at Lake Biwa*, 2015, p. 50.

[**B-2**] N. Saito, T. Ogawa, S. Asamizu and M. Haseyama, “A Confidence-aware Method for User Activity Genre Classification Based on Heterogeneous Features from Image Sharing Services,” in *Proceedings of the International Workshop on Advanced Image Technology*, 2016, p. P.1C-8.

[**B-3**] N. Saito, T. Ogawa, S. Asamizu and M. Haseyama, “A Tourism Category Classification Method Based on Estimation of Reliable Decision,” in *Proceedings of the IEEE 5th Global Conference on Consumer Electronics*, 2016, pp. 52-53.

[**B-4**] N. Saito, T. Ogawa and M. Haseyama, "Identification of Reliable Results for Rubber Material Property Estimation Using Scanning Electron Microscope Images and Mix Proportions," in *Proceedings of the 32nd International Technical Conference on Circuits/Systems, Computers and Communications*, 2017, pp. 265-266.

[**B-5**] N. Saito, T. Ogawa, S. Asamizu and M. Haseyama, "Tourism Category Classification on Image Sharing Services Through Estimation of Existence of Reliable Results," in *Proceedings of the ACM International Conference on Multimedia Retrieval*, 2018, pp. 493-496.

[**B-6**] N. Saito, T. Ogawa, S. Asamizu and M. Haseyama, "Tourism Category Classification Based on Novel Decision-making Process from Multiple Results," in *Proceedings of The 2nd GI-CoRE GSQ, GSB & IGM Joint Symposium*, 2018.

[**B-7**] M. Matsumoto, N. Saito, T. Ogawa and M. Haseyama, "Missing Image Data Estimation via Iterative Convolutional Sparse Coding," in *Proceedings of the IEEE 7th Global Conference on Consumer Electronics*, 2018, pp. 169-170.

[**B-8**] T. Matsui, N. Saito, T. Ogawa, S. Asamizu and M. Haseyama, "Gaze-based Visual Feature Extraction via DLPCCA for Visual Sentiment Estimation," in *Proceedings of the 2019 Joint International Workshop on Advanced Image Technology and International Forum on Medical Imaging in Asia*, 2019, p. 122.

[**B-9**] N. Saito, T. Ogawa, S. Asamizu and M. Haseyama, "Classification of Tourism Categories Based on Heterogeneous Features Considering Existence of Reliable Results," in *Proceedings of the 2019 Joint International Workshop on Advanced Image Technology and International Forum on Medical Imaging in Asia*, 2019, p. 223.

[**B-10**] M. Matsumoto, N. Saito, T. Ogawa and M. Haseyama, "Chronic Gastritis Detection from Gastric X-ray Images via Deep Autoencoding Gaussian Mixture

Models," in *Proceedings of the IEEE 1st global Conference on Life Sciences and Technologies*, 2019 (to appear).

[**B-11**] T. Matsui, N. Saito, T. Ogawa, S. Asamizu and M. Haseyama, "Estimation of Emotions Evoked by Images Based on Multiple Gaze-based CNN Features," in *Proceedings of the IEEE 1st global Conference on Life Sciences and Technologies*, 2019 (to appear).

(C) 技術報告

[**C-1**] 斉藤 直輝, 小川 貴弘, 浅水 仁, 長谷山 美紀, "画像共有サイトの投稿画像の高精度な分類に関する検討," 映像情報メディア学会技術報告, vol. 39, no. 7, pp. 71-75, 2015.

[**C-2**] 斉藤 直輝, 小川 貴弘, 長谷山 美紀, "生物の走査型電子顕微鏡像の撮像部位の分類に関する検討–分類精度を考慮した複数の分類結果の統合手法の導入–," 第 18 回画像の認識・理解シンポジウム, 2015.

[**C-3**] 斉藤 直輝, 小川 貴弘, 平井 悠司, 長谷山 美紀, "生物の走査型電子顕微鏡像の撮像部位の分類に関する検討–カーネル法の導入による複数の分類結果統合の高精度化–" , 映像情報メディア学会技術報告, vol. 40, no. 6, pp. 51-54, 2016.

[**C-4**] 斉藤 直輝, 小川 貴弘, 浅水 仁, 長谷山 美紀, "複数の特徴を用いた観光画像の分類における正確な分類結果の推定方法の導入に関する検討" , 第 19 回画像の認識・理解シンポジウム, 2016.

[**C-5**] 斉藤 直輝, 小川 貴弘, 平井 悠司, 長谷山 美紀, "生物の走査型電子顕微鏡像の撮像部位の分類における特徴選択方法に関する検討" , 映像情報メディア学会技術報告, vol. 40, no. 28, pp. 19-20, 2016.

[**C-6**] 平井 悠司, 奥田 直人, 斉藤 直輝, 小川 貴弘, 町田 龍一郎, 野村 周平, 大原 昌宏, 長谷山 美紀, 下村 政嗣, "マダラシミ鱗粉の摩擦特性," 高分子討論会, 2016.

[**C-7**] 平井 悠司, 奥田 直人, 斉藤 直輝, 小川 貴弘, 町田 龍一郎, 野村 周平, 大原 昌宏, 長谷山 美紀, 下村 政嗣, “昆虫の体表面摩擦力測定,” コロイドおよび界面化学討論会, 2016.

[**C-8**] 斉藤 直輝, 小川 貴弘, 長谷山 美紀, “電子顕微鏡画像および配合量を用いたゴム材料の物性値推定に関する検討,” 映像情報メディア学会技術報告, vol. 41, no. 5, pp. 41–45, 2017.

[**C-9**] 斉藤 直輝, 小川 貴弘, 長谷山 美紀, “複数の条件で撮像された走査型電子顕微鏡画像を用いたゴム材料の物性値推定に関する検討,” 映像情報メディア学会技術報告, vol. 41, no. 29, pp. 31–34, 2017.

[**C-10**] 斉藤 直輝, 小川 貴弘, 浅水 仁, 長谷山 美紀, “深層学習を用いた画像共有サービスに投稿される観光画像のカテゴリ分類の高精度化に関する検討,” 映像情報メディア学会技術報告, vol. 42, no. 4, pp. 161–165, 2018.

[**C-11**] 斉藤 直輝, 小川 貴弘, 長谷山 美紀, “電子顕微鏡画像および配合量を用いたゴム材料の特性曲線の推定に関する検討,” 映像情報メディア学会技術報告, 2019 (2019 年 2 月発表予定).

[**C-12**] 松井 太我, 斉藤 直輝, 小川 貴弘, 浅水 仁, 長谷山 美紀, “視線データを用いた画像注視時に誘起される感情の推定に関する検討–KDLPCCA に基づく特徴変換の導入による高精度化–,” 映像情報メディア学会技術報告, 2019 (2019 年 2 月発表予定).

(D) 講演発表

[**D-1**] 斉藤 直輝, 小川 貴弘, 浅水 仁, 長谷山 美紀, “画像共有サイトへの投稿画像およびそのメタデータに注目したマルチモーダル画像検索手法に関する検討,” 平成 26 年度 電気・情報関係学会北海道支部連合大会講演論文集, p. 141, 2014.

[**D-2**] 斉藤 直輝, 小川 貴弘, 浅水 仁, 長谷山 美紀, “画像共有サービスにおける投稿画像の分類結果の信頼度推定に関する検討,” 平成 27 年度 電気・情報関係学会北海道支部連合大会 講演論文集, p. 128, 2015.

[**D-3**] 松本 真直, 斉藤 直輝, 小川 貴弘, 長谷山 美紀, “反復型の Multi-channel Convolutional Sparse Coding を用いた欠損した画素値の推定に関する検討,” 平成 30 年度 電気・情報関係学会北海道支部連合大会講演論文集, pp. 143–144, 2018.

[**D-4**] 松井 太我, 斉藤 直輝, 小川 貴弘, 浅水 仁, 長谷山 美紀, “視線データを用いた画像注視時の感情推定に関する検討–正準相関分析に基づく画像特徴変換手法の導入–,” 平成 30 年度 電気・情報関係学会北海道支部連合大会講演論文集, pp. 139–140, 2018.

(E) 受賞

[**E-1**] The 2015 IEEE Sapporo Section Student Paper Contest Encouraging Prize

[**E-2**] 1st Prize IEEE GCCE 2016 Excellent Poster Award

[**E-3**] 2017 IEEE Sapporo Section Paper Awards Encouragement Award

[**E-4**] The 2019 Joint International Workshop on Advanced Image Technology & International Forum on Medical Imaging in Asia IWAIT Best Paper Award