

さまざまな分野で活用される マルチモーダル・マイニング

中野 宏毅^{†1} 阪本 正治^{†1} 森 由美^{†2}

^{†1} 日本アイ・ビー・エム (株) ^{†2} 横浜国立大学

Blog, Twitter, 製品カタログなど、インターネット上に存在する多くのドキュメントには画像とテキストデータとが含まれる。画像情報とテキスト情報とを統合したマルチモーダル・マイニングを実現することにより、テキストだけ、あるいは画像だけの解析では得られなかった新たな知見を得る可能性がある。また、対話型ロボットでは視覚情報と音声情報とを統合的に解析することにより、より効果的な対話を実現できる可能性がある。一方、医用画像の解析においても、画像以外の診断情報と統合することにより、より精度の高い診断支援システムを構築できる可能性がある。本稿では、マルチモーダル・マイニングに関連する筆者らの取り組みを紹介し、今後の展望について述べる。

1. はじめに

近年インターネットやイントラネットに存在する文書の多くは、テキストと画像や図表とが混在したものとなっている。テキストを解析して新たな知見を引き出す手法であるテキスト・マイニングでは、複数の製品が市場に出回っている。一方、大量の画像を解析して新たな知見を得る画像マイニングの研究も盛んであるが、両者を統合したマルチモーダル・マイニングは、テキストだけ、あるいは画像だけの解析では得られなかった新たな知見を得る可能性がある。一例としてSyeda-Mahmoodらによる医用画像と電子カルテなどのテキスト情報とを統合して新たな医学上の知見を得るための研究が挙げられる[1]。本稿では、画像データとそれ以外のデータを組み合わせたマルチモーダル・マイニングの実例について述べる。第2章では、マルチモーダル・マイニングを構成する要素技術としての画像マイニングとその実践例について述べる。具体的には、タブレットを用いて書籍や菓子箱などの商品を撮影し、複数の商品を同時に識別して必要な情報をタブレット画面にオーバーレイ表示する大規模商品画像識別システムを紹介する。第3章では、画像を含むTwitterメッセージに対して、第2章で紹介した商品識別システムとテキストマイニング・ソフトウェアとを組み合わせたマルチモーダル・マイニングによる、商品のマーケティング・リサーチの事例を紹介する。第4章では、医用画像のマルチモーダル・マイニングの事例を紹介する。この第3章と第4章とが、本稿の骨子となる。そして第5章では、マルチモーダル・マイニングの今後の展望を示す。

2. 大規模商品識別システム

本章では、第3章と第4章とで述べるマルチモーダル・マイニングを構成するための重要な要素技術である画像マイニングの最新事例について述べる。

近年の物体認識アルゴリズムの進歩と計算機の飛躍的な能力向上により、画像認識技術による多品種の商品の識別が可能となってきている。書店では、書籍にスマートフォンやタブレット端末をかざすことにより、本を手にとってISBNバーコードをスキャンすることなく、売り上げランキングや在庫情報を書籍画像にオーバーラップ表示させて、店頭在庫管理の効率化をはかりたいという要求がある。しかしながら、商品の品種に比例して、マスターデータ作成や照合に要する時間が増大するという課題がある。筆者らは、10万点規模の書籍マスターから1秒以内に99%以上の正答率で書籍の照合を行いたいというユーザの要求に答えるために、GIFTS (Goods Image Feature for Tree Search) と呼ぶ画像局所特徴を使用した大規模特定物体照合システムを提案した[2]。本章では、このシステムにおける実用化に向けた取り組みと工夫について述べる。

2.1 実利用のための課題

商品の認識など実利用に特定物体認識技術を応用するためには、高い認識精度や高速処理に加えて、実利用の際の利便性やインタフェースも重要である。たくさんの商品をすべて認識したい場合には、1枚の写真の中に納まる複数の商品を一度に認識できれば効率化が図れる。また、商品がある程度斜めから撮影された写真であっても認識できるようなロバスト性も実用に

は重要な要素である。

2.2 GIFTS 画像局所特徴

GIFTS 画像局所特徴はキーポイント法の一つである[3]。まず、原画像および、原画像を $1/\sqrt{2}$ 倍ずつ繰り返し7回縮小した、8オクターブの多重解像度画像を得る。それぞれの画像の各画素について、半径Rの円周上にある16画素を、濃淡差により明画素、暗画素および類似画素に分類し、明画素あるいは暗画素が9個以上連続している場合に、その画素をキーポイントとする。

図1に、注目画素pとその周囲の円周上の16個の画素(番号1~16)のpとの濃淡差について示す。

ここでは右回りに番号12~16, 1~7の12個の画素が暗画素として連続しており、注目画素pがキーポイントであることが分かる。また、キーポイントのオリエンテーションは、中心座標を起点として、半径Rの円周内部の画素値の重心座標を終点としたベクトルの向きとする。GIFTSでは、キーポイントの特徴ベクトルとして、キーポイントの周囲 31×31 画素の中で128個の画素ペアを選択し、その画素ペアの画素値の差を特徴ベクトルとしている。

2.3 KD-Tree

KD-Treeは二分木のk次元への拡張である[4]。K次元のベクトル群から木構造を生成し、クエリー対象ベクトルと最も距離の近いノードに配置されたベクトルを探索する。KD-Treeでは、 $O(M \cdot \log N)$ オーダーの探索時間でマスタ画像とクエリー画像との照合が可能と考えられる。ここでNはKD-Tree生成に使用したマスタ画像群のキーポイント総数、Mはクエリー画像のキーポイント数である。

マスタ画像が1万枚の場合、マスタ画像1枚あたり平

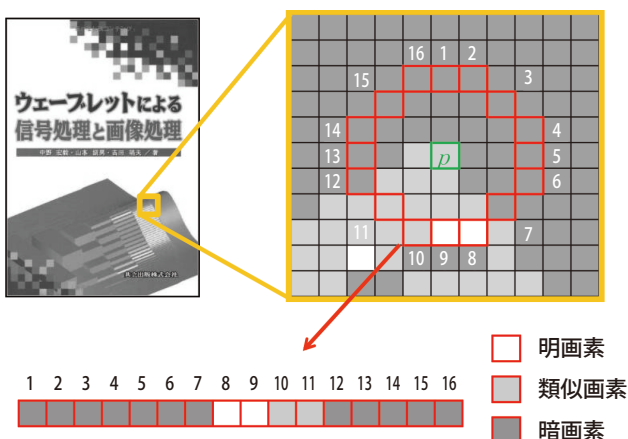


図1 GIFTSのキーポイント検出方法

均1,000個のキーポイントがあるとする、1,000万個の特徴ベクトルによりKD-Treeが生成される。照合時にはKD-Treeを探索し、クエリー画像の特徴ベクトルそれぞれについて、最も距離の近い特徴ベクトルの含まれるマスタ画像に投票する。投票数が閾値以上で最も投票数の多いマスタ画像を検索結果とする。

2.4 大規模書籍照合システム

本章で対象とする書籍は、日本で流通しているものだけで数十万種類といわれている。マスタ書籍は出版社ごと、あるいは出版時期ごとに分けて複数のKD-Treeを生成する。マスタ書籍数の増加に伴い、計算機で使用するコアを増やし、複数のコアにKD-Treeを配置して並列にクエリーを実行する。このようなシステム構成をとることにより、マスタ画像の枚数が増加しても、探索時間を一定時間以内に収めることが可能となる。

この実験では約1万枚のマスタ画像ごとにKD-Treeを生成し、プロセッサコアに配置する。照合時には各コアのKD-Treeを探索し、最も一致するキーポイントの多いマスタ画像を検索結果とする。

なお、表紙の類似した書籍が検索結果となる場合があるため、上位n個の検索結果を求める。図2にシステム構成を示す。書籍約1万冊ごとにグループ分けした表紙の画像をマスタ画像群G1, G2, ..., Gxとする。検索サーバにおいて各々の画像のGIFTS特徴量を抽出し(a), マスタ画像群ごとにKD-Treeを生成してプロセッサコアに配置しておく(b)。ユーザは検索対象の書籍の

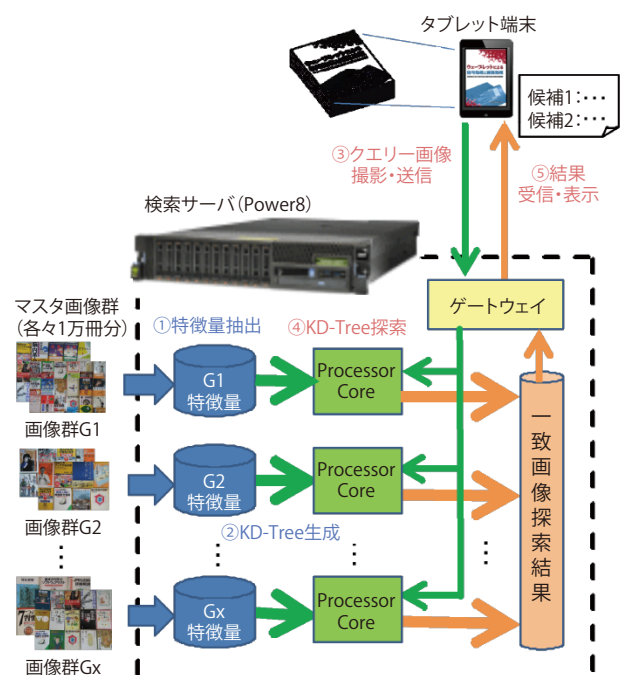


図2 システム構成法

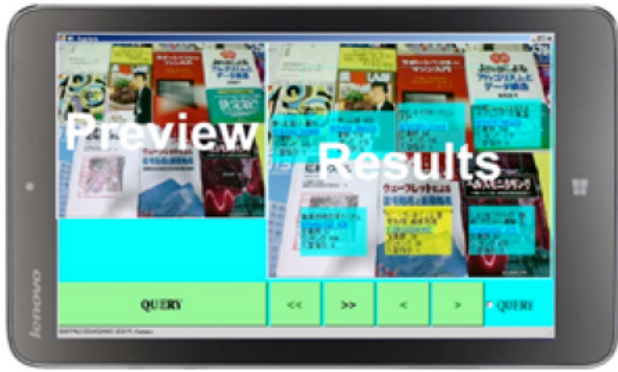


図3 タブレット端末による照合結果の表示例

表紙をタブレット端末で撮影し、それをクエリー画像として検索サーバに送信する(c)。検索サーバにおいてクエリー画像からGIFTS特徴量を抽出し、各プロセスコアでマスタ画像群のGIFTS特徴量とのKD-Tree探索による照合を行って一致する画像の候補を抽出する(d)。その結果をタブレットに送信し、タブレットは受信した探索結果を表示する(e)。本稿では、検索サーバが画像を受信してから、サーバ内部で照合結果を得るまでの時間を探索時間と呼ぶ。タブレット端末に表示された照合結果の例を図3に示す。書籍画像に検索結果がオーバーレイ表示される。表示されるデータは、書籍名・ISBN・在庫数・入れ替え指示などであり、書店での店頭在庫管理に必要な表示内容となっている。複数の書籍を一括して照合できることは、バーコードにはない利点と考えられる。

2.5 実証実験と考察

GIFTS特徴を使用した画像照合手法の認識性能と処理速度とを測定した。インターネットから書籍の表紙画像を10万点収集し、これをマスタ画像とした。マスタ画像のサイズは120×180画素から300×450画素である。実験にはIBM Power8 サーバ (3.72GHz, 20core, 512GB memory, OS: RHEL) を使用した。プロセッサの1コアごとに1万枚前後のマスタ画像を割り当て、照合精度および探索時間の実測を行った。ここではコアあたりの実験結果を示す。

マスタ画像を1万枚登録し、クエリー画像の鉛直方向からの傾きごとの、マスタに存在する4,000枚のクエリー画像に対する正答率および、マスタに存在しない4,000枚の画像のリジェクト率を測定した(図4)。正答率%は、

$$(\text{正解クエリー数} / \text{クエリー画像枚数}) \times 100$$

と定義し、リジェクト率%は、

$$(\text{正解リジェクト数} / \text{クエリー画像枚数}) \times 100$$

と定義する。

マスタ画像中の書籍とクエリー画像中の書籍との傾きの差が22.5度以下の場合には99%以上の正答率が得られ、またどの傾き角においても99%以上のリジェクト率を得た。また、マスタ画像が10万枚の場合においても同等の正答率を得ることができた。なお、実証実験では、マッチするキーポイント数がしきい値50以上の上位3個までの検索結果を表示する。同じコンテンツの書籍が単行本と文庫版で出版されている場合など、類似の表紙デザインの書籍が検索結果として複数表示されることがある。ただし、このような場合でもクエリー画像と同じ書籍のマスタ画像のスコアが最大となるため、2番目以降の順位のリジェクト率が正解となるケースは見られなかった。

図5に登録したマスタ画像枚数ごとの平均探索時間を示す。画像あたりの最大キーポイント数は、マスタ画像・クエリー画像ともに1,000とした。KD-Treeに約1万枚のマスタ画像の特徴ベクトルを配置した場合は探索時間が0.15秒となり、 $O(M \cdot \log N)$ オーダの探索時間であることが確認できた。さらに、図3のシステム構成において、10コアのKD-Treeに10万枚のマスタ画像の特徴ベクトルを配置した場合においても0.2秒以内の探索時間であることを確認した。

2.6 実運用時における正答率の改善

実証実験において、チルト角が22.5度以上の場合

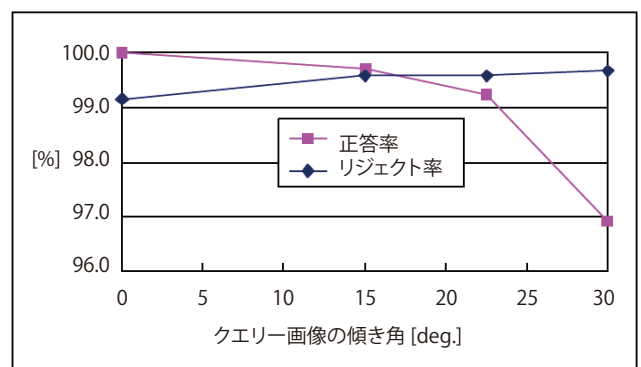


図4 クエリー画像の傾き角ごとの正答率とリジェクト率

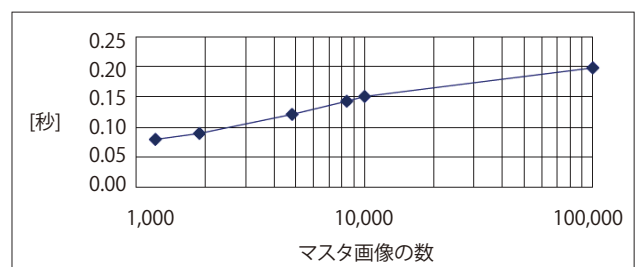


図5 マスタ画像枚数ごとの平均探索時間



図6 チルト角補正
(a) クエリー画像, (b) チルト角補正後のクエリー画像

正答率が99%未満になることが分かった。そこで、タブレット端末の傾きセンサを利用して、チルト角補正を試みた。その結果、チルト角がほぼ0のクエリー画像を得ることができ、正答率・リジェクト率とも99%以上の精度を得ることができた (図6)。

2.7 大規模商品識別のまとめ

GIFTSと命名した木構造探索に向けた画像局所特徴およびKD-Treeの組合せによる大規模特定物体照合システムを提案し、実利用可能な認識精度および照合時間が達成できるかを検証した。10万点規模の書籍の検索に提案システムを適用し、99%以上の高い正答率を得ることができ、そのときの検索サーバ内部での探索所要時間は0.2秒以内であった。実験システムはシングルユーザのため、1秒以内の応答速度でタブレット端末から照合結果を得ることができるが、システムを拡張することにより、マルチユーザの場合においても1秒以内の応答時間を得ることは可能と考えられる。

本章で述べた高精度な画像識別技術とテキスト・マイニングとを組み合わせれば、Twitterのような画像を含む文書の分析精度を高められる可能性がある。次章では、画像マイニングとテキスト・マイニングとを統合した事例について紹介する。

3. マーケティング・リサーチ

マルチモーダル・マイニングの事例として (さらに大規模商品識別システムの応用として)、SNS (Social Networking Service) における評判分析について紹介する。従来のSNS評判分析では、たとえばTwitterであればリアルタイムにつぶやきを効率良く収集してテキストマイニングを行い、世界のどこで何が話題になっているかを即座に知ることができる。メーカーにとっては、自社製品の売れ行きとともに評判も気になるところであり、Twitterの評判分析を行うことにより現在および

過去にわたって製品に関連するコメントを検索・収集して分析することが容易にできるようになった。ところで、Twitterではテキスト情報だけではなく、写真などの画像も多用されている。この場合には、画像そのものが主語や目的語となり、コメント上にはそれらが示されないことが多い。この例を図7に示す。

この図に示すように、テキスト情報の「飲みたくて探しているけど見つからない」だけでは何が飲みたいのか、何がみつからないのか、まったく分からず、情報としての価値はないに等しい。画像を見て初めて内容を理解することができる。このように画像が添付されているTwitterにはコメントだけを見ても意味をなさないものが少なくない。したがって、Twitter分析を行う場合には、テキスト情報だけではなく、添付されている画像に何が示されているかを理解し画像から得られる情報も合わせて分析することが必要となる。Twitter分析においてもテキストと画像とを用いたマルチモーダル・マイニングが欠かせない技術であると考えられる。

図8にテキスト情報と画像を含む膨大なTwitterを対象にした価値ある知見・情報抽出の概要を示す。画像をテキスト化・数値化した後、膨大な量の情報を高速



図7 画像が添付されている Twitter の例

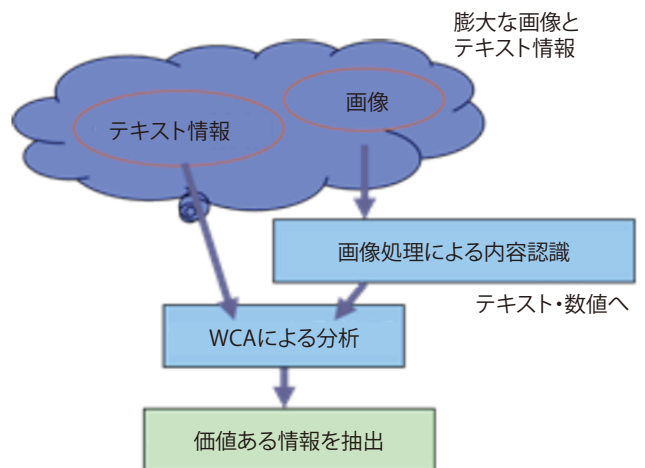


図8 WCAを使った Twitter からの情報抽出

に処理し高度な分析機能を提供するIBMのテキストマイニング・プラットフォーム・ソリューションであるWCA (Watson Content Analytics) を使って付随するコメント情報と結合し、分析を行う[5]。WCAでは自然文で書かれたテキストを入力し、テキスト部分の言語解析とメタデータ生成とを自動的に行う。言語解析では、形態素解析で単語を区切るとともに文法上の特徴を分析し、構文解析で単語間の係り受けを特定する。

図8に示す分析の準備として必要なTwitter画像のテキスト化および付随コメントとの結合について図9に示す。あらかじめ登録した商品画像のデータベースと照合して画像特徴量が最も一致する商品名を決定し、このTwitterのメタデータファセットとして登録する。付随するコメントは品詞分析および形態素解析を行ってファセット分類し登録する。この2種類のファセットを合わせてWCAによるテキストマイニングを行う。ここで、ファセットとは宝石のカット面を意味し、分析でのカテゴリを表す。

Twitterを用いた商品の評判分析は次の手順にて行うことができる。まず、Twitterのテキスト情報から目的の商品名を含むものを検索し収集する。これを第一群とする。次に大規模商品識別システムにより、Twitterから得られる画像（時期や地域を限定してTwitterから収集した画像）をあらかじめ登録した商品画像と照合し、目的の商品が含まれる画像を選び出す。その画像に商品名を関連付けし（すなわち画像情報をテキスト化し）、付随するコメントとテキスト化した画像情報とを合わせたセットを収集する。これを第二群とする。最後に、第一群と第二群とを合わせてテキストマイニングを行う。この手法により、コメントだけを抽出する評判分析（第一群のみによる評判分析）では分から

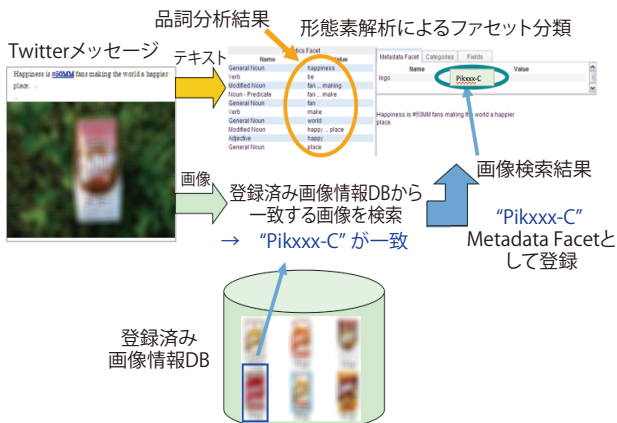


図9 WCAを使ったTwitterからの情報抽出および付随コメントとの結合

なかった知見を獲得することが可能となり、早期に商品の改良や新商品の開発につなげることができる。

この評判分析の一例として、ある商品（たとえば乳飲料の新商品）を販売開始した後、2週間の間にTwitterに投稿されたテキストと画像とを収集して該当する商品に関する評判を分析した結果を表1に示す。画像中に存在する商品とテキストに現れる形容詞を使用して、相関分析を行った。このように、2種類のカテゴリを使った相関分析をファセットペア分析と呼ぶ。この表において、各行は商品名を、各列はテキストの中で使用されている品詞の中でも特に形容詞について示している。表の各セルに示される数値は、上側が頻度、下側が相関値である。相関値は全文書集合に対する分析母集団でのキーワードの割合比で表され、行・列に示される2つの項目がペアで使用されている確率(共起確率)が母集団の場合の共起確率よりも高い場合、1.0以上の数値となる。たとえば、商品Aは、“かわいい”というコメントと一緒に使われている頻度が27である。その相関値は1.2であり、母集団の傾向よりも共起確率が高いことが分かる。また、この表に示される商品群の中ではほかを大きく上回り、最も高い共起確率であることが示されている。すなわち、商品Aは“かわいい”という評価が他製品より勝っていることが分かる。

ところで、画像情報のテキスト化においては、目的の商品名だけではなく、そのほかの情報も関連付けすることによってさらに深い分析が可能となる。たとえば、①商品の数、②一緒に写り込んでいるほかの商品名、③一緒に写り込んでいるシーン(風景、環境)や人(性別、年齢層)などの情報である。①・②の情報については大規模商品識別システムにより抽出することができる。一方、③については特定の商品識別ではない一般物体

表1 Twitterにおける写真(商品)とテキスト(形容詞)とのファセットペア分析

サブファセット値	かわいい	おいしい	大好きだ	美味しい
33	25	5	4	
商品 A	27	18	2	0
47	1.2	0.9	0.1	0.0
商品 B	3	1	0	2
24	0.1	0.0	0.0	0.2
商品 C	3	3	0	3
13	0.1	0.1	0.0	0.9
商品 D	3	1	2	0
12	0.1	0.0	0.2	0.0
商品 E	2	0	0	0
7	0.1	0.0	0.0	0.0
商品 F	3	0	0	0
5	0.3	0.0	0.0	0.0

認識やシーン認識などの画像処理技術を使用する。このような画像処理応用は大規模データを用いた機械学習や深層学習により精度向上が進んでいる。今後は、画像から抽出される多数の情報が意図的に含まれたものかそうでないのかを付随するテキスト情報から推定して取捨選択するというマルチモーダル・マイニングも必要になると考える。

マーケティング・リサーチの観点では、Twitterに添付された写真において商品のどの部分が注目されているかも非常に重要である。たとえば、商品のパッケージのある部分を中心として撮影された写真のコメントが「かわいい」であれば、商品そのものでなくパッケージの一部に興味を示され「かわいい」という感想を持たれたことが明白である。このような情報を多数収集することにより、新製品の開発のヒントを得ることもできる。

このように、マーケティング・リサーチの分野でも画像とテキスト情報とを合わせたマルチモーダル・マイニングは重要な役割を果たすことが期待できる。

4. メディカル・マルチモーダル・マイニング

医療やヘルスケアの分野においても画像を定量化しほかの情報と合わせて潜在的な知見を抽出するマルチモーダル・データマイニングの試みが始まっている。医師の経験に基づいて目視により行われてきた診断基準を数値化することによって、医用画像も客観的な数値データに置き換えることができる。医用画像はその種類により特徴が異なり、また何を知りたいかによって抽出する特徴量も変わると考えられるため、画像の種類に応じて専門医師の診断に基づいた構造化データとしての定量化が必要となる。

医用画像の中でも内視鏡画像は定量的な解析が最も遅れているといわれる。CT, MRI, PET, 超音波診断装置などでは直接観察できず、データ変換して可視化することが必要なこともあり視認性を高めるための多くの研究がなされてきた。一方、カメラの画像を直接取得できる内視鏡画像においては、FICE (Flexible spectral Imaging Color Enhancement) 拡大内視鏡を使った拡大・強調画像での癌の有無の自動判別研究[6]の事例があるものの、通常の広い視野で使用する内視鏡においては定量化の研究がほとんどなされていない。ここで1つの研究事例について紹介する[7]。

胃における扁平な腫瘍には、白色光による通常内視

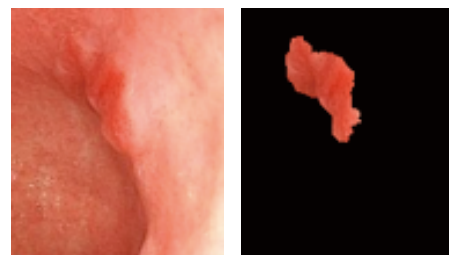
鏡観察によって色の観点から以下のような種類に分類され、組織学的な診断をゴールドスタンダード（すなわち、顕微鏡下で細胞レベルでの癌かどうかの判定を基準として）として、癌の有無（EC：early cancer / LGA：low grade adenoma）を感度64%、特異度94%で画像から診断できることが報告されている[8]。

- red-color group : 悪性度が高い
- not-red-color group : 悪性度が低い

この場合の“赤い／赤くない”の判定は通常2名以上の専門医により目視によって定性的になされている。赤いほど悪性腫瘍、白いほど良性腫瘍の確率が高いとの報告がある。この判定を画像分析により定量的に行うことを目的として、内視鏡の原画像から専門医の注目する個所（腫瘍の疑いのある個所）を切り出し、その個所の画像特徴量の抽出を試みた。

(1)と(2)の例においては専門医が注目する領域と、切り出した腫瘍個所とは比較的合致した(図10, 図11, 図12)。H (Hue: 色相) 色成分を比較すると(1)の方が(2)よりも小さい値を示しており、これは(1)の

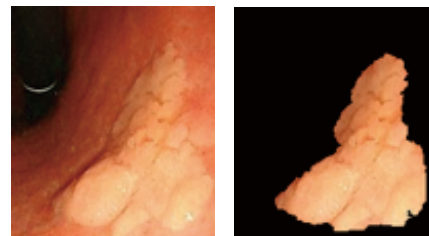
(1) red-color groupの胃の腫瘍の抽出



(a) 原画像 (b) 腫瘍個所の画像

図10 red-color groupの胃の腫瘍

(2) not-red-color groupの胃の腫瘍の抽出



(a) 原画像 (b) 腫瘍個所の画像

図11 not-red-color groupの胃の腫瘍



(1) H: 0.014 (2) H: 0.069

(色相H: 0~1に正規化された値を使用)

図12 腫瘍個所の平均色の比較

方が赤いことを示している。H色成分を使って赤さの程度を判定することが可能である。

このように腫瘍個所を画像から抽出し定量化できる可能性があり、これによって従来、定性的に“赤い／赤くない”と説明していたものを数値で表すことが可能となる。色以外の特徴量として、注目個所の面積や位置、また経時変化などの項目も有用である。これらのデータはほかの医療データと合わせてメディカル・マルチモーダル・データマイニングに活用できると考える。

患者の自覚症状や検査結果（生体データや医用画像）、診断結果、治療効果、さらに患者の性別・年齢・嗜好・住む地域などパーソナルな情報などと合わせて関係性を見出すことによって、それは価値あるエビデンスとなり、医師や研究者が効果的に効率的に情報を共有し、治療支援や教育支援、医学研究の役に立つものとなり得る。

5. 今後の展開

第2章および第3章において、商品識別技術および、その技術とテキスト・マイニングとを統合したマーケティング・リサーチの事例について紹介した。また第4章ではマルチモーダル・マイニングを医用画像に適用した例を紹介した。本章では、今後の展開として、画像情報と音声情報とを組み合わせた対話型ロボットによる商品レコメンデーション技術の展望について述べ、あわせて、近年ディープ・ラーニング技術の進展により飛躍的な精度の向上が期待される医用画像解析とほかの情報とを統合した診断支援システムの展望について述べる。

5.1 画像と音声情報とを組み合わせたマルチモーダルな対話型の商品レコメンデーション

第2章および第3章で述べた画像識別技術と音声認識技術とを組み合わせると、マルチモーダルな対話型の商品レコメンデーションに利用できる可能性がある。たとえば、顧客が本を持参して書店に出向き、書店に設置されている対話型ロボットと対話することにより、持参した書籍のカテゴリで顧客の嗜好に即したほかの書籍をレコメンデするというユースケースが想定される。具体的には、まず、対話型ロボットがカメラにより顧客の性別・年齢層を推定し、あわせて、持参した書籍を識別する。次に、顧客との対話により嗜好を聞



図13 性別・年齢層の推定および書籍の識別例

き出すことにより、レコメンデする書籍の棚に顧客を案内するというシナリオである（図13）。

Watsonと呼ばれるコグニティブ・システムが提供する Visual recognition API, Speech to text API, Dialog API および Text to speech API を対話型ロボットが呼び出すことにより、このようなシナリオが実現可能である[9]。

5.2 医用画像とほかの情報とを組み合わせた診断支援

近年、物体や音声の認識において深層学習が成功を収めており、医用画像処理においても畳み込みニューラルネットワーク（Convolutional Neural Network, 以下CNN）と呼ばれる多層ニューラルネットワークの利用が盛んになっている[10]。CNNは順伝播型ニューラルネットワークの一種であるが、畳み込み層とプーリング層の連結が繰り返される特別な構造を持つ。畳み込み層はフィルタを構成しており、入力から特徴を抽出する働きがあり、プーリング層は畳み込み層の直後に挿入され、畳み込み層で抽出された特徴の位置感度を低下させる働きがある。畳み込み層とプーリング層の繰り返しの後には、隣接層間のユニットがすべて密に結合した全結合層が複数連続して配置される。

内視鏡画像における胃の腫瘍の悪性度の定量化でも述べたように、これまでは、病変部の色や形態的特徴を抽出するアルゴリズムの実装が性能を左右したが、深層学習ではそれを必要としない。すなわち、医学的な知識なしに、多数のデータを用いて畳み込みニューラルネットワークを訓練するだけで病変部を判別するモデルを構築できる。

最新の低線量X線CT装置は、高精細な画像（たとえばスライス厚1mm間隔）を短時間に撮影でき、1回のスキャンあたり300枚程度のスライス画像を生成でき

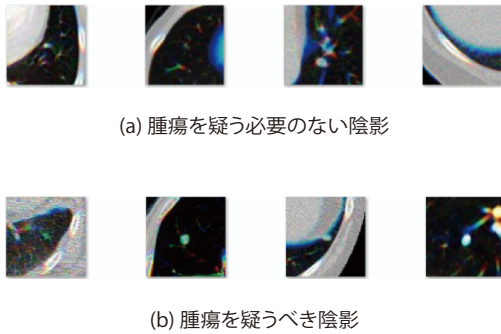


図 14 腫瘍候補画像 64 pixel × 64 pixel × 3 slice

る。図 14 は最新の低線量 X 線 CT 装置で撮影した胸部 CT スキャン画像データセット [11] から抽出した肺癌のスクリーニング用の画像の例である。(a) は各画像の中心部に特徴的な陰影があるが、それは腫瘍を疑う必要のない陰影であり、(b) の各画像の中心に見られる特徴的な陰影は腫瘍を疑うべき陰影である。

しかし、肺 CT 画像中に表れる陰影の大半は血管や肋骨など腫瘍とは無関係な陰影であり、腫瘍を疑われる陰影はごく少数である。したがって、CT 画像をくまなく読影し、稀に見られる大きさが数ミリ程度の初期の腫瘍を疑うべき陰影を見つけ出すことは、放射線科医の負担が大きく、胸部 CT 画像から腫瘍を疑う陰影を自動的にスクリーニングする Computer-aided Diagnosis (CAD, コンピュータ支援診断) system の開発が望まれている。

Setio らの報告 [12] によれば、CNN が胸部 X 線 CT 画像から肺癌を疑うべき部位をスクリーニングする性能は、特徴量を抽出するこれまでの方法を上回っている。

ほかの研究事例として、IBM Research では、米国において主要な死因の 1 つとされる心疾患に焦点を当て、電子カルテや健康診断結果などの医療データを分析し、早期のリスクの発見を促し、より正確な臨床診断を支援するための有用な示唆を提示する Advanced Analytics for Information Management (AALIM) の開発を 2007 年から継続的に進めている [1]。電子カルテデータあるいは健康診断データは本質的にマルチモーダルであり、単純 X 線画像、心電図、心音、X 線 CT 画像、MRI 画像、PET 画像、血液検査結果、医師による所見、病歴、投薬など多岐に渡るデータが含まれている。これらマルチモーダルデータを分析し、各モーダルデータの特徴量を抽出し、それらを統合し、過去の心臓病罹患者の膨大な情報から類似度の高い症例を探索し、早期に病気の兆候を捉えたり、治療方針や処方箋を提示するシステムが開発されている。

当初、心臓病に特化して開発が進められてきたが、今後、癌や糖尿病あるいはアルツハイマー病などの疾患にも適用することができると考えられ、同様のデータの予測分析によって早期発見と治療への誘導が迅速に行われることが期待されている。

6. おわりに

本稿では、筆者らが取り組んでいるマルチモーダル・マイニングの事例を紹介し、今後の展望について述べた。マルチモーダル・マイニングは、マーケティング・リサーチや診断支援の分野で利用が始まっており、対話型ロボットによる商品レコメンデーションなどの新しい応用分野において、これからますます画像情報とほかの情報とを統合したマルチモーダル・マイニング技術の発展およびその実用例が広がることが期待される。

謝辞 特に第 4 章メディカル・マルチモーダル・マイニングの研究にあたり、ご指導・ご助言いただきました横浜市立大学 医学教育学名誉教授 後藤英司先生、医学教育学教授 稲森正彦先生、消化器内科学教授 前田慎先生、医学教育学講師 飯田洋先生に深く感謝申し上げます。

参考文献

- 1) Syeda-Mahmood, T., Wang, F., Beymer, D., Amir, A., Richmond, M. and Hashmi, S.: AALIM: Multimodal Mining for Cardiac Decision Support, *Computers in Cardiology*, pp.209-212 (2007).
- 2) Nakano, H., Mori, Y., Morita, C. and Nagai, S.: Large Scale Specific Object Recognition by Using GIFTS Image Feature, *ICIAP*, No.2, pp.36-45 (2015).
- 3) Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol.60, No.2, pp.91-110 (2004).
- 4) Arya, S., et. al.: An Optimal Algorithm for Approximate Nearest Neighbor Searching in Fixed Dimensions, *Journal of the ACM*, Vol.45, No.6, pp.891-923 (1985).
- 5) Zhu, W., et. al.: IBM Watson Content Analytics: Discovering Actionable Insight from Your Content, *An IBM Redbooks Publication*. ISBN-10:0738439428 (2014).
- 6) Miyaki, R., Yoshida, S., Tanaka, S., Kominami, Y., Sanomura, Y., Matsuo, T., Oka, S., Raytchev, B., Tamaki, T., Koide, T., Kaneda, K., Yoshihara, M. and Chayama, K.: Quantitative Identification of Mucosal Gastric Cancer under Magnifying Endoscopy with Flexible Spectral Imaging Color Enhancement, *J Gastroenterol Hepatol*, Vol.28, No.5, pp.841-847 (2013).
- 7) 森 由美: 機能性ディスぺプシアにおける内視鏡的十二指腸炎のインパクト: 上部消化管内視鏡画像の定量的評価とメディカル・マルチモーダル・マイニング, 横浜市立大学大学院医学研究科博士論文, pp.17-20 (2015).
- 8) Maki, S., Yao, K., Nagahama, T., Beppu, T., Hisabe, T., Takaki, Y., Hirai, F., Matsui, T., Tanabe, H. and Iwashita, A.: Magnifying Endoscopy with Nar-

row-band Imaging is Useful in the Differential Diagnosis between Low-grade Adenoma and Early Cancer of Superficial Elevated Gastric Lesions, Gastric Cancer, Vol.16, No.2, pp.140-146 (2013).

- 9) 清水周一, 井上忠宣, 高野光司: コグニティブ・ロボット最前線—IBM 東京基礎研究所による「賢い」ロボットへの挑戦, PROVISION, No.90, pp.30-32 (2016).
- 10) Krizhevsky, A. et al.: ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems, No.25, pp.1097-1105 (2012).
- 11) Lung Nodule Analysis 2016, Retrieved June 24, 2016, <http://luna16.grand-challenge.org/> (2016年12月20日現在)
- 12) Setio, A., et al.: Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks, IEEE Transactions on Medical Imaging, Vol.35, No.5, pp.1160-1169 (2016).

中野 宏毅 (非会員) hnakano@jp.ibm.com

日本アイ・ビー・エム (株) 東京ラボラトリー・ワトソンヘルス開発所属. 1983年大阪大学基礎工学部卒業. 同年, 日本アイ・ビー・エム (株) 入社. 画像認識技術およびマルチモーダル・マイニングの研究開発に従事. 2015年より現職. 学術博士, 電子情報通信学会, 精密工学会各会員.

阪本 正治 (非会員) sakamoto@jp.ibm.com

日本アイ・ビー・エム (株) 東京ラボラトリー・ワトソンヘルス開発所属. 1989年電気通信大学大学院通信工学専攻修士課程了. 同年, 日本アイ・ビー・エム (株) 東京基礎研究所入所. 音声情報処理, 並列計算処理, セキュリティ, 半導体パターンニング, 3D プリテンディングなどの研究に従事. 2016年より現職. 数理科学博士, IEEE 会員.

森 由美 (非会員) moriyumi@ynu.ac.jp

横浜国立大学 未来情報通信医療社会基盤センター所属. 2015年横浜市立大学大学院医学研究科博士課程修了. 医工融合, レギュラトリーサイエンス, メディカル・マルチモーダル・マイニングの研究開発に従事. 2015年より現職. 博士 (工学), 博士 (医学), 映像情報メディア学会, 感察工学研究会各会員.

採録決定: 2016年12月20日

編集担当: 齋藤 忍 (日本電信電話 (株))