

画像特徴とテキスト特徴の統合による Web スポーツニュース画像のイベント分類

北原 章雄[†] 奥山 幸也[†] 柳井 啓司[†]

[†] 電気通信大学大学院 情報工学専攻 〒182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: [†]{kitaha-a,okuyama-k}@mm.cs.uec.ac.jp, ^{††}yanai@cs.uec.ac.jp

あらまし 現在, Web 上に配信されている画像付きニュースは, ニュース内容や画像の内容によって細かく分類して配信されていないことが多いため, ユーザが目的のニュースを検索することが困難となっている. そこで, 本研究では Web 画像ニュースの中のスポーツニュースを対象とし, これらスポーツニュースをスポーツの種類と, 画像の内容によりプレイ中か非プレイ中かのスポーツイベント分類を行う方法を提案する. 2 段階手法はテキスト特徴によりスポーツ分類を行った後に画像特徴を Multiple Kernel Learning (MKL) によって統合してイベント分類を行う方法であり, 1 段階手法はテキスト特徴と画像特徴を同時に MKL によって統合することでスポーツイベント分類を同時にを行う方法である. 実験では, Yahoo!JAPAN 写真ニュースの記事を対象に野球, ゴルフ, F1, サッカー, テニス, 相撲のプレー中/非プレー中の 12 種類に分類した. 2 段階手法スポーツ分類では分類率 99.33%を達成し, イベント分類では野球 81.5%, ゴルフ 67.5%, F1 91.0%, サッカー 81.0%, テニス 71.0%, 相撲 92.5%という分類率になり, ゴルフ以外で特徴統合により特徴単独よりも高い分類率を得た. 一方, 1 段階手法では全体の分類率 77.08%を得た.

Event Classification of Sport News Photos by Integrating Textual and Visual Features Akio KITAHARA[†], Kouya OKUYAMA[†], and Keiji YANAI[†]

[†] Department of Computer Science, The University of Electro-Communications

Chofugaoka 1-5-1, Chofu-shi, Tokyo, 182-8585 Japan

E-mail: [†]{kitaha-a,okuyama-k}@mm.cs.uec.ac.jp, ^{††}yanai@cs.uec.ac.jp

Abstract There are many Web news sites which deliver us news photos everyday. However, since they classified news photos only very roughly, it is not easy for us to search for photos we want from a huge amount of news photos. In this paper, we treat with event classification of sports news photos as an instance of researches on more sophisticated search methods for large-scale photo news databases. We propose two methods to classify sports news photos into one of six sports genres and to discriminate in-play photos from not-in-play ones. One is the two-step method which classifies sport genres first and recognizes in-play conditions next, and the other is the two-step method which classifies them simultaneously. In the proposed methods, we integrate textual features extracted from news articles and image features extracted from photo images by Multiple Kernel Learning (MKL). In the experiment of the two-step method, we obtained 99.33% as the classification rate for the genre classification which is the first step and 80.75% for the in-play classification which is the second step. On the other hand, in the experiment of the one-step method, we obtained 77.08% which was a little less than the result by the two-step method.

1. はじめに

現在, Web 上において毎日大量のニュース記事が配信されており, その速報性から新聞やテレビニュースよりも Web から情報を得ている人も多い. また, 近年のハードディスクの大容量化により, Web ニュースを収集・蓄

積することで一般個人がニュースライブラリ (データベース) を持つことも可能になった. また, データベースに登録する際にインデックス付けなどを行っておけば, Web ニュースは新聞やテレビニュースに比べて遥かに容易に情報検索ができる. このような大規模で容易に検索可能なニュースライブラリの作成は, 新聞やテレ

ビニュースではできないことであった。以上を考えると Web ニュースは今後最も有用なニュースデータベースになっていくと思われる。

Web ニュースには文章だけでなく画像を含んだ Web 画像ニュースも存在する。画像は文章だけではわからない内容を視覚的に捉えることができるため、ニュース記事内容のさらなる理解の手助けとなる。今後 Web 画像ニュースはますます増加すると考えられる。しかしこれらの Web 画像ニュースはニュース内容により詳しく分類して配信されていないことが多く、また画像の内容によっても分類されていない。ニュースデータベースを検索するときのクエリとしては、「○○に関するニュース」「○○の画像がついているニュース」などが考えられるが、分類されていない Web 画像ニュースをそのままデータベースに登録してしまうと検索が困難となってしまう。これを解決するためには、1つ1つの Web 画像ニュースを手で分類してデータベースに登録するしかないが、非常に手間のかかる面倒な作業となってしまう。そこで Web 画像ニュースを自動で認識・分類して簡単に検索できるようにしたいというニーズが高まっている。このような自動分類が高精度で行えるようになれば、将来的に「イチローがボールを打っている画像の付いたニュース」というクエリによる検索も可能となる。

そこで、本研究では Web 画像ニュースの中のスポーツニュースを対象とし、これらスポーツニュースをスポーツの種類と、画像の内容によりプレイ中か非プレイ中かのスポーツイベント分類を行う。その方法として、2段階手法と1段階手法の2種類の手法を提案する。2段階手法では、まずニュース記事から抽出したテキスト特徴によりスポーツ分類を行い、次に画像から抽出したいくつかの画像特徴を重み付きで統合して使用することにより分類を試みる。1段階手法ではテキスト特徴と画像特徴を重み付きで同時に統合することにより、スポーツイベント分類を1度に行う。実験では Yahoo!JAPAN 写真ニュース [1] に対して2つの手法を適用し、その有効性を確認する。分類するスポーツは野球、ゴルフ、F1、サッカー、テニス、相撲の6種類を用いる。

ここで、本研究で対象とする Yahoo!JAPAN 写真ニュースについて紹介する。Yahoo!JAPAN 写真ニュースは国内、海外、経済、エンターテインメント、スポーツ、テクノロジー、地域の7種類のカテゴリで配信されているが、それ以上詳細な分類はされておらず、また画像の内容に関連した配信も行われていない。本研究ではスポーツカテゴリとして分類されているニュースを対象として扱う。各スポーツニュースは図1のように配信されている。

2. 関連研究

Quattoni ら [2] は構造的学習フレームワークの導入により Web 画像ニュースのジャンル分類を行っている。彼らは局所特徴の bag-of-keypoints 表現に、ニュースキャプションから作成した補助問題から得た行列を掛け合



図1 Yahoo!JAPAN 写真ニュース スポーツ

せることで、テキストまで含めた新しい画像表現方法を提案している。分類対象はスポーツに限定していない。Web 画像ニュースのスポーツ分類の研究としては Jain らの研究 [3] がある。彼らは観客席から一般人が撮影したスポーツ画像を Web から収集し、画像に写っているスポーツ競技面を認識することによりスポーツ分類を行った。彼らのデータセットは競技面から離れた観客席から取られた写真のため、画像全体に占める競技面の割合が大きく、写っている選手が小さいという特徴がある。しかし本研究で使うデータセットは Web ニュース画像であるため、プロカメラマンがスポーツ選手の近くで撮影したものが多い。そのため競技面は大きく写っておらず、彼らの手法では分類が行えないと思われる。Web 画像ニュースではないがスポーツ分類の研究として Li らの研究 [4] がある。彼らは画像全体のシーン認識と画像領域の物体認識を統合することで、Web から収集したスポーツ画像の分類を行った。彼らはクエリ画像に対して、スポーツ分類結果だけでなく、シーンと物体の認識結果も同時に示すことを目指している。ここで紹介した研究では、Web 画像ニュースや Web から収集した画像を、画像特徴やテキスト特徴を用いてジャンル分類を行っている。これらの研究と本研究の1番の大きな違いは、ジャンル分類を行った後に画像内容によりさらに詳しく分類しているかどうかである。本研究では、Web 画像ニュースを各スポーツに分類するだけでなく、プレイ中であるかどうかまで認識し分類を行う。

3. 提案手法概要

本研究ではスポーツイベント分類を行う方法として、2段階手法(図2)と1段階手法(図3)の2種類を提案する。

2段階手法では、まずスポーツ分類ステージにおいてニュース記事からテキスト特徴を抽出する。テキスト特徴としては bag-of-words を用いることにする。そして抽出したテキスト特徴を用いて Support Vector Ma-

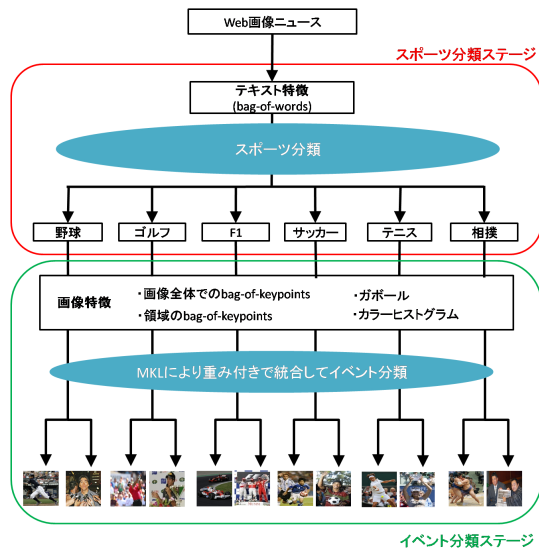


図 2 2段階手法の処理の流れ

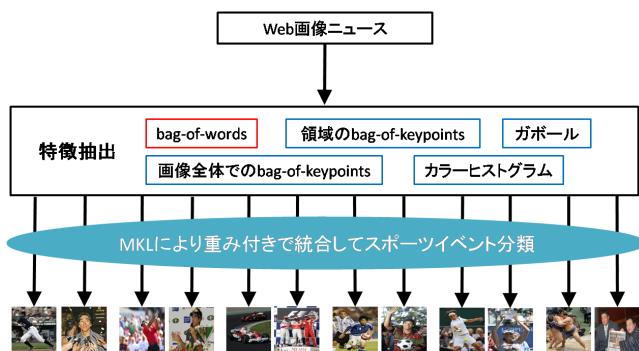


図 3 1段階手法の処理の流れ

chine (SVM) によりスポーツをマルチクラス分類する。次にイベント分類ステージにおいてニュース画像から画像特徴を抽出する。画像特徴としては、画像全体の bag-of-keypoints、画像領域での bag-of-keypoints、画像を 4×4 グリッドに分割し各グリッドから作成したガボール特徴ヒストグラム及びカラーヒストグラムの4種類を用いる。これらの特徴を Multiple Kernel Learning (MKL) により重み付きで統合することにより、それぞれのスポーツ画像を各イベントへとシングルクラス分類する。画像全体から抽出した bag-of-keypoints と、領域分割した領域から抽出した region-based bag-of-keypoints を MKL を用いて統合することは従来行われておらず、初めての試みといえる。

1段階手法では、ニュース記事から抽出したテキスト特徴と、ニュース画像から抽出した画像特徴を MKL により同時に統合することで、スポーツイベント分類を1度で行う。各スポーツのプレイ中/非プレイ中のことをクラスと呼ぶことにするとクラスは6スポーツ $\times$ 2 イベント=12個あるため、このスポーツイベント分類は12クラスのマルチクラス分類となる。

次章以降で各特徴と統合方法について詳しく述べる。

手順 1 bag-of-words によるテキスト分類

step1: 単語頻度カウント

全テキストを品詞分解し、テキストごとに単語頻度を得る。

step2: codebook 作成

学習テキストの単語頻度をもとに、codebook を作成する。

step3: 単語ヒストグラム作成

codebook をもとに学習テキスト、テストテキストの単語頻度ヒストグラムを作成する。

step4: 分類

ヒストグラムを分類器にかけ、テキスト分類を行う。

4. 提案手法詳細

4.1 特徴量

4.1.1 Bag-of-words

Bag-of-words は、正確に言うとはテキスト特徴ではなくテキスト分類手法である。これは文章を単語 (word) の集合とみなし、出現順を無視した単語の頻度をそのテキストの特徴として分類を行う手法である。アルゴリズムは簡単であるが精度が高いことが知られている。手順 1 に bag-of-words によるテキスト分類の流れを示す。まず全テキストを品詞分解し、各テキストごとに出現する単語の頻度を数える。次に学習テキストの単語頻度上位 N 個の単語を codebook とする。そして codebook 中の単語がいくつ出現したかを数えることにより、学習テキスト、テストテキストの単語頻度ヒストグラムを作成する。最後にこれらの単語頻度ヒストグラムを分類器によって分類する。

2段階手法のスポーツ分類ステージは、bag-of-words の全 step と等しいものとなっている。また、1段階手法において bag-of-words をテキスト特徴で使うときは、step3 までで得られたヒストグラムをその Web 画像ニュースのテキスト特徴とする。

単語頻度カウントでは、日本語形態素解析ツール「茶筌」[5] を使って学習・実験ニュース記事テキストをすべて品詞分解する。その結果、品詞が名詞、動詞、形容詞、副詞、形容動詞の単語の出現頻度を数える。ただしストップワードと呼ばれる、どのテキストにもよく出てくる単語は、分類精度を低下させる原因となるので除外することにする。今回用いるストップワードは以下のものとする。

ストップワード

ひらがな・カタカナ1文字のみ、する、なる、れる、いる、られる、ため、できる、せる、いう、ない、やすい、これ、それ、あれ、どれ、ある、おる、どう、ここ、そこ、あそこ、どこ、この、その、あの、どの、的、場合、以下、とき、さらに、ところで、そして、もし、しかし、だが、行、もう、そう、たち、以外、考える、思う、年、月、日、人、さん、氏、こと、EOS、 、REUTERS、ロイター、配信、時事通信社、毎日新聞、スポーツナビ、西日本新聞社、EPA =

ロイター、時事通信社、毎日新聞などはニュースの配信元であり、ニューステキスト中に頻繁に登場するためストップワードとして登録してある。また、本研究で利用する Web 画像ニュースデータセットは、事前に html タグを除去したニュース記事本文のみを使用しているが、除去ミスの際に残ってしまった html タグもストップワードに含んでいる。これらは本研究データセット独自のストップワードといえる。

4.1.2 画像全体での bag-of-keypoints

画像から SIFT 特徴を抽出し bag-of-keypoints 表現で表すことは、現在の一般物体認識において最もスタンダードな方法であり、精度も良いとされている。bag-of-keypoints は画像を局所特徴 (keypoint) の集合とみなし、位置情報を無視した局所特徴の出現頻度を用いて画像の分類を行う手法である。まず全学習データから抽出した局所特徴をクラスタリングして codebook を求める。Codebook の各ピンは visual word と呼ばれ、画像から抽出した局所特徴を codebook 中の一番近い visual word に投票することで各画像の局所特徴ヒストグラムを作成する。このヒストグラムは、各ピンをその画像の局所特徴の総数で割ることにより正規化が必要である。局所特徴としては SIFT (Scale Invariant Feature Transform) [6] がよく使われる。SIFT は画像のスケール変化 (拡大縮小)、回転や視点変化に不変な局所特徴であり、特徴点の周りの局所画像パターンを 128 次元の特徴ベクトルとして表現する。SIFT 特徴の抽出にはフリーソフト SIFT++ [7] を使用し、検出方法にはランダム点検出を用いる。ランダムに検出する特徴点の数は 2000 個とする。

a) 領域分割を用いた bag-of-keypoints

Bag-of-keypoints では画像全体から 1 つの特徴ヒストグラムを作成するため、画像中の認識対象の領域以外の背景領域からも特徴を抽出している。そこで、認識対象の領域のみから特徴を抽出する方法を考える。具体的には、画像を領域分割して分類に重要な領域を推定し、その領域を認識対象の領域と捉えることにする。分類に重要な領域を選択したら、その領域を bag-of-keypoints により表現して画像特徴として利用する。

画像の領域分割には教師なし領域分割手法である JSEG [8] を用いる。本研究では Web 上に公開されているソースコードに手を加え、画像全体に占める最小領域の割合と、分割する領域数を指定して分割を行う。ただし画像によっては指定した領域数と等しい分割が行われないこともある。分割を行った後は学習画像にしか写っていないような分類に有効な領域を選択する必要がある。その方法として、本研究では mi-SVM [9] を用いる。mi-SVM ではポジティブバッグとネガティブバッグを SVM により学習し、学習したモデルを使ってポジティブバッグを分類する。その結果、各画像で最も SVM 出力値の高い領域がポジティブインスタンス、つまり重要領域であるとする。mi-SVM による重要領域の選択の様子を図 4 に示す。

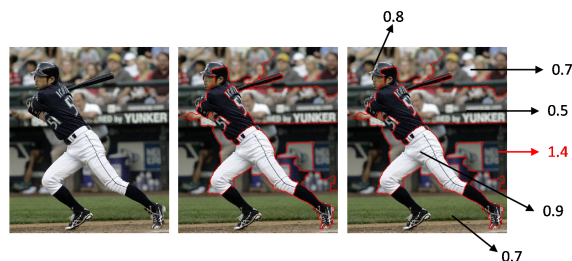


図 4 mi-SVM による領域の選択 (左:元画像. 中:領域数を 5 としたときの JSEG 結果. 右:mi-SVM によって選ばれた領域)

mi-SVM によって学習画像から重要領域を選択したら、その領域から抽出した SIFT 特徴を使い、codebook を作成する。その後の処理は通常の bag-of-keypoints と同様である。

b) ガボール特徴

ガボール関数を用いて画像をたたみこむ (ガボールフィルタリングをかける) ことにより、画像から抽出された局所的な濃淡情報のことである。本研究では 6 方向、4 周期のフィルタカーネルを用いてガボール特徴を抽出する。さらに画像を 4×4 に分割して各ブロックでガボール特徴を抽出することにより、画像から $6 \times 4 \times 4 \times 4 = 384$ 次元のガボール特徴を抽出する。

c) カラーヒストグラム

本研究では一般的によく使われる RGB 色空間によるカラーヒストグラム (RGB ヒストグラム) を用いる。RGB 色空間では、各色は R(赤), G(緑), B(青) の 3 つのパラメータにより $(R,G,B)=(0,0,0) \sim (255,255,255)$ として表される。これをそのまま特徴ベクトルとして表現すると、ベクトルの次元が 256^3 次元になってしまい実用的ではない。そこで、本研究では各色を 4 分割し、0~63 を 0, 64~127 を 1, 128~191 を 2, 192~255 を 3 として表すことにより、 $4 \times 4 \times 4 = 64$ 次元で表すことにする。さらに画像を 4×4 に分割して各ブロックでヒストグラムを作成することにより、画像を $64 \times 4 \times 4 = 1024$ 次元で表すことにする。

4.2 分類手法 (Support Vector Machine)

本研究では、2 段階手法のイベント分類においてシングルクラス分類を、2 段階手法のスポーツ分類ステージと 1 段階手法においてはマルチクラス分類を行う必要がある。その分類器として Support Vector Machine (SVM) を使用する。SVM は 2 値分類問題を解くために考えられた学習アルゴリズムであり、基本的に線形識別器である。SVM では、分離超平面と最も近いテストデータとの距離であるマージンが最大になるように分離超平面を求める。このとき、ある程度の分離ミスは認めるというソフトマージンが通常は用いられる。SVM はカーネルトリックにより非線形識別器に拡張することが可能である。本研究では、2 段階手法のイベント分類用シングル

クラス分類器に線形カーネルを使用する。

SVM を使用してマルチクラス分類をするためには 1-vs-rest 分類法を行う必要がある。つまり、あるクラスを正、それ以外を負としてシングルクラス分類器を学習するというを、すべてのクラスに対して行うのである。そしてテスト画像をすべての分類器に入力し、最も出力値の大きいクラスに分類する。本研究では χ^2 カーネルを用いたシングルクラス SVM を使用して 1-vs-rest マルチクラス分類を行う。 χ^2 カーネルの式は以下で表される。

$$K(a, b) = \exp(-\gamma * \sum_{i=1}^n \frac{|a_i - b_i|^2}{a_i + b_i}) \quad (1)$$

4.3 特徴統合手法 (MKL)

4 種類の画像特徴を統合する方法として Multiple Kernel Learning (MKL) を用いる。MKL とは複数のカーネル (サブカーネル) を線形結合することにより、新たなカーネルを求める方法である。定式化すると以下のようになる。

$$K_{combined}(\mathbf{x}, \mathbf{x}') = \sum_{j=1}^K \beta_j k_j(\mathbf{x}, \mathbf{x}') \\ \sum_{j=1}^K \beta_j = 1, \beta_j \geq 0 \quad (2)$$

MKL では、各サブカーネルに対する重み β_j をどのように設定するかにより、結合してできた新たなカーネルの出力値が異なってくる。そのため最適なサブカーネルの重み β_j をどのように求めるかが問題となる (MKL 問題)。これを解決する最も簡単な方法は、すべての β_j の組み合わせを実際に試して、最もよい結果の重みを採用することである。しかし、サブカーネルの数が増えるにつれて β_j の組み合わせの総数は指数的に増加する。また、 β_j の刻み幅を小さくすればするほど、 β_j の組み合わせの総数はさらに増加する。このように計算量の問題で、重みの総当たりによる解決法では最適な重みを求めることは実際には困難である。そこで Sonnenburg ら [10] によって、SVM 学習を反復することによって最適な β_j を求める方法が提案されている。Varma ら [11] は Sonnenburg らの方法を用いて、画像認識における特徴統合に MKL を用いた。Varma らは MKL のサブカーネルを各画像特徴と対応付けることで、MKL によるカーネルの統合を特徴の統合としている。本研究でもこれと同様に MKL による画像特徴の統合を行う。

5. 実験

5.1 データセット

収集した Yahoo!JAPAN ニュースの中から、学習データとして 2007 年、テストデータとして 2008 年のニュー

スをデータセットとして利用した。スポーツは野球、ゴルフ、F1、サッカー、テニス、相撲の 6 種類である。ニュースは各スポーツごとにプレイ中 100 個、非プレイ中 100 個とし、合計で 100×2 (プレイ中、非プレイ中) $\times 6$ (野球、ゴルフ、F1、サッカー、テニス、相撲) $\times 2$ (学習、テスト) = 2400 個を用意した。これらのニュースはどのスポーツか、プレイ中か非プレイ中かどうかを人手により事前にラベル付けし、ground truth としてある。プレイ中か非プレイ中かは著者の判断によって決定した。そのスポーツの競技フィールド内で、そのスポーツをしている最中ならば基本的にプレイ中としてある。

5.2 評価方法

分類が正しくできたかどうかを測る指標として、本研究では分類率を用いることにする。分類率は、分類率 = (正しく分類されたニュースの数) / (分類するすべてのニュースの数) で表される。

5.3 2段階手法

5.3.1 スポーツ分類実験

スポーツ分類では、codebook の大きさ (単語頻度上位 N) を 1000, 1200, 1400, 1600, 1800, 2000, 2500 の 7 種類用意し、それぞれの値でマルチクラス SVM によるスポーツ分類を行った。また、全テキスト中の上位単語を集めて codebook とするだけでなく、全テキスト中の上位単語と各スポーツの上位単語を合わせて codebook を作る方法も試した。このとき全体の上位は 1000 位までに固定し、各スポーツの上位 100, 200, 300 単語を codebook とした。分類には線形カーネルを用いた SVM を使用し、ソフトマージンの値は 10000 に固定して実験を行った。

5.3.2 イベント分類実験

イベント分類では、特徴を統合する前に各特徴単体でどれだけ精度良くイベント分類ができるかどうかを実験した。各スポーツごとにシングルクラス SVM を用意し、SVM カーネルには χ^2 カーネルを使用した。ソフトマージンの値 C と χ^2 カーネルで用いる値 γ は、設定したいいくつかの値を総当たりで実験することにより最適な値を求めた。設定する値は、C=1~100000 まで 10 倍ごと、 γ =0.1~1.0 まで 0.1 刻みとした。ここで、領域での bag-of-keypoints 特徴の単体実験では、負領域例として mi-SVM で求めた正画像内のネガティブ領域だけでなく、負画像のすべての領域も負例領域として学習した。そして領域の分類結果をもとに、テスト画像内の領域が 1 つでもポジティブ領域ならばそのテスト画像はポジティブであると判定した。

次に、MKL によって 4 種類の画像特徴を最適な重みで統合してイベント分類を行った。ここでは各スポーツごとに MKL-SVM を使うが、MKL ではソフトマージンの値 C は全特徴で共通の値を用いなければならないため、実験では C=10 とした。 χ^2 カーネルで用いる値 γ は各特徴ごとに選択することができるため、特徴単

体のときと同様にいくつかの値を総当たりで実験した。設定する値は $\gamma=0.01, 0.05, 0.1, 0.5, 1.5, 10$ の 7 種類とし、 $7^4 = 2401$ 通りの γ の値の組み合わせで実験を行った。ここでの領域による bag-of-keypoints 特徴では、提案手法通りに負領域を選択した。また、画像全体での bag-of-keypoints 特徴、ガボール特徴、カラーヒストグラムは 1 枚のテスト画像から 1 つだけ特徴ベクトルを抽出するが、領域での bag-of-keypoints 特徴はその画像の領域数だけ特徴ベクトルが抽出される。これらのテストデータを MKL に入力する方法として、本実験では領域での bag-of-keypoints 特徴以外の 3 特徴の特徴ベクトルを、その画像の領域数と同じだけ複製して MKL に入力した。そして MKL の出力が 1 つでもポジティブならば、その画像はポジティブであると判定した。

5.3.3 1 段階手法

1 段階手法では、各クラスの MKL-SVM のカーネルに χ^2 カーネルを使用し、ソフトマージンの値 C は $C=10$ に固定して実験を行った。 χ^2 カーネルで用いる値 γ は、Zhang ら [12] の研究に従って、全学習画像のベクトルの組み合わせの χ^2 距離の平均の逆数を設定した。また MKL では、収束条件パラメータ MKL_C の値を 0 とすると重み付けがスパース（重みが小さいカーネルはすべて重みが 0 になる）になり、大きくすると重みが均等に近づく。本実験ではこの MKL_C の値を 0~100 まで 10 おきに設定し実験を行った。MKL に入力するテストデータについては、2 段階イベント分類のときと同様にデータ入力を行った。テスト画像の分類結果は、テスト画像の全領域を各クラスの MKL-SVM に入力し、（領域数*クラス数）個ある出力値の中で最も値の高いクラスに分類することで得ることとする。

5.4 実験結果

5.4.1 2 段階分類のスポーツ分類結果

各 codebook でスポーツ分類の結果を表 1 に示す。codebook1000 は全体の上位 1000 単語で codebook を作ったことを示し、codebook 1000-300(1600) は全体の上位 1000 単語と各スポーツの上位単語 300 単語を組み合わせで codebook を作ったことを示す。括弧の中身は codebook に用いた合計の単語数である。表 1 から、スポーツ分類ステージにおいて、テキストを使って最も良い精度で 99.33% という高い分類率でスポーツ分類することができたことがわかる。これは bag-of-words 手法がテキスト分類手法として非常に優れている結果であると考えられる。また codebook の作り方を変えていくつか実験を行ったが、分類結果にはほとんど影響しなかった。これは本研究においては codebook サイズが 1000 以上であれば十分な結果が得られるということであろう。おそらく codebook のサイズを 1000 以下に設定すれば、分類精度は低下すると思われる。SVM のカーネルには最も簡単な線形カーネルを用いているため分類精度が悪いことが予測されたが、線形カーネルでも高い分類率を得る

表 1 各 codebook でスポーツ分類の結果

codebook	分類率
1000	99.00%
1200	99.00%
1400	99.17%
1600	99.25%
1800	99.00%
2000	99.17%
2500	99.00%
1000-100(1600)	99.25%
1000-200(2200)	99.25%
1000-300(2800)	99.33%

表 2 codebook1000-300 の分類結果の混合行列

	野球	ゴルフ	F1	サッカー	テニス	相撲
野球	195	0	4	0	0	0
ゴルフ	0	200	0	0	0	0
F1	0	0	200	0	0	0
サッカー	0	0	0	199	0	1
テニス	1	1	0	0	198	0
相撲	0	0	0	0	0	200

ことができた。線形カーネルよりも χ^2 カーネルを用いた方が分類精度はよいといわれているが、そもそも線形カーネルでの精度が良いためこれ以上の精度向上は見込めないと推測し、 χ^2 カーネルでの実験は行っていない。

また、表 2 に、最も分類精度が良かった codebook 1000-300 の結果を混合行列に表したものを示す。スポーツ分類は 6 種類に分類するマルチクラス分類なので、この混合行列を見ればニュースがどのスポーツに分類されているかを確認することができる。ゴルフ、F1、相撲においては他のどのスポーツとも間違えることなく、100% 正しく分類することができたことが確認できる。

表 3 に各特徴を単体で使ったときと MKL により統合したときの、各スポーツのイベント分類結果で最も精度が良かったものを示す。“BoK” は画像全体の bag-of-keypoints、“領域 BoK” は領域での bag-of-keypoints、“色” はカラーヒストグラム特徴をそれぞれ表す。図 5 は表 3 をグラフ化したものである。表 3 と図 5 から、F1 や相撲は各特徴単体でも高い精度で分類できるのに対し、ゴルフやテニスは特徴単体では分類精度が低いことがわかる。また、特徴同士で比較すると、領域特徴が最も分類精度が低く、ガボール特徴が最も精度が高いことがわかる。また、MKL で特徴を重み付き統合することにより、ゴルフ以外で特徴を単体で用いたときより分類精度が向上することが確認された。

図 6 に MKL により各特徴を重み付きで統合したときのイベント分類率と各特徴の重みを示す。F1、テニス、相撲は最高分類精度の各特徴の重みが 2 通りずつ存在した。これらはガボールと色特徴を必ず必要とし、残りの BoK と領域はどちらでもよい。これは BoK と領域 BoK が特徴量として基本的に同一なことから考えると考えられる。ゴルフは分類率の向上が見られず、最大の重みがつ

表 3 各特徴単体と MKL により統合した場合のイベント分類の分類率

	BoK	領域 BoK	ガボール	色	MKL
野球	76.0%	71.5%	80.5%	72.0%	81.5%
ゴルフ	60.5%	58.0%	66.5%	67.5%	67.5%
F1	85.5%	82.5%	84.0%	86.0%	91.0%
サッカー	74.5%	73.5%	73.0%	77.5%	81.0%
テニス	63.0%	57.5%	68.5%	65.0%	71.0%
相撲	87.0%	78.5%	91.0%	91.0%	92.5%
平均	74.42%	70.25%	77.25%	76.50%	80.75%

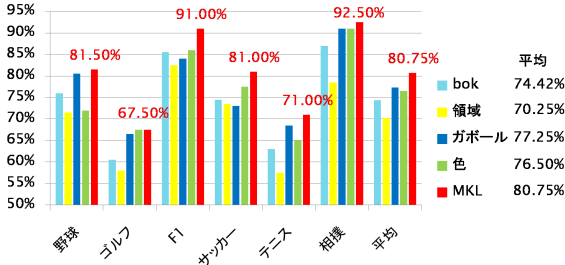


図 5 各特徴単体と MKL により統合した場合のイベント分類の分類率

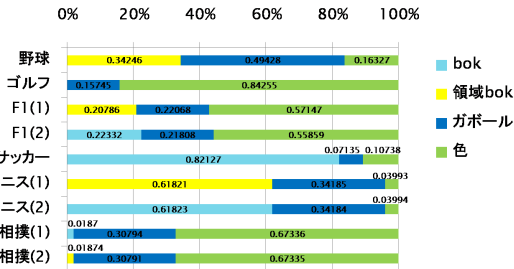


図 6 MKL によるイベント分類の各特徴の重みと分類率

いた色特徴単体での分類率と同じ結果であった。ガボール単体では色特徴単体と同程度の精度であるが、ガボールの重みが約 0.15 と小さいため全体の精度向上にはつながらなかった。サッカーは BoK の重みが最も大きく、他のスポーツとの違いが見て取れる。相撲も 2 通りの重み付け方法があった。しかし、ガボールや色特徴単体ですでに 91% の分類率があるため、特徴統合による精度向上率は低かった。図 7 に分類結果の良かった F1 の分類の様子を、図 8 に分類結果の悪かったゴルフの分類の様子を示す。

5.4.2 1 段階分類のスポーツイベント分類結果

表 4 に 1 段階手法における MKL.C の値とその分類率、また MKL によって得られた重みの例として野球プレイ中の各特徴の重みを示す。MKL.C の値を大きくするにつれて分類率が良くなり、最高で 77.08% の分類率を得ることができた。また MKL.C の値が大きくなるにつれて、各特徴の重みが均一になっている。本論文には他クラスの重みは載せていないが、重みが均一になる現象はどのクラスでも起こっていた。このことから、MKL によって各特徴を重み付きで統合したが、本実験

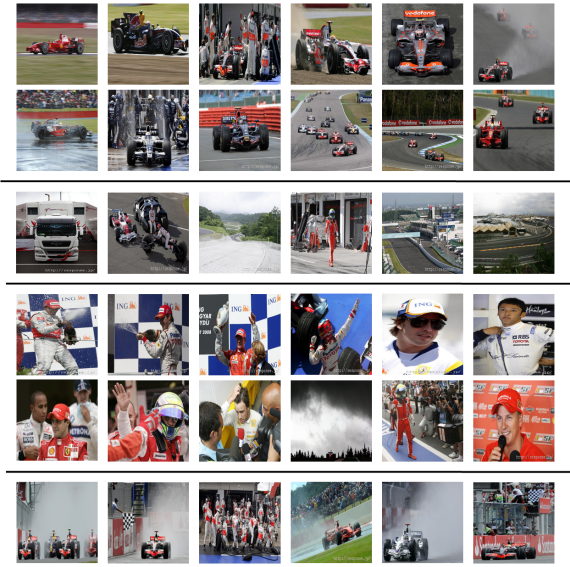


図 7 F1 の分類結果 (1,2 段：プレイ中の成功例, 3 段：プレイ中の失敗例, 4,5 段：非プレイ中の成功例, 6 段：非プレイ中の失敗例)



図 8 ゴルフの分類結果 (1,2 段：プレイ中の成功例, 3,4 段：プレイ中の失敗例, 5,6 段：非プレイ中の成功例, 7,8 段：非プレイ中の失敗例)

においては重みを付けずに各特徴を均一に統合した方が結果がよいということがわかる。表 5 には、各特徴を単体で用いたときと統合したときの 1 段階手法の比較を示す。表 5 を見ると、すべての特徴を統合することで特徴単体よりも分類率が向上していることがわかる。詳しく見ると、BoK 単体の分類率が 72.08% なので BoK の性能を他の特徴が向上させていると考えられる。各クラスごとの 1 段階手法分類結果を表 6 に示す。あるスポーツの

表 4 1 段階手法の MKL_C の値と結果, および得られた各特徴の重み

MKL_C	分類率	BoW	BoK	領域 BoK	ガボール	色
0	68.00%	0.61687	0.00000	0.16136	0.08261	0.13916
10	71.08%	0.56720	0.06240	0.11919	0.11919	0.13202
20	73.17%	0.49127	0.12718	0.12718	0.12718	0.12718
30	75.92%	0.41920	0.14520	0.14520	0.14520	0.14520
40	77.00%	0.35616	0.16096	0.16096	0.16096	0.16096
50	76.75%	0.30392	0.17402	0.17402	0.17402	0.17402
60	77.00%	0.26121	0.18470	0.18470	0.18470	0.18470
70	76.92%	0.22611	0.19347	0.19347	0.19347	0.19347
80	77.08%	0.20000	0.20000	0.20000	0.20000	0.20000
90	77.08%	0.20000	0.20000	0.20000	0.20000	0.20000
100	77.08%	0.20000	0.20000	0.20000	0.20000	0.20000

表 5 特徴単体と特徴統合をした 1 段階手法の結果

特徴	分類率
BoW	10.50%
BoK	72.08%
領域 BoK	67.67%
ガボール	39.58%
色	49.17%
MKL(MKL_C=100)	77.08%

表 6 1 段階手法 (MKL_C=100) の混合行列 (b:野球, g:ゴルフ, f:F1, so: サッカー, t:テニス, su:相撲, p:プレイ, n:非プレイ)

	b_p	b_n	g_p	g_n	f_p	f_n	so_p	so_n	t_p	t_n	su_p	su_n
b_p	68	32	0	0	0	0	0	0	0	0	0	0
b_n	7	93	0	0	0	0	0	0	0	0	0	0
g_p	0	0	56	44	0	0	0	0	0	0	0	0
g_n	0	0	22	78	0	0	0	0	0	0	0	0
f_p	0	0	0	0	81	19	0	0	0	0	0	0
f_n	0	0	0	0	7	93	0	0	0	0	0	0
so_p	0	0	0	0	0	0	85	15	0	0	0	0
so_n	0	0	0	0	0	0	22	78	0	0	0	0
t_p	0	0	0	0	0	0	0	0	17	83	0	0
t_n	0	0	0	0	0	0	0	0	7	93	0	0
su_p	0	0	0	0	0	0	0	0	0	0	87	13
su_n	0	0	0	0	0	0	0	0	0	0	4	96

テスト画像をプレイ中と非プレイ中のイベント分類として間違えることはあっても, 別のスポーツと間違えることはなかった. 2 段階手法のスポーツ分類ではテキストだけで 99.33% の分類率であることを踏まえると, MKL を使ってテキスト特徴に画像特徴を統合することによってスポーツ分類は 100% を達成したといえる. また, 1 段階手法においてもゴルフの分類精度が悪く, 特にゴルフプレイ中の精度が悪いことがわかる.

6. おわりに

本研究では Web 画像ニュースの中のスポーツニュースを, スポーツの種類 (野球, ゴルフ, F1, サッカー, テニス, 相撲) と画像の内容によりプレイ中か非プレイ中かのスポーツイベント分類を行った. 手法として, ニュースのテキスト特徴を使ってスポーツ分類を行ってから画像特徴を統合して使うことによりイベント分類を行う 2 段階手法と, テキスト特徴と画像特徴を同時に統合してスポーツイベント分類を 1 度に行う 1 段階手法を提案した. 特徴統合方法には MKL を採用し, 重み付きで統合した. Yahoo!JAPAN 写真ニュースによる評価実験の結果, 2 段階手法のテキストによるスポーツ分類率は 99.33% を

得た. イベント分類では野球 81.5%, ゴルフ 67.5%, F1 91.0%, サッカー 81.0%, テニス 71.0%, 相撲 92.5% という結果になり, ゴルフ以外で特徴統合により特徴単体よりも高い分類率を得ることができた. 1 段階手法では全体のスポーツイベント分類率 77.08% を得た. ただし学習した各特徴の重みは均一だったため, 結果的にすべての特徴を平等に統合したことになった.

今後さらに分類精度を向上させるためには, 本研究で用いたものとは異なる識別能力を持つ特徴量の追加が考えられる. また, プレイ中によく見られる選手のポーズを認識することによっても精度向上が期待できる. 分類する対象についても, 本研究ではスポーツの種類が 6 種類しかないため, これをさらに増やした場合の実験が必要である. 特に野球とソフトボール, サッカーとラグビーなど画像が似ているスポーツについての分類を検討している.

文 献

- [1] Yahoo!JAPAN 写真ニュース: <http://headlines.yahoo.co.jp/h1?ty=p>.
- [2] Quattoni, A., Collins, M. and Darrell, T.: Learning Visual Representations using Images with Captions, *Proc. of IEEE Computer Vision and Pattern Recognition* (2007).
- [3] Jain, V., Singhal, A. and Luo, J.: Selective Hidden Random Fields: Exploiting Domain-Specific Saliency for Event Classification, *Proc. of IEEE Computer Vision and Pattern Recognition* (2008).
- [4] Li, J. and Li, F.: What, where and who? Classifying events by scene and object recognition, *Proc. of IEEE International Conference on Computer Vision* (2007).
- [5] 茶釜: <http://chasen.naist.jp/hiki/ChaSen/>.
- [6] Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110 (2004).
- [7] Vedaldi, A.: SIFT++, <http://vision.ucla.edu/~vedaldi/code/siftpp/siftpp.html>.
- [8] Deng, Y. and Manjunath, B. S.: Unsupervised Segmentation of Color-Texture Regions in Images and Video, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 8, pp. 800–810 (2001).
- [9] Andrews, S., Tsochantaridis, I. and Hofmann, T.: Support Vector Machines for Multiple-Instance Learning, *Advances in Neural Information Processing Systems*, pp. 577–584 (2003).
- [10] Sonnenburg, S., Rätsch, G., Schäfer, C. and Schölkopf, B.: Large Scale Multiple Kernel Learning, *The Journal of Machine Learning Research*, Vol. 7, pp. 1531–1565 (2006).
- [11] Varma, M. and Ray, D.: Learning The Discriminative Power-Invariance Trade-Off, *Proc. of IEEE International Conference on Computer Vision* (2007).
- [12] Zhang, J., Marszalek, M., Lazebnik, S. and Schmid, C.: Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study, *International Journal of Computer Vision*, Vol. 73, No. 2, pp. 213–238 (2007).