新聞記事における写真と言語表現の対応の学習

山田 剛一　　杉山 一成　　中川 裕志
横浜国立大学 工学部
{aron,ksugi,nakagawa}@naklab.dnj.ynu.ac.jp

1 はじめに

マルチメディアの最大の魅力は、メディア間の意味的な統合により新たな意味の世界が出現しうる点にある。メディアの統合が意味的であるほど、その検索やブラウジングの質も向上する。このような観点から、メディア間の関係を解析し、その意味的融合を図る研究を行っている。

我々が本研究で対象としているのは、写真の入った新聞記事である。新聞は古くから存在する主要なメディアの一つであるが、近年はカラー写真も増え、物理的に見るとテキストと画像が複合したメディアとなっている。この複合したメディアの間の意味的な関係を解析するのが本研究の目的である。

新聞記事では、当然、写真と記事本文の間には意味的な関係があるのだが、本文と写真が直接対応しているのではない。新聞記事においては写真は補足情報としての意味合いが強く、写真に写っているモノ(あるいはコト)と対応する語句は本文のほんの一部分である。この点で、百科事典における図と本文や、写真とキャプションとの関係とは異なっており、まず写真と対応する語句を特定することが必要である。

図 1: 画像と対応するのはテキストの一部

本稿で提案するのは、写真と対応する語句とそうでない語句、それぞれの語句の周辺の言語表現の特徴を学習することにより、これを判別するという手法である。今回は、対象を画像中の人物と記事本文中の人名との関係に絞り、記事中の人名の表す人物が写真に現れているか否かを判断するシステムを構築し、実際の新聞記事を用いて評価した。

2 画像と言語の意味的統合

ここで目指している、画像と言語の意味的統合の簡単な例を図2に示す。双方のメディアの情報を統合することにより、新たな単一の意味を生成している。対応づけの対象は、画像側では写っている物体、テキスト側ではその物体と対応する単語である。本稿で提案するのは、テキスト中の各単語が画像の物体と対応するかどうかを言語表現の学習により判断しようという手法である。

まずそれぞれのメディアで対応づけの対象の候補を抽出し、次にそれぞれの特性に応じた解析を行なってその意味情報を充実させていく。画像であれば対象領域の特徴量の抽出、テキストでは、核となる語の周辺の局所的な意味解析を行なう。画像側、テキスト側とも、可能であればモノの情報だけでなくコトの情報も意味情報として抽出する。意味情報の記述には、言語構造の階層性を考慮して素性構造を用いる。

一般には、画像から抽出される領域も、画像と対応すると判断される語も複数であるため、対応づけは、素性構造のコストつき単一化などを用い、最も可能性の高い対応づけを行なう。

なお、本研究では顔画像と人名の対応づけを最初の課題として設定している。

3 言語表現の学習手法

新聞記事の本文には多くの人名が現れる。その人名の表す人物は、写真に写っている人物であることもあれば、そうでないこともある。しかし、写真に写っている人物は記事の内容において重要な人物であることが多く、また本文中で明示的に写真を参照する場合も

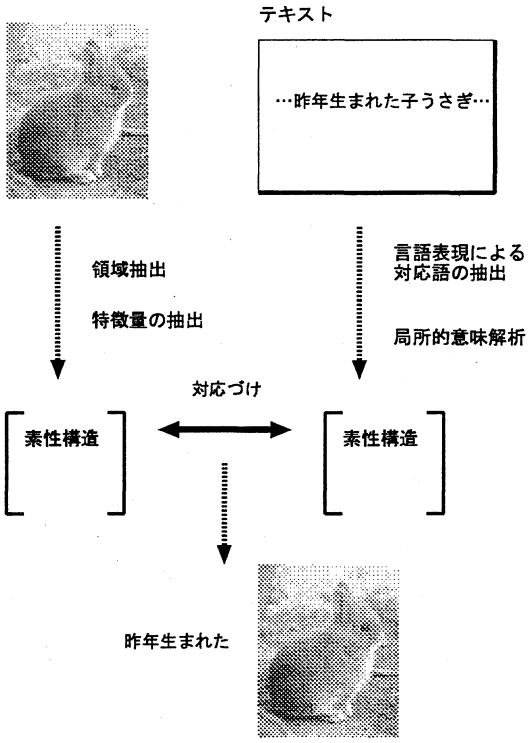

図 2: 意味構造の生成過程

あるので、人名が写真に写っている人物を指している場合には、その周囲の言語表現に何か特徴があると考えられる。本手法ではこの考えに基づき、人名の前後の言語表現、およびそれが写真中の人物であるか否かを 1 つの事例とした学習を行なう。

3.1 学習アルゴリズム

学習アルゴリズムにはさまざまなタイプのものがあるが、本研究では、分類モデルを決定木およびプロダクションルールで表現するタイプの学習プログラムである C4.5(release 8)[1] を用いている。本手法ではクラスが離散であらかじめ定義されており、決定木を生成するタイプの学習アルゴリズムに適していること、また、C4.5 は他に比べ属性間の従属性にシビアでなく属性を追加しやすいことが、C4.5 を選択した理由である。なお、C4.5 の欠点の一つに属性数が固定でなければならないという制約があるが、本手法では固定された範囲内の言語表現を学習するため障害にはならない。

3.2 属性として用いる言語表現の特徴

言語表現の特徴を学習する際に重要なのは、どこまで処理したものを学習させるかである。例えば、文字の列として学習させるのでは特徴がなかなかつかめないし、真面目に構文解析を行うと曖昧性の扱いに苦慮することになる。また、今回は人名に限定しているものの、大量の新聞記事を処理する必要があるためコストの高い処理は実用上問題がある。

そこで本研究では、記事本文を形態素解析し、形態素の列として、その見出し語、品詞などを学習の際の属性として用いることにした。ただし、学習結果に構文的な情報が現れやすくするため、複合名詞[1]はあらかじめ結合し一つの形態素として扱った。

具体的な属性としては、以下のものを用意した。

着目する語の前後の語の情報

着目する語 (人名) の周辺の言語表現の特徴をつかむため、着目する語の前後 n 形態素の見出し語 (基本形) と品詞情報を、相対位置の情報も含めて使用することにした。

品詞情報については、使用している日本語形態素解析システム JUMAN version 3.4[2] の品詞体系に基づき、品詞と品詞細分類の 2 つを属性として利用した。ただし、用言については品詞だけとした。よって、見出し語と合わせて属性の数は計 3 つ (用言の場合は 2 つ) である。

なお、着目する語 (人名) 自体の情報は用いていない。これは、人名の周辺の言語表現の特徴をつかむのが目的であるからである。実際には、着目している語の見出し語、つまり人名は強力な情報であり、例えば画像に登場しやすい首相や横綱の名前は、当然画像の人物と対応しやすい。ただし、一般に話題の人物の移りかわりは激しいので、学習してもその効果は持続しないと考えられる。

着目する語の出現位置の情報

着目する語の出現位置は、文書の構造を近似的に反映するものとして重要である。まず、タイトル内の語であれば当然重要であり、画像に現われる可能性は高いと予測される。また、記事本文中の語であっても、特に新聞記事では大事なことを最初の文に書くといった特徴があるため、語の出現位置は重要な情報である。そこで、着目する語の位置がタイトル内か、ある

[1] 人名は他の名詞と複合しないものとした。

いは本文なら、何文目、何段落目であるかを属性として用意した。

語と写真の物理的な距離は、対象が書籍の場合には有効な情報として機能すると考えられるが、新聞では写真の位置が記事のレイアウトにより定まるうえ、写真は記事全体の補足情報という位置づけのため、記事の参照部分との位置関係はあまり考慮されていないと思われる。よって本手法では物理的な位置情報は使用していない

表 1: 属性とその数

属性		数
周辺表現 (前後 n 形態素)	見出し語	$3 \times 2n$
	品詞	
	品詞細分類	
出現位置	何文目か	1
	何段落内か	1

4 評価

本手法を評価するため実際の新聞記事を解析し、その精度を確認した。

4.1 データ

今回の評価に用いたデータは、毎日新聞 AULOS の写真ニュースである。これは毎日新聞社が Web で公開しているホームページ「AULOS」[3] の一部で、毎日 10 前後の記事が画像つきで登録されており、過去の記事も参照できるというものである。写真ニュースの性格上、内容は幅広いが、長い記事は少なく、リードは存在しない。また、写真にはキャプションはついていない。通常、1 つの記事に画像は 1 つであるが、まれに複数の画像を持つものがある。

今回の評価では、1997 年の 5 月と 6 月の記事のうち、画像を複数持たないものを利用した。なお、本手法では属性として段落情報を用いるが、これには、HTML 形式で書かれた記事の段落タグによる区切りをそのまま利用している。

4.2 評価方法

本手法では記事中の人名が 1 事例となり、クラスは、人名と対応する人物が画像に存在する / 存在しない、の 2 クラスである。

1997 年の 5 月と 6 月の総記事数は 508 記事であるが、画像中の人物の人名を含む記事の数は 227 記事である (表 2)。これを学習およびテストに利用した。

表 2: 画像中の人物の人名を含む記事の数

	総記事数	画像中の人物の人名を含む記事の数
1997 年 5 月	232	123
1997 年 6 月	276	104

記事集合を 5 月と 6 月の 2 つのブロックに分割し、ブロック数 2 の交差検定を行った。実際の事例は記事単位ではなく人名ごとであり、学習 / テストに用いた事例数は表 3のとおりである。各ブロックの事例数はほぼ均等になっている。なお、学習の際には、各クラスに属する事例数を等しくしている。これは、C4.5 利用上の制約による。

なお、形態素解析結果の誤りは修正していないので、JUMAN の辞書に人名と登録してあるもののみから事例が構成されている。ただし、カタカナの人名は JUMAN の辞書に登録されている数が少ないため、カタカナすべてを人名候補とみなしてテストデータとする実験も行った。

表 3: 用いた事例数

テスト 1	期間	画像と対応	非対応
トレーニング	1997 年 5 月	144	144
テスト (カタカナ除外)	1997 年 6 月	140 (100)	795 (299)

テスト 2	期間	画像と対応	非対応
トレーニング	1997 年 6 月	140	140
テスト (カタカナ除外)	1997 年 5 月	144 (122)	792 (318)

着目する語の前後の形態素数 n は 5 とした。属性数は合計 32 である。

4.3 評価結果

まず、構築された決定木およびプロダクションルールのエラー率を示す (表 4)。決定木よりもルールの方がエラー率が低いことがわかる。

表 4: 決定木 / ルールのエラー率

	カタカナ含む		カタカナ除外	
	決定木	ルール	決定木	ルール
テスト 1	0.32	0.21	0.42	0.34
テスト 2	0.30	0.27	0.41	0.24
平均	0.31	0.24	0.42	0.29

画像と対応する語であるというクラスの再現率 / 適合率を求めた結果が、表 5, 6である。ただし、どちらも枝刈りの際の信頼度 CF の値を 35% としている。画像と対応する語であるというクラスに属する事例の比率が小さいため、エラー率から想像されるよりも値が低くなっている。

この結果を見ると、決定木を用いた方が再現率が高く、ルールを用いた方が若干適合率が高いという傾向がわかる。対応づけの候補を抽出するという目的からすると再現率が高いことが望ましく、決定木の方が適しているといえるが、いずれの手法でも適合率が低く、まだ改善の必要性があるといえる。なお、カタカナを除外すると再現率 / 適合率とも向上していることから、カタカナが人名か否かまでを判断することは困難であることがわかる。

表 5: 決定木の場合の適合率 / 再現率

	カタカナ含む		カタカナ除外	
	再現率	適合率	再現率	適合率
テスト 1	0.70	0.24	0.74	0.29
テスト 2	0.75	0.27	0.83	0.41
平均	0.73	0.25	0.78	0.35

表 6: ルールの場合の適合率 / 再現率

	カタカナ含む		カタカナ除外	
	再現率	適合率	再現率	適合率
テスト 1	0.46	0.35	0.60	0.39
テスト 2	0.56	0.29	0.61	0.56
平均	0.51	0.32	0.60	0.48

生成された決定木を見ると、見出し語で枝分かれしていることが多く基本的に語彙主導で判断されていることがわかるが、構文的な要素も若干含まれている。例えば、動詞が前からかかっている場合 (「…会議に出席した～外相は…」)、2つ後ろの形態素が読点である場合 (「…の～選手、脱税で逮捕」) などでは、画像と対応している場合が多いといったことが見てとれた。

属性として有効であった周辺表現はほぼ前後 3 形態素内に存在するものであった。また、語の位置情報はあまり有効ではなかった。

5 おわりに

本稿では、学習により表層の簡単な情報から画像と対応する人名か否かを判断する手法を提案した。今回の手法で属性として用いた情報は形態素解析結果と語の位置情報だけであったが、語の出現形態を表現するより多くの属性を用いることにより、精度向上が図れると考えている。カタカナの人名辞書が存在しない問題に対しては、本手法とほぼ同様の手法による人名周辺の言語表現の学習を行なうことによって対処する実験を進めている。

また、今回は対象を人物に絞ったが、オブジェクト一般やイベントに対しても実験を行っていきたいと考えている。

本研究の目的は画像と言語の双方から意味を構築することであり、画像解析、および画像と言語の双方から得られる情報を統合するモデルの設計を並行して行なっている。今後の展開としては、統合された意味の世界における検索 / ブラウジングのシステムを構築し、ユーザにメディアの境界を意識させない意味の世界を提供したいと考えている。

謝辞 充実した写真ニュースを Web 上で発信している毎日新聞社に敬意を表すると同時に感謝いたします。

参考文献

[1] J.Ross Quinlan. *C4.5: Programs for Machine Learning.* Morgan Kaufmann Publishers, Inc., 1993. (古川康一 監訳. AI によるデータ解析. トッパン, 1995).

[2] 黒橋禎夫, 長尾真. 日本語形態素解析システム JUMAN version 3.4. 京都大学大学院工学研究科, Oct 1997.

[3] 毎日新聞社. 毎日新聞 AULOS 写真ニュース. http://aulos.mainichi.co.jp/, 1997.

[4] 渡辺靖彦, 長尾真. 画像の内容を説明するテキストを利用した画像解析. 人工知能学会誌, Vol. 13, No. 1, pp. 66–74, Jan 1998.

[5] 佐藤真一, 中村裕一, 金出武雄. Name-It: 動画像処理と自然言語処理の統合による映像内容アクセス手法. 第 3 回知能情報メディアシンポジウム論文集, pp. 187–194. 知能情報メディア時限研究専門委員会 電子情報通信学会, Dec 1997.