新聞記事における写真と言語表現の対応の学習

4Q－8

山田 剛一　杉山 一成　中川 裕志

横浜国立大学 工学部

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、写真に写っているもの(こと)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。

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、これを判断するという手法である。今回行った実験では、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システムに判断させた。その結果、約75%の確率で正しく判断することができた。

2 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。本研究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つの事例とする学習データを用意し、分類モデルを決定木で表現するタイプの学習プログラムであるC4.5 [1]を用いて学習することにした。

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掴めないし、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、以下のものを用意した。

- 着目する語(人名)の前後n形態素の見出し語/品詞/品詞細分類[2]
- 着目する語の位置がタイトル内か、あるいは本文なら、何文目、何段落目であるか

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[3]。

また、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、語の出現位置も属性として用意した。

Learning Relations Between Expressions And Pictures In Newspaper.
Koichi Yamada, Kazunari Sugiyama and Hiroshi Nakagawa.
Division of Electrical and Computer Engineering,
Faculty of Engineering, Yokohama National University

[1]人名は他の名詞と複合しないものとした。

[2]この品詞体系は日本語形態素解析システム JUMAN version 3.4[4] のものである。

[3]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。

3 評価方法と評価結果

本手法の評価のため、毎日新聞AULOSの写真ニュース[5]を用いた。これは毎日新聞社がWebで公開しているもので、すべて記事に画像がついていて、毎日10記事前後が新たに登録されている。今回の評価では、1997年5月1日から6月7日までの記事を用いた[4]。総文数は1,439であり、この中で写真中の人物と対応する語(人名)は171個現れている[5]。この語の数が均等になるように記事を2つのブロックに分け、ブロック数2の交差検定を行った。

表1: 各ブロックの大きさ

	期間	写真と対応する語の数
ブロック1	5/1-5/24	86
ブロック2	5/25-6/7	85

言語表現の多様さを考えるとデータの数が少ないため、C4.5で木を生成する際に、すべての属性によるテストが、少なくとも2つの出力値に対して10個の事例を持たなくてはならないように指定し、ノイズの影響を押さえるようにした。枝刈りの際の信頼度CFの値はデフォルトの25%を用いた。

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。

表2: テスト結果

データ		エラー率	
学習データ	テストデータ	枝刈り前	枝刈り後
ブロック1	ブロック2	25.7%	20.4%
ブロック2	ブロック1	29.2%	29.2%

交差検定の結果は表2のとおりであり、約75%の精度で推定できることがわかった。

生成された決定木は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。また、語の位置情報はあまり有効ではなかった。

[4]今回は画像が複数ある記事は対象外とした。

[5]これはJUMANの辞書にある人名、およびカタカナの人名だけの数字である。

4 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。その評価実験では、約75%の推定精度であった。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。

また、今回の実験で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。また、これを元にした検索/ブラウジングシステムを構築し、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。

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] 渡辺靖彦, 長尾真. 図鑑の解説文から内容抽出を行うための専門知識の構築. 人工知能学会誌, Vol. 11, No. 3, pp. 451–460, May 1996.

[3] 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.

[4] 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.

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