

テレビの情報番組における指示詞の照応先同定

西田 奉示[†] 徳永 健伸^{††} 山田 一郎[‡]

[†] 東京工業大学 工学部 情報工学科

^{††} 東京工業大学 大学院情報理工学研究科 計算工学専攻

{nishita,take}@cl.cs.titech.ac.jp

[‡] 日本放送協会放送技術研究所

1 はじめに

近年、記憶装置や記憶媒体の大容量化、多様化に伴い、音声や映像といった大容量のデータを扱う事が容易になった。また、デジタル放送やネットワーク配信といった放送される情報の大量化、多様化により、視聴者は視聴したい情報、番組の取舍選択が必要になってきており、放送、配信済みの映像資源の管理、再利用等の点においても、映像、番組の自動要約技術が要求されている。さらに、1997年の放送法改正により、字幕付き放送を増やすことに関する努力規定が定められ、クローズドキャプション付きのテレビ番組等が増えている。

このような背景をふまえ、我々は、クローズドキャプションを利用して番組の要約を生成することを目指している。本稿ではそのための基礎技術としてクローズドキャプション中の指示詞がテキスト中の先行詞を指しているか、映像を指しているかを判定する手法を提案する（番組の要約を作る際、テキスト中の指示詞が映像を指している場合は要約にもその映像を付与する必要がある）まず、NHK 番組「ためしてガッテン」のクローズドキャプションから、コソア系の指示詞を収集し、その出現傾向を分析した。また、その分析に基づき指示対象の同定、特に外界照応を識別するための素性を抽出し、機械学習により学習モデルを作った。

2 関連研究

映像に対して要約、構造解析等を行っている先行研究としては、ニュース番組を対象とした伊藤ら [4] の研究や料理番組を対象とした柴田ら [5] の研究がある。本稿では情報番組の要約を目的とし、クロー

ズドキャプション中で映像を参照している箇所を抽出することに焦点を置いており、これらの研究は直接の参考にはなりづらい。

また、指示詞の照応解析に関しての先行研究は多くあるが、コーパスとしてテレビ番組のクローズドキャプションを使っているもの、外界照応を積極的に識別することを目的としたものはほとんどない。

3 照応タグ付け

3.1 分析対象

人手で5番組分¹のクローズドキャプション中の指示詞に対して、照応対象の分類をしたタグをつけた。テキストを日本語形態素解析プログラム茶筌 [10] を用いて形態素解析し、品詞タグを手がかりに、指示詞を抽出した。対象となる指示詞は以下の通りである。

コ系: これ、この、ここ、こう、こんな、こちら、こっち

ソ系: それ、その、そこ、そう、そんな

ア系: あれ、あの、ああ、あんな、あっち

3.2 タグの種類

上記のコソア系指示詞に対して、タグ付けを行い、コーパスを作った。タグの種類と例は以下の通りである。

o: 文脈上の名詞句を指す。

当て布をして<o その>上から …。

¹NHK. 2004. 2/25, 6/16, 6/23, 7/14, 8/4 放送分

e: 文脈上のイベントあるいは説明を指す。
まわりは染料が濃くなって残ります。<e それ>が輪染みです。

v: 映像を指す。
<v これ>は先ほどの革の構造の写真です。

o+v: o と v の両方を指す。
尿酸を溶かした水溶液です。<o+v ここ>に酸性の水溶液を加えます。

e+v: e と v の両方を指す。
茶色い毛が生えています。さて<e+v この>茶豆 3, 4 倍のお値段がいたします。

x: 映像以外の状況を指す
さて、<x ここ>で問題。

n: 指示対象なし
<n そんな>に酸っぱくないですね。

表 1: タグ付け結果

個数	文内	映像	文+映	他	合計
これ	21	53	33	3	110
この	17	49	17	12	95
その	39	3	8	2	52
それ	39	1	3	5	48
ここ	0	20	8	18	46
こちら	0	30	5	0	35
こう	6	17	2	5	30
そう	16	0	0	1	17
そこ	9	1	3	2	15
あの	1	6	2	1	10

表 2: 指示詞内訳

	コ系	ソ系	ア系
内訳	67 %	30 %	3 %

3.3 コーパス分析

上記のタグを以下のように「文内」「映像」「文+映」「他」の4種に分類した。各分類の指示詞の出現頻度の高いものを表1に示す。また、指示詞のコソア別の分布を表2に示す。

文内: o, e

映像: v

文+映: o+v, e+v

他: x, n

その結果、合計 516 個の指示詞が抽出され、その内訳を見ると、コ系、ソ系の出現頻度が高く、ア系の出現頻度が低いことが分かった。また、ソ系は 80 % が文内照応で外界照応しているものは 5 % 程度であり、コ系が外界照応をしているものは 50 % 程度だった。そこでコ系に対して外界照応を識別する学習モデルを作ることを試みた。

4 決定木による照応解析手法

コ系のうち特に出現頻度の高い「これ」「この」に対して、照応対象を識別する機械学習モデルを作った。照応対象の分類は、映像だけを指す外界照応か、そうでないか、の二値とした。上記のタグにおける、v とそれ以外に相当する。それぞれの素性は以下の通りである。

「これ」

- 発話者 (スタジオ, VTR)
- 後接助詞等 (は, が, を, なし, その他)
- 文の述語タイプ (名詞 + 断定, 形容詞, 動詞, 他, ご覧+, ご覧-)

文の述語タイプは

- 「これ」が主語の場合は、名詞 + 断定, 形容詞, 動詞, 他の値を取る。
- 「これ」が主語でない場合は、ご覧+, ご覧- の値を取る。ご覧+, ご覧- とは、述語に「ご覧」という語を含むかどうかである。

「この」

- 発話者 (スタジオ, VTR)
 - 後接名詞が5分前までに発言されているか (真, 偽)
 - 後接名詞の分類語彙表 [9] における分類 (1.10, 1.11, 1.12, 1.13, 1.14, 1.15, 1.16, 1.17, 1.18, 1.19, 1.2, 1.30, 1.31, 1.32, 1.33, 1.37, 1.38, 1.4, 1.50, 1.51, 1.52, 1.53, 1.54, 1.55, 1.56, 1.57)
- 分類語彙表による分類について5つに分けて例を示す.

1.1: 抽象物・関係

結果 (1.11), 効果 (1.12), 時 (1.16)

1.2: 人間活動の主体

人 (1.20), 女性 (1.20), 村 (1.25)

1.3: 人間活動, 精神及び行為

実験 (1.30), グラフ (1.31), 仕事 (1.33)

1.4: 生産物及び用具

革製品 (1.42), 水 (1.43), 機械 (1.46)

1.5: 自然物及び自然現象

物質 (1.51), 細菌 (1.54), 病 (1.57)

5 評価実験

これらの素性を使い, 15 番組分²(「これ」388 「この」341) の指示詞を用い, Weka[1][2] の J4.8 により決定木を作り, 10 分割交差検定を行った. この結果を表 3 に示す.

また, 比較実験として, 被験者二人にクローズドキャプションのみを見せ, 5 番組分³(「これ」110, 「この」95) の指示詞に対して人手によるタグ付け実験を行った. この結果を表 4 に示す.

評価に使う正解率, 精度, 再現率, F 尺度は以下の通りである. 外界照応の指示詞を判別することを目的としているため, v と判定できる精度に注目する.

$$\text{正解率} = \frac{\text{正しく同定できた個数}}{\text{指示詞の総数}}$$

²NHK. 2004. 2/25, 3/3, 3/10, 3/17, 4/7, 4/21, 4/28, 5/12, 5/19, 6/2, 6/9, 6/16, 6/23, 7/14, 8/4 放送分

³NHK. 2004. 2/25, 6/16, 6/23, 7/14, 8/4 放送分

表 3: 学習モデルによる実験結果

	正解率	精度	再現率	F 尺度
これ	58.2 %	56.6 %	59.0 %	57.8
この	61.9 %	63.5 %	66.3 %	64.9

表 4: 被験者による実験結果

	正解率	精度	再現率	F 尺度
これ	71.8 %	80.5 %	59.6 %	68.5
この	76.3 %	76.9 %	61.2 %	68.2

$$\text{精度} = \frac{\text{正しく } v \text{ と同定できた個数}}{\text{システムが } v \text{ と判断した個数}}$$

$$\text{再現率} = \frac{\text{正しく } v \text{ と同定できた個数}}{\text{照応対象が } v \text{ である指示詞の総数}}$$

$$F \text{ 尺度} = \frac{2 \times \text{精度} \times \text{再現率}}{\text{精度} + \text{再現率}}$$

6 考察

被験者の正解率や精度に比べ再現率が低いことからわかるように, 人間にとってもクローズドキャプションのみから指示詞が文内照応であることを識別することに比べ, 指示詞が外界照応であることを識別することは難しい問題であることがわかる. 人間は会話が行われている状況が見えない時には, 会話中の指示詞が物を指し示していることに気がつきづらいということがわかった. 人間と学習モデルの結果を比較すると, 精度に関しては学習モデルに改善の余地があるものの, 再現率に関してはこれ以上の改善は困難であることがわかる.

「この」は後接名詞の分類語彙表による分類が主に効いており, 「これ」に比べよく学習できている. 「これ」は「この」における語彙分類のように特に効いているというような素性を見つけることができなかった. 「この」は後接名詞によって指示対象がテキストからもわかることがあるが, 「これ」は表層的な情報だけでは何を指しているのかはわからない. 「これ」に関しては述語の情報が重要と思われる. 述語

の形容詞や動詞の格フレームから「これ」の情報を求め、それを素性として利用することで性能を改善できると考えられる。

さらに性能を改善するためには、画像認識により映像を利用する事が考えられる。例えば、「この」の後接名詞が具体名詞であると語彙分類された場合や、「これ」が主語であった場合、人間が映っていないシーンの有無等が素性として利用できると考えられる。

7 まとめ

NHK 番組「ためしてガッテン」のクローズドキャプションに対してコソア系の指示詞に指示対象をタグ付けしたコーパスを作り、その出現傾向を分析した。その結果、コ系、ソ系の出現頻度が高く、ア系の出現頻度は低いことがわかった。またソ系が外界照応をしているものは5%程度であり、コ系が外界照応をしているものは50%程度だった。コ系のうち特に出現頻度の高い「これ」、「この」に対して、解析し外界照応を識別するための機械学習モデルを作った。学習モデルは再現率に関しては人間とほぼ同じ程度の性能を得られたが、精度についてはさらに改善の余地がある。

本研究ではクローズドキャプションの表層的な情報を用いて照応解析を行った。さらに性能を改善するためには、既存の照応解析の手法のように先行詞候補との同定や格フレームを利用したり、画像認識により映像を利用することが考えられる。

参考文献

- [1] University of Waikato. Weka. Open source data mining software written in Java. <http://www.cs.waikato.ac.nz/ml/weka/>
- [2] Ian H. Witten, Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniques (Second Edition). Morgan Kaufmann, 2005.
- [3] Stuart Russel, Peter Norvig. エージェントアプローチ 人工知能. 古川康一監訳, 共立出版, 1997.
- [4] 伊藤一成, 酒井康旭, 齋藤博昭. 音声と映像の一貫性を考慮したニュース動画の要約. 日本データベース学会 Letters Vol.3, No.1.
- [5] 柴田知秀, 立木将人, 河原大輔, 岡本雅史, 黒橋禎夫, 西田豊明. 言語情報と映像情報の統合による教示発話の構造解析. 言語処理学会 第10回年次大会, pp.532-535 (2004.3).
- [6] 飯田龍, 乾健太郎, 松本裕治. 照応性判定を含む名詞句照応解析の実験と分析. NL160-15, 2005.
- [7] 徳永健伸. 情報検索と言語処理. 東京大学出版会, 1999.
- [8] 神尾昭雄. 情報のなわ張り理論, 言語の機能的分析. 大修館書店, 1990.
- [9] 国立国語研究所. 分類語彙表. 秀英出版, 1993.
- [10] 松本 裕治, 北内 啓, 山下 達雄, 平野 善隆, 松田 寛, 高岡一馬, 浅原 正幸 日本語形態素解析プログラム茶釜 version 2.3.3 奈良先端科学技術大学院大学情報科学 研究科自然言語処理学講座 松本研究室. 2003.