

## 8M-4 映像説明テキストに含まれる特定表現の抽出によるインデックス作成手法

宮原 伸二 内野 一 古瀬 蔵 外村 佳伸

NTT サイバーソリューション研究所

### 1. はじめに

近年、記憶媒体の大容量化やネットワークの高速化により、大容量のマルチメディアコンテンツが容易に扱えるようになってきた。それに伴い、それらのコンテンツが広く普及し、様々な分野で利用されている。そのため現在では、一般ユーザが手軽に利用できるようにするため、コンテンツのより多様な提供方法が求められている。

その手段として、マルチメディアコンテンツにコンテンツ情報や他のコンテンツへのリンク情報などを付加することにより、ユーザが多様に利用する方法が考えられている。例えば、映像内の人物や建物にリンク情報を付加し、WWW ページなど他の情報と統合させることにより、ユーザへの提供の幅を広げている。また、ユーザによるコンテンツ操作として映像検索が挙げられ、様々な研究が行われている。例えば、スポーツ映像に対し、映像内の対象物の動きなどを解析して検索する手法[1]、ドラマ映像に付加されたテキスト情報を解析し、それらを基に映像を検索する手法[2]が提案されている。しかし、これらの手法では、いずれも映像内容の分野を事前に限定しており、幅広い分野の映像内容への対応が困難である。

本稿では、幅広い分野の映像検索を目的とした、映像の内容を説明するテキスト情報に含まれる単語の属性をインデックスに用いる手法について述べる。また、本手法をドラマ映像に適用し、特定の語彙グループに含まれる属性をキーとする映像検索の例を示す。

### 2. 単語の属性を用いたインデックス作成による映像検索

映像内容を説明するテキスト情報を解析し、それらを含む単語の属性をシソーラスを用いて調べ、その属

性をインデックスとして作成する。日本語単語の属性を示す意味カテゴリを分類・階層化したシソーラスとして、日本語語彙大系[3]がある。日本語語彙大系は、単語を約 3000 の意味属性で分類している。この日本語語彙大系を用いて、映像を説明するテキスト情報内の単語の属性を調べることにより、様々な映像検索を行う。例えば、鳥カテゴリ、獣カテゴリを属性にもつ単語からは、映像に動物が映っていると予想できる。また、悲喜カテゴリ、怒りカテゴリを属性にもつ単語からは、人物が感情を表現している映像内容だと予想できる。

このように、テキスト情報内の単語の属性を用いた映像検索が可能であり、また、ある特定の属性をもつ単語を含む複数の映像を抽出することで、ユーザの様々な意向に応じた映像を提供できると考えられる。例えば、ドラマにおける登場人物の心理が変化する映像は重要であるため、それらの映像を抽出することには大きな意味がある。そこで、心理が変化する映像にはテキスト情報に感情カテゴリを属性にもつ単語が含まれると考え、それらの単語を含む映像を抽出する。

### 3. インデックス作成方法

図1のように複数の映像に分割された映像について考える。テキスト情報はその映像の先頭に付加されているものとする。まず始めに、インデックスを作成するため、すべてのテキスト情報に対し、ALT-J/E[4]の形態素解析を用いて単語単位に分割する。形態素解析で得られたすべての単語を日本語語彙大系の木構造意味カテゴリを使って、単語のもつ属性を調べ、それを映像のインデックスとする。

### 4. 適用例

ここでは、2つのドラマ映像 A, B に対し、インデックス作成を試みる。映像を説明するテキスト情報として台本のト書きを用いている。台本のト書きは、映像内の登場人物に関する行動への指示が記されており、

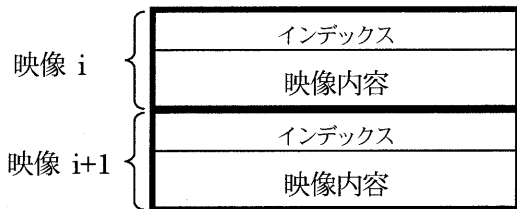


図 1: 映像分割例

映像内容を説明するテキスト情報として用いるのに適当と考えられる。また、映像へのテキスト情報の付加方法として、登場人物がト書きにある行動を起こした時間を区切りに映像を分割している。3 節の方法に基づき映像に対しインデックスを付与しておく。

ドラマ映像 A の内容は、男性 1 と女性の出会いがあるが、女性は過去の記憶に苦しみ自殺を図るストーリーである。ドラマ映像 B の内容は、女性の自殺を阻止したことを機会に男性 1、その友人男性 2 と女性が仲良くなり、自殺の理由を聞くことと、男性 2 と女性が親密になるストーリーである。この 2 つのドラマ映像から、感情カテゴリノード以下に含まれる属性(1253-1342)をインデックスとする映像を抽出した結果、それぞれの映像から主として下記のようなインデックスをもつ映像が得られた。

#### (ドラマ映像 A)

[映像 A1 インデックス] 1342(上機嫌)

内容 男性 1 と女性の出会いで盛り上がり、男性 1 がにこやかに話している。

[映像 A2 インデックス] 1342(寂しさ)

内容 女性が過去を思い出し、部屋で 1 人膝を抱えている。

[映像 A3 インデックス] 1263(怒り)

内容 男性 1 と男性 2 が喧嘩している

[映像 A4 インデックス] 1259(苦しみ)

内容 女性が自殺未遂を計り、床に倒れている。

#### (ドラマ映像 B)

[映像 B1 インデックス] 1263(嬉し)

内容 女性を部屋に招き入れ、男性 1 が食事を作っている。

[映像 B2 インデックス] 1265(驚愕)

内容 女性から自殺の理由を聞き、男性 1 が驚く。

[映像 B3 インデックス] 1265(呆返り)

内容 男性 1 は男性 2 に女性の気をひかれ呆然としている。

[映像 B4 インデックス] 1265(呆れ)

内容 男性 1 は女性の男性 2 への好意に呆れている。

上記において、数字はインデックスとして用いたカテゴリ番号であり、括弧内の単語は感情カテゴリノード以下の属性をもつト書き中の単語を示している。この結果から、第 1 回、第 2 回の両方のドラマ映像から、

ストーリー上重要な映像が抽出されていることがわかる。また、この他にも感情カテゴリをもつ単語を含む映像が抽出されたが、多くの映像が上記の映像の時間的前後に位置している。

## 5. 考察

4 節の例では、感情カテゴリを属性にもつ単語を含む映像を抽出した。また、その映像がストーリー上で重要であることがわかった。ここで、抽出された映像とドラマストーリーの関係について考える。抽出した映像内における登場人物の様子は、喜び、悲しみなど感情の変化を演技で表しているものが多い。ドラマでは、登場人物の人間関係や心理の変化を中心にストーリーが構成されているものが多い。そのため、一般に登場人物の感情の変化が起こっている場面が、ストーリー上で重要となっている。このことを考慮すると、ドラマにおけるト書きに感情表現が含まれる映像は、ドラマストーリーの概要を短時間で理解できる映像として利用できると考えられる。その他の適用例として、酒の属性を検索対象とすれば、バーやレストランの映像を中心として抽出が可能である。

## 6. まとめ

本研究では、映像に付加された説明テキスト情報を形態素解析にかけ、得られた単語の属する語彙グループを解析しインデックスを作成した。実験では、ドラマ映像を対象にし、映像を説明するテキスト情報としてドラマ台本のト書きを利用した。本手法は、感情表現を表すグループに属する単語を含むシーンを抽出することで、それらの抽出したシーンがドラマ要約に利用できることがわかった。今後の課題として、より様々な分野の映像に対し、要約映像を生成できるような抽出方法の検討が挙げられる。

## 参考文献

- [1]越後富夫: 視覚的な手がかりによる動画検索, IPSJ Magazine, Vol. 41, No. 4, pp. 351-356 (Apr. 2000).
- [2]張文利, 柳沼良和, 坂内正夫: 映像と関連テキストのマッチングによるドラマ映像記述システム, 電子通信学会, TECHNICAL REPORT OF IEICE, NLC99-72, PRMU99-255, pp. 69-74 (Mar. 2000).
- [3]池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦(編): 日本語語彙大系, 岩波書店 (Sep. 1997).
- [4]S.Ikehara, S.Shirai, A.Yokoo, and H.Nkaiwa: Toward an MT system without pre-editing -Effects of new method in ALT-J/E -, In Proc. of MT SUMMIT '91, pp. 101-106, (1991).