E-059

コミュテンツ (TM) 記事からのシーンメタデータ抽出実験と評価

山崎 智弘† 筒井 秀樹† 浦田 耕二† 福井 美佳†

The experiment and the evaluation of video scenes' metadata extraction from articles published by using Commutents(TM)

YAMASAKI Tomohiro TSUTSUI Hideki URATA Koji FUKUI Mika

1. はじめに

近年のネットワーク帯域の拡大に伴い、YouTube のような個人が簡単に動画を投稿・共有することができるサイトが増加している。こうした投稿型動画共有サイトは好きなときに自由に動画を見ることができるため非常に多くの人気を集めており、動画の投稿数も飛躍的に増加している。

これに伴い動画自体あるいは動画内のシーンの検索ニーズは高まっている。動画を見ずに特定のシーンを検索するためには、シーンごとに誰が何をどうしたのかを表すメタデータが必要である。またネタばれや伏線を検索するためにはシーン同士がどういう関連を持っているかを表すメタデータが必要である。しかしながら投稿動画のような制作者が不明確な動画に限らず DVD のように制作者が明確な動画であってもこのようなメタデータはほとんど存在しない。そのため顔認識やテロップ認識、字幕情報の解析などを利用してシーンごとのメタデータを抽出する研究が行なわれている。

一方インターネット利用者のブログ開設率は 36.5% となる [1] など、個人からの情報発信は活発に行なわれている。そこで我々は動画とブログシステムを融合し、特定のシーンを話題にしたコミュニケーションを活性化することでユーザの自然な視聴スタイルからコンテンツのメタデータを獲得することができると考え、“コミュテンツ (Commutents)[4]” を開発した。

動画の特定のシーンを話題にしたコミュニケーションを活性化するシステムとしては SceneNAVI[3], ニコニコ動画 [5], Synvie[2] などが知られている. SceneNAVI ではシーンごとに用意された掲示板に対して感想を投稿する. 掲示板は映像と同期しており, シーンの経過に応じて次々に切り替わっていく. ニコニコ動画や Synvie では任意のシーンに対して感想を投稿する. 特にニコニコ動画は短い感想を字幕として表示することで不特定多数と一緒に盛り上がる雰囲気を生み出している. これらのシステムでは特定のシーンに対して感想を投稿するため自分が気になったシーンのほかの利用者の感想は発見しやすいが、一人の利用者の感想がさまざまなシーンに分散してしまうため関連したシーンに対する感想をまとめて読みづらい。また「キター！」のように短くて情報量の少ない感想が多くなりやすい。

我々が必要とするシーンごとの「誰が何をどうしたのか」、あるいはシーン同士の関連を表すメタデータは、レビュー記事のように長くて情報量の多い感想からの方が抽出しやすいと考えられる。そこで我々は感想を一箇所にまとめられ、その中でシーンに言及することができるブログというモデルを採用し、ブログ記事の中にシーン情報を埋め込むことができる仕組みを開発した。図 1 はコミュテンツを用いてシーン情報を埋め込まれたブログ記事の例である。挿入されたアイコンは参照先コンテンツの ID と時刻を保持している。具体的なシステム構成、動作については [4] を参照されたい。

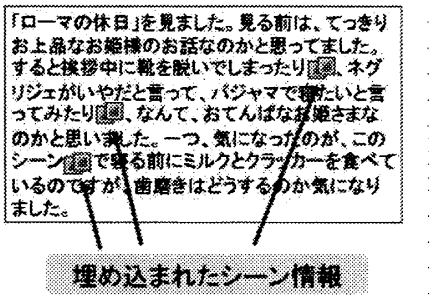

図 1: シーン情報が埋め込まれた記事の例

本稿ではコミュテンツを利用して執筆されたブログ記事からシーンごとの「誰が何をどうしたのか」というメタデータを抽出する方法について説明し, 実際の利用者 24 人分の記事から抽出したメタデータについて精度を検証する.

2. メタデータ抽出実験

本章ではシステムの利用実験、ならびにシステムを利用して執筆された記事からシーンごとのメタデータを抽出した実験の結果について説明する。表 1 はシステムを利用した 24 人によって 1 ヶ月間に執筆された記事の収集結果である。23 コンテンツに対し 259 記事が執筆され、10690 シーン情報が埋め込まれたことを示している。

表 1: 記事収集結果

コンテンツ	記事	シーン	文字	返信
オズの魔法使	8	271	22644	49
バック・トゥ・ザ・フューチャー	18	404	33847	55
パイレーツ・オブ・カリビアン	25	850	67127	88
ローマの休日	28	776	54530	67
...	...	...	...	...
合計	259	10690	852384	1034

本システムではコンテンツは共有せずシーン情報を埋め込んだブログ記事のみを共有する. ブログ記事はシーンに対する短い感想ではないため、文章が長く複数のシーンに言及することが多い。したがってシーンごとの

†株式会社東芝 研究開発センター Corporate Research & Development Center, TOSHIBA Corporation

メタデータを抽出するためには感想からそれぞれのシーンに対応する記述を切り出す必要がある。メタデータ抽出処理の流れを図 2 に示す.

今回の実験では感想テキストを形態素解析し、埋め込まれたシーン情報の前後 n 形態素の範囲をそのシーンに対応する記述とした。続いて切り出されたシーン記述のうち、時刻の差が閾値以下など同じシーンに言及していると考えられるものをひとまとめにし、それぞれのシーンごとに頻出するキーワードを抽出する。その後、抽出されたキーワードが主体、対象、行動などのどれに相当するかを国立国語研究所資料集１４「分類語彙表」を用いて分類する。分類の元データは分類語彙表の分類番号、分類項目ごとに妥当と思われるものを用意した。

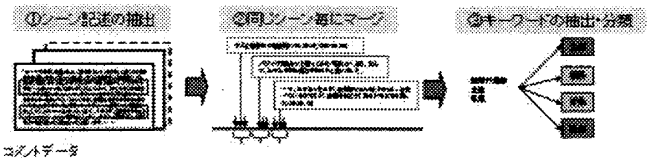

図 2: メタデータ抽出処理の流れ

また「・・・○○シーン (場面、台詞)」や「・・・○○。このシーン (場面、台詞)・・・」のように、感想ではなくシーンを説明する表現であることが明示的に示されている箇所はシーン検索のメタデータとして非常に有用なため，その表現をシーン説明表現として抽出する。

表 2 はローマの休日に関連する記事からシーンを表すキーワードとシーン説明表現を抽出した結果の例である.

表 2: 「ローマの休日」

00:56:27	主体	アン王女, おばさん、ブラドリー
	場所	魚屋, 街中, 市場
	行動	買い物, 探索, カット
	説明	王女が街中を冒険するシーン
01:15:37	主体	アン王女, ジョー, 警察
	場所	街中, ヴェネツィア広場
	行動	暴走, 運行
	説明	説明不要なほど有名な街中で スクーターを乗り回すシーン

3. 評価

一つの動画の中で感想を書きたくなるシーンは利用者間で重複することが多いと考えられるため、記事に埋め込まれるシーン情報の分布は粗密があると考えられる。コミュテンツによる記事に埋め込まれていないシーンのメタデータは本手法では原理的に抽出できないため、カバーできない範囲が広いことは望ましくない。そこでそれぞれのコンテンツについてカバーできない範囲がどの程度あるかを調査した。図 3 に誤差の許容量とカバーできなかった範囲の関係を示す。どのコンテンツでも誤差 5 秒以内ではすべてのシーンの 50〜60% をカバーできないが、誤差 15 秒以内では 20〜30%、30 秒以内では 10% 程度に抑えられることがわかる。そのため記事に埋め込まれるシーン情報からも現実的な範囲で任意のシーンのメタデータを抽出することができるものと結論付けることができる.

続いて，抽出すべきシーン代表語の正解をバック・トゥ・ザ・フューチャーとオズの魔法使のすべてのシーンについて作成し，シーン代表語を選択する範囲をシーン情報の前後 $n = 3 \ldots 20$ 形態素まで変化させたときの抽出結果と比較して精度を検証した結果を図 4 に示す.

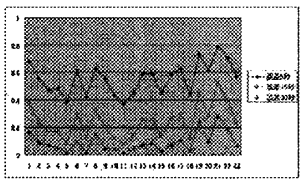

図 3: 誤差の許容量とカバー失敗率

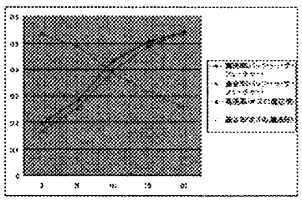

図 4: 代表語を選択する範囲と正解精度

抽出されたシーン代表語は比較的妥当である印象を受けるが、今回の実験ではどちらのコンテンツも抽出精度は 0.4 程度となった。これはブログ記事のテキストがフォーマルでないため、形態素解析に失敗しやすいこと、同義語や複合語の表記にゆれがあることなどが原因としてあげられる。特に登場人物は、利用者によって登場人物名，役者名，ニックネームなどさまざまに表記されるため、精度向上のためには異表記の同一性判定が重要である。

4. おわりに

本稿では映像コンテンツとブログシステムを融合したコミュテンツ (TM) によるブログ記事から実際に、誤差 30 秒では 90% の範囲でコンテンツのメタデータをシーンごとに獲得できることを示した．また抽出したメタデータの精度は前後 10 形態素を用いた場合再現率も適合率もほぼ 0.4 となった.

精度を下げている大きな要因は同義語・類義語・言い換えなどであると考えられる．今後はシソーラスやオントロジーを導入して精度を向上して行く予定である.

参考文献

[1] 「ブログ」に関する C-NEWS 生活者評価. http://www.info-plant.com/research/00332.html, 2006.

[2] 山本大介ほか: Synvie:映像シーンの引用に基づくアノテーションシステムの構築とその評価, インタラクション 2007, pp.11–18 (2007).

[3] 東正造ほか: 多種類のテレビ映像を対象とした映像シーン連動型掲示板におけるコミュニケーションの分析, 情処 GN 研 2006, pp.31–36 (2006).

[4] 筒井秀樹ほか: ブログと映像コンテンツを介したコミュニケーション支援システム “コミュテンツ (TM)” の開発, 情処 HCI 研 2007

[5] ニコニコ動画, http://www.nicovideo.jp.