

Synvie: ビデオブログコミュニティからの知識獲得とその応用

山本 大介[†] 増田 智樹^{††} 大平 茂輝[‡] 長尾 確^{††}[†]名古屋大学 情報科学研究科 ^{††}名古屋大学 工学部 電気電子・情報工学科[‡]名古屋大学 エコトピア科学研究所 ^{††}名古屋大学 情報メディア教育センター

1 はじめに

近年, YouTube や Google Video などといったビデオ共有サービスが盛んに提供されている. これらのサービスでは, ビデオコンテンツの投稿や閲覧のための機能のみを提供するだけではなく, ビデオコンテンツに対するコメント付与・投票・評価・タグ付け等を行うための機能や, ブログエントリー上でのビデオコンテンツの埋め込みを支援する機能を提供しているものも多い. 当研究室では, Synvie[1] というビデオ共有サービスを試験的に提供[†]している. Synvie では, 「ビデオコンテンツを中心としたブログコミュニティの形成を支援し, そのコミュニティにおける知的活動からコンテンツに関連する知識をアノテーションとして収集すること」を目的とし, 既存のビデオ共有サービスにはない新しい機能やサービスを提供している. 特に, 任意の映像シーンに対するユーザコメントの投稿やボタン評価などのアノテーション機能や, 任意の映像シーンをブログへ引用するためのインタフェースを提供する機能などが新しい. また, それらを全ての編集履歴をアノテーションとして蓄積している. 本稿では, ユーザによる編集履歴と対応するコンテンツ(のシーン)とを詳細に関連付けたデータをアノテーションと定義する. Synvie によって収集された, これらの多様なアノテーションに基づくアプリケーションシステムを構築し, これらのアノテーションに基づくアプリケーションの有用性について議論する.

2 Synvie によって収集されるアノテーション

Synvie によって収集されたアノテーションを, 対象単位, 行為, データ型で分類したものを表 1 にまとめる. 映像の任意のシーンに対するアノテーションをシーンアノテーションと呼び, 特に, コメント投稿をシーンテキストアノテーション, 画面矩形領域に対するコメント投稿をシーン領域テキストアノテーション, ボタンによる評価をシーンボタンアノテーションと呼ぶ. 映像の任意のシーンを引用してブログエントリー上でコメントを記述し映像シーンとコメントとの関連付けすることをシーン引用アノテーションと呼び, 特に, 連続するシーンを引用した引用アノテーションを連続シーン引用アノテーション, 連続しない複数シーンを引用した引用アノテーションを非連続シーン引用アノテーションと呼ぶ. また, 映像コンテンツ全体に対するコメント付与をコンテンツテキストアノテーション, タグ付与をコンテンツタグアノテーションと呼ぶ. それぞれのアノテーションに対応した専用のインタフェースを開発し, ユーザがそれぞれのアノテ

ーションに最適化されたツールを用途や目的によって使い分けることができる. また, タイトルやコンテンツの紹介文などあらかじめ入力されているメタデータ情報もアノテーションとして利用する.

Synvie によって取得されるアノテーションには以下のような特徴がある. アノテーションの量はコンテンツの面白さや話題性などに依存しやすい. 一般ユーザにとって関心のあるシーンにより多くのアノテーションが集まりやすいため, 時間軸に対するアノテーション密度の偏りが大きい. また, アノテーションの質や信頼性は, アノテーションを付与する人の知識や性格に依存し, また, シーンアノテーションよりもシーン引用アノテーションの方がよりテキスト品質が高いなどアノテーションの編集法に依存することが分かっている. そのため, アノテーションの量や偏りにロバストであること, アノテーションの質は人や編集スタイルに依存しやすいという特徴を考慮する必要がある.

3 アノテーションの解析

一般的に映像コンテンツの要約や検索, 推薦といったようなアプリケーションを構築する場合, それぞれのアノテーションやシーン間の関連性や重要性(重み)の値を計算することが重要である.

3.1 ブログエントリーに基づく意味的な関連性

Synvie の特徴の一つとして, パラグラフ単位に構造化された, ビデオシーンを引用したブログエントリー(ビデオブログ)を執筆できる点にある. 一つのパラグラフ内において, 複数の映像シーンを引用することが可能であり, そのパラグラフ内のコメントの意味内容に基づき, それらのシーンは何らかの関連性があると推定できる. さらに, ビデオブログは映像シーンを引用した複数のパラグラフから構成される. ユーザは, 何らかのテーマに基づいてブログエントリーを執筆しているため, それらのパラグラフ間には何らかの関連性が期待できる. 特に, 二つのパラグラフ間で共通するキーワードが多く存在する場合, それらのパラグラフ間の関連性は強いと考えられ, キーワードの共起関係からパラグラフ間の関連性を推定する. このことは間接的に, それぞれのパラグラフの引用元映像シーン間に共起したキーワードに基づく関連性があるといえる. これらの関係をグラフで表現すると図 1 のようになり, グラフ間の距離がそれぞれのアノテーションやシーンの意味的な関連性を表し, 検索や推薦などと

表 1: 公開実験によって取得されたアノテーション

対象単位	行為	データ型	アノテーションタイプ
コンテンツ	投稿	単語	コンテンツタグ
	—	文	コンテンツテキスト コンテンツメタデータ
シーン	投稿	ボタン情報	シーンボタン
	引用	文	シーン領域テキスト シーンテキスト 連続シーン引用 非連続シーン引用

Synvie: Video Applications Based on the Knowledge Acquired from Video Blog Communities.

[†] YAMAMOTO, Daisuke(yamamoto@nagao.nuie.nagoya-u.ac.jp)^{††} MASUDA, Tomoki(masuda@nagao.nuie.nagoya-u.ac.jp)[‡] OHIRA, Shigeki(ohira@nagao.nuie.nagoya-u.ac.jp)^{††} NAGAO, Katashi(nagao@nuie.nagoya-u.ac.jp)

Nagoya Univ. (†)

Furocho, Chikusa-ku, Nagoya, Aichi 464-8603, Japan

[†] <http://video.nagao.nuie.nagoya-u.ac.jp/>

いったアプリケーションに利用可能である。

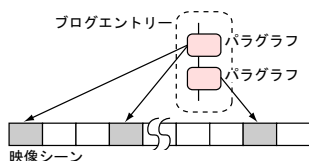


図 1: ブログエントリーとビデオのリンク

3.2 アノテーションとシーンの重み

本稿では、アノテーションの重みは、アノテーションの対象粒度、アノテータの信頼性、アノテーションタイプから推定する。つまり、信頼できる人がより正確にアノテーションを作成できるツールを用いて、より粒度の細かい対象（コンテンツよりもシーン、長いシーンよりも短いシーン）に対するアノテーションを付与した場合に、より高い重みを与える。本来ならばアノテーションの意味内容を加味したアノテーションの重み付けをすることが望ましいが、本稿では意味内容を考慮したテキスト解析は一般に困難であるため見送った。将来的には実データを基に学習アルゴリズムを用いることによってより正確な値を算出する予定である。

また、映像シーンの重みは、より多くの、アノテーションの重みが大きいアノテーションから参照されているシーンほど重要であると仮定し、それぞれのシーンを参照するアノテーションや引用の数、アノテーションの重みに基づいて決定する。

4 アノテーションに基づく応用

本実験によって取得されたアノテーションに基づく応用の例として、新しい発想に基づく、ビデオシーン検索システム、ビデオ推薦システム、ビデオスキミングシステムを提案する。従来のシステムには、自動映像解析技術に基づくものや半自動アノテーション技術に基づくもの等が存在するが、映像解析技術の精度やコスト等で問題がある。本稿で提案する応用は、ユーザから収集したアノテーションテキストや、アノテーションとシーン間の関連性や重みに基づいて実現しており、映像解析技術の精度や人的コストの問題が発生しないという利点がある。

4.1 映像シーン検索

ユーザは、検索したい対象を検索するために、一つないし複数の検索キーワードを入力する。それぞれの検索キーワード毎に全文検索を行い、そのキーワードを含むアノテーションを列挙する。アノテーションを検索することによって、間接的にビデオを検索しようとする試みである。

コンテンツ毎に、全文検索にマッチしたアノテーションについて考える。関連性の高い複数のアノテーションに対応する映像シーンは同じトピックに属している可能性が高いため、それらの映像シーンを一つの検索結果候補として扱う。検索結果候補内に属するアノテーションの重みの合計が、その検索結果候補の重みとする。ただし、アノテーション数が少なく分散しており、検索結果のシーンを特定できない場合は、シーンを絞ることは困難であると判断し、コンテンツ全体を検索結果とする。これにより、十分なアノテーションが存在しない場合にも検索漏れを防ぐことが可能になり、再現率の向上が期待できる。検索結果候補の重みに基づき、検索結果のランク付けを行う。

検索結果の内容を理解するために、検索シーンと検索キーワードに適したサムネイル画像を表示することが有効である。サムネイル画像はそれぞれの検索結果に対して一定個数を表示する。なお、本システムでは5個表示する。

サムネイル画像は、検索キーワードに合致したアノテーションに関連付けられているシーンに属するサムネイルを候補とする。ただし、サムネイル画像が一定個数以上存在する場合には、そのサムネイル画像が属する映像シーンの重みに基づいて絞る。ビデオシーン検索システムのインタフェースを図2に示す。



図 2: ビデオシーン検索インタフェース

4.2 その他の応用

映像と同期して関連性のある他のコンテンツのサムネイル画像とキーワード、及びその根拠となるブログエントリーを表示することによって、ビデオ推薦を行うシステムを開発している。従来からあるソーシャルフィルタリングを用いたコンテンツ推薦システムでは、映像のシーン単位で推薦を行うのは困難であること、サンプル数が少ない場合には必ずしも精度が良くないこと、さらに、推薦となる根拠を統計的にしか示すことができないなどの欠点がある。本システムでは、ユーザが複数コンテンツを引用したビデオブログエントリーを執筆した情報を用いて、統計情報に頼らない詳細なコンテンツ推薦を実現している。

さらに、ビデオスキミングシステムを開発している。ビデオスキミングとは、映像の重要なシーンのみを通常の速さで再生し、それ以外のシーンを早送りで再生する仕組みである。これにより、映像の内容を短時間で把握するのに適している。具体的には、映像シーンの重みや意味的な関連性に基づき再生速度を決定する。映像シーンの関連性は、重要なシーンに関連しているシーンは重要であるという考えに基づき関連する映像シーンの重みを伝播させる場合や、関連しているシーンは内容の重複があるという考えに基づき映像シーンの重みを減衰させる場合など、ユーザの目的や短縮したい時間によって使い分ける。

5 おわりに

本稿において、ビデオブログコミュニティから抽出されたアノテーションに基づく応用システムをいくつか紹介した。今後の課題として、応用システムの評価に基づく Synvie のアノテーションシステムとしての有用性の検証と改良がある。

参考文献

- [1] 山本大介, 清水敏之, 大平茂輝, 長尾 確: Synvie: ログの仕組みを利用したマルチメディアコンテンツ配信システム, 情報処理学会第 58 回グループウェアとネットワーク研究会, pp. 13-18 (2006).