

ニコニコ探検くらぶ： ソーシャルアノテーションとキーワード群に基づく動画要約

多胡 厚津史[†] 中川 博之[†]
田原 康之[†] 大須賀 昭彦[†]

NicoNico Exploration Club: Summarizing Videos by Keywords based on Social Annotation

ATSUSHI TAGO,[†] HIROYUKI NAKAGAWA,[†] YASUYUKI TAHARA[†]
and AKIHIKO OHSUGA[†]

1. はじめに

記憶装置の大容量化やブロードバンドインターネットの急速な普及を受け、YouTube などの動画共有サービスが出現し、世代によってはTVの視聴割合が減り、インターネット上の動画サイトの視聴割合が増えている¹⁾など、動画共有サービスはテレビと並び主要なメディアとしての地位を確立しつつある。さらに、ニコニコ動画をはじめとする動画に再生時間と同期したコメントを付加できるサービスが、双方向性の高いコミュニケーションメディアとして注目されている。その一方で、手軽に動画を Web 上に公開できることから、動画数は急激な増加傾向にあり、限られた視聴時間内で動画を効率良く閲覧する必要性が出てきている。

その1つの解決方法として、動画要約がある。動画要約は動画内の各シーンがどのような内容を表現しているかを分類することによって動画を要約し、動画1つ当たりの再生時間を短縮して、短時間での視聴を可能にするものである。動画要約を実現するためには、要約に必要な各シーンを切り出して組み合わせるため、シーン検索の実現が必要である。さらに Web 上に存在する無数の動画を対象とするため、アノテーションの付与を低コストで行うことができ、またコン

テンツに依存しない汎用性の高い検索手法が望まれる。

そこで本研究では、ニコニコ動画のコメントを時間軸に沿って意味的、量的に分析し、動画の各シーンに対するソーシャルアノテーションとみなすことで、シーン検索・動画要約を行うシステム「ニコニコ探検くらぶ」を提案する。本システムでは、従来コストのかかっていたアノテーション付与を自動化できる他に、コメントの多様性を生かすことで従来手法では検索が難しかった感性的なキーワードによるシーン検索や盛り上がり方を考慮したシーン検索が行える。

2. ニコニコ探検くらぶ

本研究では、ニコニコ動画のコメントをソーシャルアノテーションとみなしてシーンの検索、動画の要約を実現する「ニコニコ探検くらぶ」を提案する。ニコニコ探検くらぶは、Flex により作成したユーザーインターフェイスを ExternalInterface を用いて C# でラッピングすることで実装されており、動画検索タブとシーン検索タブから成る。また、サムネイルの作成には ffmpeg を用いている。

ニコニコ探検くらぶの利用方法は以下の通りである。

- (1) 動画検索タブにおいて任意のキーワードを入力して動画を検索し、要約したい動画を画面右下の「箱」にドラッグアンドドロップして指定する。動画は複数指定することも出来る(図1)。
- (2) シーン検索タブにキーワードを入力すると、システムではその語に関連するオントロジーによって定義されたキーワード群をコメントから

[†] 電気通信大学大学院情報システム学研究科
Graduate School of Information Systems, University of
Electro-Communications
<http://youtube.com/>
<http://www.nicovideo.jp/>



図 1 動画の指定 .
Fig.1 Video selection.

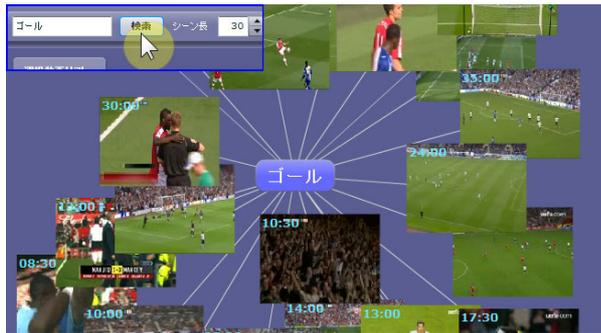


図 2 シーンを検索 .
Fig.2 Scene search.

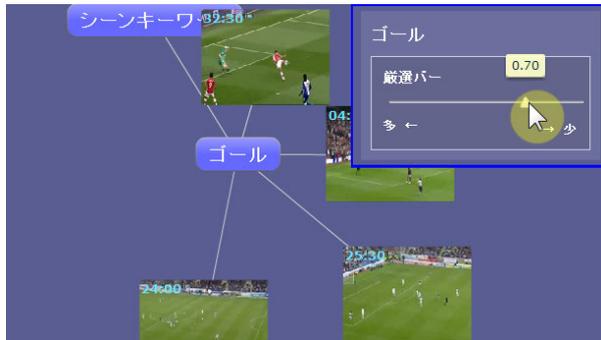


図 3 「厳選バー」による検索結果の変化 .
Fig.3 Changing search results caused changing "top pick bar" value.

- 探し出し、キーワード群からシーンを検索して、SpringGraph により視覚化する (図 2)。
- (3) キーワードの「厳選バー」を操作し、キーワードとの関連性の強弱によって表示するシーンの数を増減することができる (図 3)。また、キーワードは複数指定でき、「面白い」や「下手」といった感性的なキーワードも使用できる (図 4)。
 - (4) 検索結果から要約に用いたいシーンを選択し、連続再生することで動画要約結果として出力する (図 5, 図 6)

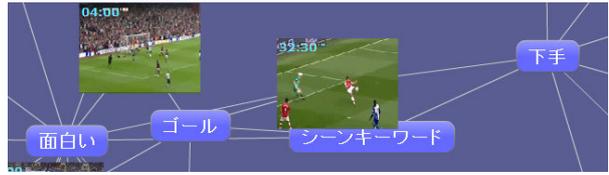


図 4 キーワードの例 .
Fig.4 Example of keywords.



図 5 シーンを選択 .
Fig.5 Scenes selection.



図 6 要約動画の再生 .
Fig.6 Playback of summarized video.

3. 手 法

3.1 コメントに基づくユニットのスコアリング

ニコニコ探検くらぶでは動画を任意の単位時間でユニットに区切り、各ユニットをコメントの内容とコメントの量、すなわち盛り上がりに基づいてスコアリングし、閾値以上のスコアを持つユニットをシーン検索結果として返す (図 7)。以下にその手順を示す。

- (1) 長さ T 、コメント総数 C の動画を、ジャンルや種類に応じて変えられる任意の単位時間 t のユニット U_i ($i = 1, 2, \dots, N$) に分割する。
- (2) ユニット U_i に含まれるコメント総数 c_i ($\sum_{i=1}^N c_i = C$) と、キーワードが含まれるコメント数 K_i を数える。キーワードはキーワードオントロジーのキーワード群に含まれる全て

ニコニコ探検くらぶ：ソーシャルノテーションとキーワード群に基づく動画要約

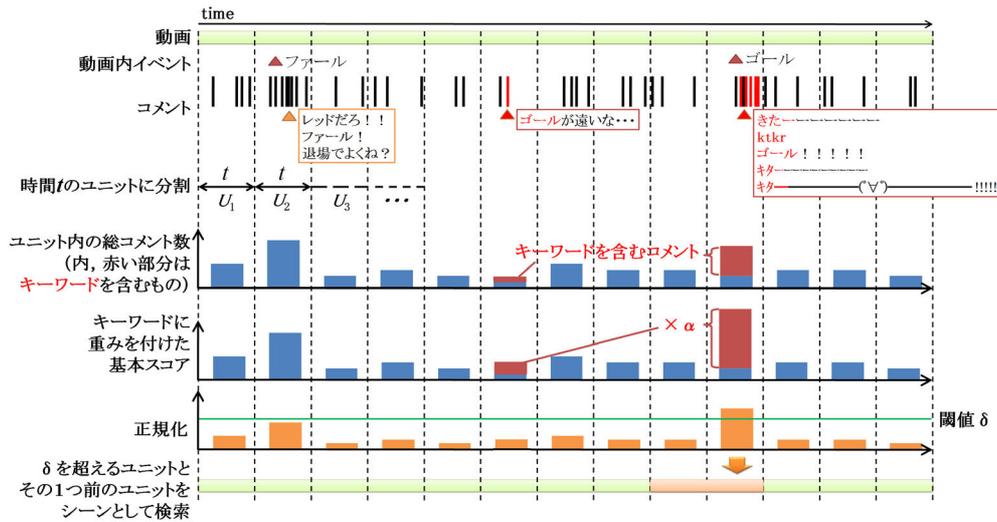


図 7 コメントに基づくユニットのスコアリングとシーン検索.
Fig. 7 Scene search and unit scoring based on comments.

- を探す.
- (3) K_i と重み α の積をとり, c_i に加えることで基本スコア $S_i (= c_i + \alpha K_i)$ を求める. ここでいう重みは厳選バーによって変化する値である.
 - (4) 動画のコメント総数 C やユニット長 t による検索結果のブレを防ぐため, 1 ユニット当たりの平均コメント数 Ct/T を求め, これにより正規化された S_i をユニットスコア S'_i とする.

$$S'_i = \frac{T}{Ct} S_i \quad (1)$$

- (5) ユニットスコア S'_i が閾値 δ を超えたユニット U_i の近傍を検索結果として返す.

3.2 キーワードオントロジーの利用

ニコニコ探検くらぶではキーワードと多様なコメントを結びつけるのにキーワードオントロジーを用いる. 例えば, キーワードとして「ゴール」が入力された時, コメントに書かれた「ゴール」の語の他に, 「キター」や「おおお」などといった語も検索対象とする (表 1). このキーワードオントロジーは, ニコニコ動画をよく利用し, かつその動画及びシーンに対して知識を持つ視聴者が協動的に編集する. オントロジーを用いることで人手による精度のよいキーワード群を手軽に構築でき, さらに様々なジャンルのシーンに対して構築できるなど汎用性も期待できる. これは, ニコニコ動画でのコメントの書かれ方に詳しい視聴者の知識を利用することで有用なものになると考えられる.

4. 精度評価

4.1 実験方法

本手法の有効性を評価するために, 長さ 20 分~65

表 1 キーワードオントロジーの例.
Table 1 Example of keyword ontology.

シーン属性	キーワード群
ゴールシーン	{ "ゴール", "キター", "おおお", ... }
下手	{ "下手", "へたくそ", "決めろ", ... }
麻生太郎	{ "麻生", "首相", "総理", ... }
高槻やよい	{ "やよい", "うっうー", "もやし", ... }
...	...

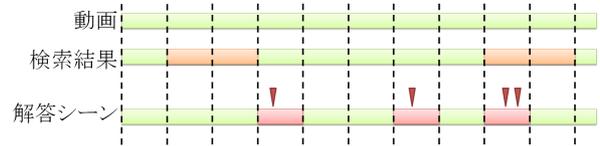


図 8 解答の定義.
Fig. 8 Definition of answer.

分のサッカーの動画 17 本に対しゴールシーンを人手によって探し出した結果と, ニコニコ探検くらぶがゴールシーンとして返した結果を比較して適合率, 再現率を計測 する実験を行った.

実験に際し, キーワード群はすべて, ニコニコ動画を頻繁に利用するユーザが予め作成したものを用い, 評価は F 値を用いて行った. 評価にあたり, キーワードを含むコメントへの重み α とスコア閾値 δ をそれぞれ $\alpha = 0, 3, 5, 10, 17$, $\delta = 1, 2, 3, 5, 7, 10, 13, 16, 20$ と変えて計測した.

4.2 実験結果

実験結果を図 9 に示す. この結果から, $\alpha = 10$,

例えば図 8 では, 正解は 1, 外れは 1, 解答は 3 で, 適合率は 50% (1/2), 再現率は 33.3% (1/3).

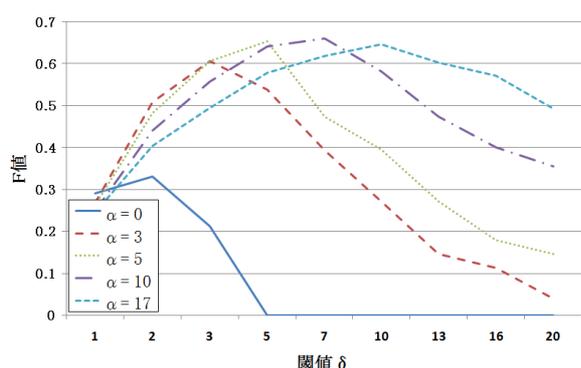


図 9 ゴールシーン検索の F 値 .
Fig. 9 F-value of searching goal scene.

$\delta = 7$ において F 値 ≈ 0.66 であり、一定の精度で検索が可能であることが分かる。しかし、 α の値により F 値も大きく変化することから高精度な結果のためには動画に応じた α を求める必要がある。また、 α が大きくなるにつれて再現率は高くなる傾向にあった。ニコニコ探検くらぶでは厳選バーによって α を変えることで、検索結果をユーザが選択しやすい数量に調整できるユーザインターフェイスを採用している。

5. 関連研究

アノテーションとは特定の情報をもとにシーンの内容を判断してメタデータとして付加することを指すものであり、様々な手段を用いた手法が提案されている。杉山ら²⁾ はニュースの音声に対し話者分割を行うことでアノテーションを行い一定の成果を挙げている。新田ら³⁾ はスポーツ動画に対し映像解析を行うことでどのようなプレイが行われたかを調査するシステムを提案している。ただし、これらの手法はコンテンツに依存したものであり、Web 上に無数に存在する動画に対して適用することは困難である。

その一方で、集合知を利用したソーシャルアノテーションも注目されている。Miyamori ら⁴⁾ はインターネット掲示板の書込と TV を同期させ、書込を TV へのアノテーションとみなすことで、アメリカンフットボールの試合を要約している。これらの手法は低コストや高い汎用性というメリットがある反面、これらを直接適用したとしてもノイズが多いことから適切な情報を取得できない、情報が膨大なことから特定のシーンを判断・抽出することが難しい、そもそもアノテーションが集まらないといった問題がある。

ニコニコ動画を扱った研究も近年見られるようになってきた。青木ら⁵⁾ はコメントを定量的に評価することで動画のサビの検出を試みている。Hamasaki

ら⁶⁾ は、ニコニコ動画上のユーザ投稿コンテンツの繋がりを追うことで、ユーザの創作活動の繋がりを解析している。

これらの研究に対して本手法では、コメントの多様な意味と盛り上がりを生かした検索が行える点、双方向性メディアに蓄積されたコンテンツに対し自動でアノテーション付与が行える点が大きな特徴となっている。

6. まとめ

本論文ではニコニコ動画のコメント量と内容を統合的に評価することで、シーンの検索、動画要約を実現するシステム、ニコニコ探検くらぶを提案した。ニコニコ探検くらぶでは直感的な操作を多く取り入れるとともに、実験により一定の精度でのシーン検索が可能であることを示した。今後の課題としては公開を目的とした、検索精度の向上、動画の長さを指定した要約を検討している。

参考文献

- 1) NTT レゾナント, 三菱総合研究所: 第 6 回ブロードバンドコンテンツに関する調査, 報道発表資料 (2007).
- 2) 杉山義昭, 西川昌史, 有木康雄: Video assembler: 動画処理と音声処理の統合によるニュース記事内容の分割, 第 4 回知能メディアシンポジウム論文集, pp.23-30 (1998).
- 3) 新田直子, 馬場口登, 北橋忠宏: 放送型スポーツ映像の構造を考慮した重要シーンへの自動アノテーション付け, 電子情報通信学会論文誌, Vol.J84-D2, No.8, pp.1838-1847 (2001).
- 4) H. Miyamori, S. Nakamura and K. Tanaka: Generation of views of TV content using TV viewers' perspectives expressed in live chats on the web, *Proceedings of the 13th annual ACM international conference on Multimedia*, pp.853-861 (2005).
- 5) 青木秀憲, 宮下芳明: ニコニコ動画における映画要約とサビ検出の試み, 情報処理学会研究報告 2008-HCI-128/2008-MUS-75, Vol.2008, No.50, pp.37-42 (2008).
- 6) M. Hamasaki, H. Takeda and T. Nishimura: Network Analysis of Massively Collaborative Creation of Multimedia Contents - Case Study of Hatsune Miku videos on Nico Nico Douga -, *in First International Conference on Designing Interactive User Experiences for TV and Video (uxTV2008)*, pp.165-168 (2008).