

説明文を付した動画ファイルに対する 文章比較モデルを用いた検索

白井宏一

f-shirai@nifty.com

岡山理科大学

椎名広光

shiina@mis.ous.ac.jp

総合情報学部

1 まえがき

近年、カメラ付き携帯電話やデジタルビデオカメラが普及してきており、またコンピュータのハードディスクドライブの大容量化やネットワーク回線の高速化によって、個人でも動画データをコンピュータ上に保管し、それらを検索して再生することも多くなった。自分で作成した動画を Web サーバに保存し、その中から必要な動画を検索して再生するためのサービスが今後もっと必要になると考えられる。

また、Web 上における情報検索 [2] では、単語頻度をベースに実現される例が多く、画像検索においても画像が保存されている HTML に対する単語を手がかりに検索結果を提示している。文の構造を用いるには文章の大きさが不均衡であったりし、Web 検索を用いるのは難しいと考えられる。

本研究では、動画ファイルサーバを XML サーバ [3, 4] を用いて作成し、動画ファイルの検索を容易にするために、あらかじめ動画ファイルにつけられた説明文 (以下、動画説明文) と検索する自然言語文 (以下、検索文) との類似性を求めて動画ファイルを提示するシステムを作成した。

特に、動画説明文と検索文の類似については 2 種類の方法を提案している。1 つ目の手法として、動画説明文と検索文に対して、形態素解析、係り受け解析を行い格フレーム形式に変換後、係り受け間の重みを計算して類似性を求める方法である。2 つ目の手法として動画説明文と検索文に対して、アニメーションの人の動作を表現を 4 つのモデルに分ける文章構造モデルを当てはめ、そのモデル毎の類似を求め最も高いものを類似値として計算する方法である。また、最後に 2 つの手法に対する評価について述べる。

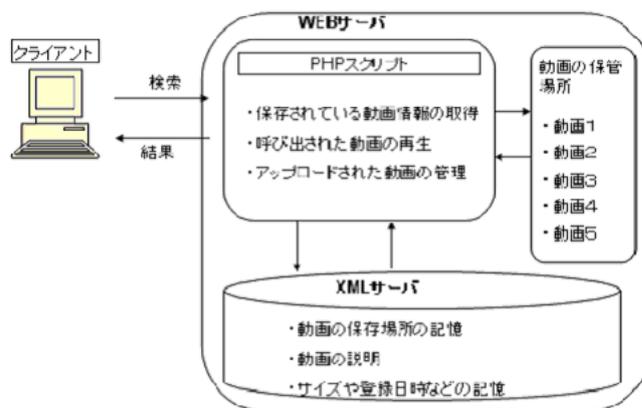


図 1: 検索システムの概要

検索ワードを入れてください:
検索: バク転をする。

動画 NO.	説明文	スコア	
M0	ジェームズがバク転をする。	1.00000	<input type="button" value="再生する"/>
M2	軽べジェームズがジャンプをする。	0.66667	<input type="button" value="再生する"/>
M3	ジェームズが縄のりの練習をしている。	0.66667	<input type="button" value="再生する"/>
M4	ジェームズが手話の練習をしている。	0.66667	<input type="button" value="再生する"/>

図 2: 検索結果の表示

2 動画サーバの検索システムの概要

本研究では、システムを Web サーバ (Apache) から CGI スクリプトを通じて XML サーバ (Xpriori[4]) に問い合わせを行う形式で構成し (図 1)、結果は XML サーバから取得できた情報を HTML に CGI スクリプトで編集した上で提示している。

XML サーバでは、動画ファイルの説明文、説明文を解析したデータ、動画の登録場所を管理しており、実際の動画ファイル自体は XML サーバに格納しない方

式をとっている。また、検索の提示は CGI スクリプトが XML サーバに問い合わせた結果からファイルの場所を取り出し、ユーザに動画再生を提供する方式である。

システムの機能としては、下記に示すと 4 つの機能を作成している。

- (1) 動画ファイルの登録・削除
- (2) 動画ファイルに対するコメントの修正
- (3) 格フレーム間の重みによる類似性による動画検索 (方法 1)
- (4) 文章構造モデル化による類似性による動画検索 (方法 2)

なお、検索に関する仕組みについては、次のとおりである。

- 動画ファイルを登録するとき、その動画ファイルを説明を格納しておくことで、XML サーバから検索を行う際には、説明を手がかりに動画ファイルを提示する。動画説明文を手がかりにするにあたって動画説明文をあらかじめ形態素解析、係り受け解析をおこない、そのうえで動画説明文を格フレームで登録する。リンクを除いた格フレーム形式は、XML データ形式とは等価に扱うことができると考えられるので、動画説明文の格フレーム形式と動画説明文の XML データを同じものとして取り扱う。
- 検索では、文を用いて検索を行い、検索文と類似する動画を提示する。検索文と動画ファイルとの類似については、検索文も形態素解析、係り受け解析を行い格フレーム形式に変換後、検索文の格フレームと動画ファイルの説明文の格フレームを間の類似性と計算して、検索文からみて最も高い類似した動画を提示する。

3 格フレーム間の重みによる類似性の計算 (手法 1)

検索手法の 1 つ目としては、動画説明文と検索文に対して、形態素解析、係り受け解析を行い格フレーム形式に変換後、係り受け間の重みを計算して類似性を求める方法である。

3.1 格フレームの作成

本研究では、動画説明文や検索文を形態素解析と係り受け解析の結果を用いて、係られる文節の子要素として係る文節が入るような構造を作成する。以下では文 1 を格フレームに変換をする (図 3,4)。

文 1: ジェームスがうまく手話で話します。



図 3: 文 1 に対する係り受けの例

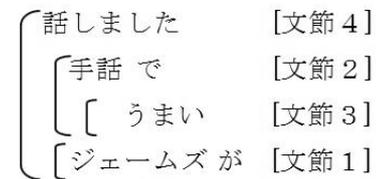


図 4: 文 1 に対する係り受け情報に基づいた格フレームの例

3.2 格フレームを使った類似度の計算方法

類似度の計算方法は、検索文、動画説明文それぞれの格フレームを並列する格フレーム同士で比較をし、一致した場合、その格フレーム重みを類似度に加えて求めている。格フレームごとの重みは、図 5 に示すような同格の格フレームに対して、主となる格フレームの重み α とその他の格フレームの重み $(\frac{1-\alpha}{\text{その他の各フレーム数}})$ で類似性を一律に決めている。また、それぞれの格フレームが複数の子供の格フレームに分割されている場合は、その親の格フレームの重さを引き継ぎ、親と同様にその子供の格フレームを主たる部分とその他に分けて重みを付けている。

例えば、文 1 では係り受けの関係から文節 1,2,4 が同列で、文節 3 が文節 2 に含まれる構造を持っている。この場合、動作主を主となるフレームとするので文節 4 「話す」の重みは α 、文節 1 「ジェームズ」、文節 2 「手話で」の重みは $\frac{1-\alpha}{2}$ となる。また、文節 2 は、「手話」と文節 3 「うまい」を含み、「手話」の重みは $\frac{1-\alpha}{2} \cdot \alpha$ 、「うまい」の重みは $\frac{1-\alpha}{2} \cdot (1-\alpha)$ となる。

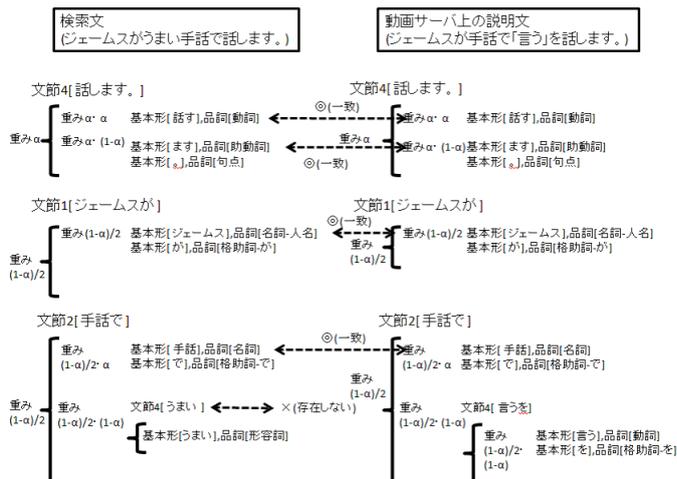


図 5: 格フレームの重みと文間の一貫

上記の格フレームの重みを用いて、動画説明文と検索文間の格フレーム間の重みによる格類似度 P_1 は次の通りである。

$$P_1 = \sum_{i=\text{一致する格}} \text{格フレーム } i \text{ の重み}$$

4 文章のモデル化による類似性の計算 (手法 2)

検索手法の 2 つ目は、動画を説明する文は、アニメーションの人の動作を説明することに限定していると考え、動画説明文は幾種類かの文構造を持つと想定できるのではないかと発想である。本研究では、次の 4 種類のモデルを想定し、動画説明文と検索文の類似度を計算している。

4.1 文章構造のモデル化

動画を説明する文のモデル (文章構造モデル) を以下に示す。

- (1) [動作主] が [名詞] に [述部] .
- (2) [動作主] は [並列句] と [述部] .
- (3) [動作主] が [動作対象] を [述部] .
- (4) [動作主] が [名詞] へ [動作対象] を [述部] .

検索文と動画説明文の形態素解析を行い、文節ごとに該当する文章構造モデルに当てはめる。検索例として「ジェームズが手話で太郎と話をする」に対して 4 種類の文章構造モデルに当てはめたものを図 6 に示す。

- (1) $\left[\begin{array}{l} \text{が} \quad [\text{ジェームズ}] \\ \text{に} \quad [\quad] \\ \text{述部} [\text{話す}] \end{array} \right.$ (2) $\left[\begin{array}{l} \text{が} \quad [\text{ジェームズ}] \\ \text{と} \quad [\text{太郎}] \\ \text{述部} [\text{話す}] \end{array} \right.$
- (3) $\left[\begin{array}{l} \text{が} \quad [\text{ジェームズ}] \\ \text{を} \quad [\quad] \\ \text{述部} [\text{話す}] \end{array} \right.$ (4) $\left[\begin{array}{l} \text{が} \quad [\text{ジェームズ}] \\ \text{へ} \quad [\quad] \\ \text{を} \quad [\quad] \\ \text{述部} [\text{話す}] \end{array} \right.$

図 6: 文章構造モデルへの適用例

4.2 文章モデルに当てはめた文間の類似度

文章モデルに当てはめた文間の類似度の計算は、検索文、動画説明文ともに同一モデル毎に類似度の計算を行い、一番類似度の高かった組をその動画の類似度とする。類似度: P_2 動画検索文の格助動詞: t とし、述部の重みを β とした場合、文章構造モデルにおける類似度を以下のように計算する。

$$P_2 = \text{述部の一致数} \times \frac{\beta}{\text{述部の形態素数}} + \sum_{i=1}^t \text{文節}_i \text{の一致数} \times \frac{1 - \beta}{t \times \text{文節}_i \text{の形態素数}}$$

5 動画説明文と検索文に対する類似度の評価

本研究では、2 種類の類似度の計算方法を評価のために、動画データベースに動画説明文を含む 50 件の動画データを保存し、それに対して全く同じ文章を検索文として使用した。なお、手法 1 の格フレーム間の重み α は 0.5、手法 2 の述部の重み β も 0.5 としている。

本稿では検索文の例として、検索文 1 「ジェームズが太郎とサッカーをしている」と検索文 2 「太郎とジェームズが将棋をしている」の 2 つに対する類似度の計算結果と検索順位 (表 1.2) の抜粋を示す。

5.1 格フレーム間の重みによる類似性の計算 (手法 1) の評価

格フレーム間の重みによる類似性の計算 (手法 1) では文章の係り受け関係が大きく評価に関わってきている。

検索文 1 の表 1 では M1 が最も高い類似度 1.00 である。M3 は接続助詞の「と」が存在しないため、その部分の点数だけ下がり 0.83 となっている。しかし、手法 1 では係り受け構造で類似性を見ているため M8 でも

表 1: 検索文 1 に対する類似度と検索順位

ID	動画説明文	P_1	順位	P_2	順位
M1	ジェームズが太郎とサッカーをしている	1.00	1	1.00	1
M2	太郎がジェームズとサッカーをしている	0.83	5	0.85	5
M3	ジェームズがサッカーをしている	0.83	5	1.00	1
M4	花子が太郎とサッカーをしている	0.91	3	0.87	5
M5	太郎が花子と歩いている	0.41	13	0.85	11
M6	太郎が机を蹴飛ばす	0.16	29	0.25	29
M7	ジェームズがボールを蹴る	0.25	25	0.37	25
M8	太郎とサッカーをジェームズがしている	1.00	1	1.00	1

表 2: 検索文 2 に対する類似度と検索順位

ID	動画説明文	P_1	順位	P_2	順位
M1	ジェームズが太郎とサッカーをしている	0.91	2	1.000	1
M2	太郎がジェームズとサッカーをしている	0.75	5	0.750	8
M3	ジェームズがサッカーをしている	0.750	5	0.87	4
M4	花子が太郎とサッカーをしている	0.83	4	0.85	4
M5	太郎が花子と歩いている	0.41	13	0.58	11
M6	太郎が机を蹴飛ばす	0.12	29	0.250	29
M7	ジェームズがボールを蹴る	0.12	23	0.37	25
M8	太郎とサッカーをジェームズがしている	0.91	3	1.00	1

係り受け先は検索文と同じ「している」の部分に全ての文節が係っているため類似度は 1.00 となっている。また、検索文 2 でも M1 と M8 が 0.91 と高い値を出している。これは「将棋を」の文節以外が高い一致を示しているからだと考えられる。

5.2 文章のモデル化による類似性の計算 (手法 2) の評価

文章のモデル化による類似性の計算手法 2 では用意した 4 つの文章構造モデルに当てはめそれらの中から類似度 P_2 が高いものを類似度としている。検索文 1 では M1 と M3, M8 が類似度 1 となり最大となっている。M1 と M8 には文章構造モデル (2) が、M3 には文章構造モデル (3) がそれぞれ適用された結果である。手法 2 では述部と動作主が揃えば高い評価が出やすい傾向がある。また、検索文 2 では M1 と M8 が類似度 1.00 となり、続いて M3, M4 となっている。手法 1 と違い高スコアの検索結果が多いのは、4 つの構造モデルによって、取捨される文節があり必ずしもすべての文節が比較対象になっていたためととえられる。

5.3 2つの手法の比較

格フレーム間の重みによる類似性の計算手法 1 では係り受け構造によって、係られる側の述部の比重が高く、それを補強するための係る側の比重は低くなるため動作が影響する割合が高い。

一方、文章のモデル化による類似性の計算手法 2 では手法 1 と違い係り受け構造の階層構造がないため、述部以外の文節の重さはすべて同じである。従って動作主に対しての評価が手法 1 よりも比較的上がって

くることが挙げられる。

6 まとめ

本研究では XML サーバ上に動画ファイルを説明する XML データベースを作成し、動画の説明文と検索文の類似性を考慮して動画ファイルを検索するシステムを作成した。類似性を求めるために格フレームによる係り受け情報を用いる方法と、文章構造モデルを使用した 2 つの手法を用いた。

述部を重要視する場合には格フレームを用いる手法 1 が適性が高く、動作主など述部以外の重要度は手法 2 を用いた検索をするよいと考えられる。

しかしながら、本研究では 2 つの手法とも単語の一致度を格フレームや構造モデルの一致としているため、類義語が考慮されていないことやまた、手法 2 に関しては構造モデルのパターンが十分でないため 2, 3 種類の構造モデルを追加することで精度が向上すると考えられる。また、手法 1 と手法 2 を合わせたモデルによる類似度計算についても検討し、精度の向上に行いたい。

参考文献

- [1] CaboCha, <http://chasen.org/taku/software/cabocha>
- [2] 北, 津田, 獅々子, 情報検索アルゴリズム, 2002
- [3] 石川, 次世代データベースとデータマイニング, CQ 出版, 2005
- [4] 山田, XML データベース入門, 翔泳社, 2006