

説明文を付したアニメーションに対する 複文による検索

白井宏一¹椎名広光²

chankou@3ls.sakura.ne.jp

shiina@mis.ous.ac.jp

岡山理科大学大学院 総合情報研究科¹ 岡山理科大学 総合情報学部²

1 まえがき

近年, YouTube やニコニコ動画に代表される動画投稿サービスが増えてきている. これには, 個人で動画を撮る環境が身近になったことに加え, ニコニコ動画などでは他の人が作成した動画コンテンツを再利用する文化などが出来上がり, 動画を作りそして公開するというニーズが高まってきたためだと考えられる. そのため, 多くの動画の中から自分の必要としているものがを探したり, 再生させたりするためのサービスが重要になってくると考えられる.

本研究で, 想定するシステムは, 動画に説明文を付して投稿をし, 検索を文によって, その検索文に類似する動画を複数個提示するシステムを想定している.

著者らは, これまで動画の説明文 (以下, 説明文) と検索文 (以下, 検索文) の類似度を計算するのに, 文を格カテゴリーに分け, その精密度と再現率の調和平均を格カテゴリーの有無による補正する手法を提案した [4].

これまでの手法は単文間の類似度を計算しているが, アニメーションの 1 つの動作でだけではないので, 複数の動詞がある文や複文による説明文や検索文への対応が必要である. そこで, 本稿では動画説明文と検索文の類似度の計算方法については, まず, 文が連続する複数の文からなるものであれば, 動作ごとに区切った文を作る. その後, それぞれの文を 1 つの動作モデルとして, 動作主, 動作対象, 動作のカテゴリーを格カテゴリーとしてわけ, その格カテゴリーごとに検索文の単語が動画説明文の単語に含まれている度合いを適合率 P , 動画説明文の単語が検索文の単語に含まれている度合いを再現率 R として計算する. 次に各カテゴリーの存在していない部分を補正した後に調和平均 F 値を用いて, 1 動作ごとに分かれた分の類似度を求める. また, 複文や複数文のように動作が複数ある時, 比較対象の動作の数が等しいときは, 相加平均を用いる. 相違する時は動作の数が多いうほうから, 少ないほう

の数の組み合わせを取り出し動作の数が等しい類似度の計算し, また組み合わせの中の最大のものを類似度とする方法を提案する.

2 動画サーバの検索システムの概要

本研究では, データベースに保存されている動画を検索するために日本語文を用いる. 検索文が 1 つの動作からなるなら, 類似する 1 つの動画を提示する. 複数の動作からなるなら, 複数の動作からなる 1 つの動画または複数の動画を取り出し, 動画を生成して提示する.

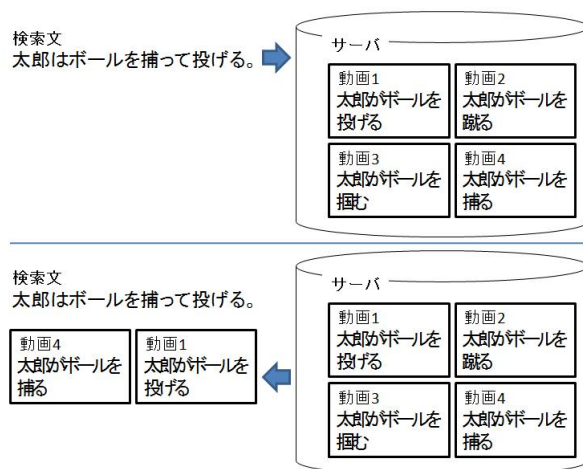


図 1: 検索概要

3 動作モデル

3.1 単一動作に対する動作モデル

アニメーションの 1 つの動作の説明は, 図 2 のように動作主, 動作対象, 動作に分けられると考えられる. そこで, 動作主, 動作対象, 動作のカテゴリーを日本語の文の助詞を手がかりにして, 次の (1) ~ (4) の 4 つの格カテゴリーに分けるものとする. カテゴリー内は, 動作主のように複数いることから「と」「や」などで並列に接続していると想定する. ただし, この段階で

は単語を修飾する形容詞や副詞や格カテゴリー内での助詞による係り受けはないものとする。

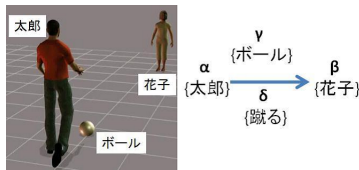


図 2: アニメーションと動画説明文のモデル化

- (1) 動作主 α : 格の「が」、係助詞の「は」に接続する。
 $\alpha = \{a_1, a_2, \dots, a_i \in \text{名詞}\}$.
- (2) 動作対象₁ : 二格の「に」、デ格の「で」、へ格の「へ」に接続する。
 $(= \{b_1, b_2, \dots, b_j \in \text{名詞}\})$.
- (3) 動作対象₂ γ : ヲ格の「を」に接続する。
 $\gamma = \{c_1, c_2, \dots, c_k \in \text{名詞}\}$.
- (4) 動作 δ : 文末の動詞。
 $\delta = \{d_1\}, d_1 \in \text{名詞または動詞}\}$.

次に例文 S_1 から例文 S_3 のように、動作主、動作対象、動作の格カテゴリー α, γ, δ の要素は、1 つだけでなく複数個存在する可能性がある。

例文 S_1 : 太郎が椅子に座る。

例文 S_2 : 太郎は花子にボールを投げる。

例文 S_3 : 太郎と花子は一郎と次郎にボールを投げる。

例文 $S_1 \sim S_3$ を格カテゴリーごとに分けた 1 動作モデル $m_S = (\alpha_S, \gamma_S, \delta_S)$ で表わすと次のようになる。

$$m_{S_1} = (\{\text{太郎}\}, \{\text{椅子}\}, \{\text{座る}\})$$

$$m_{S_2} = (\{\text{太郎}\}, \{\text{花子}\}, \{\text{ボール}\}, \{\text{投げる}\})$$

$$m_{S_3} = (\{\text{太郎}, \text{花子}\}, \{\text{一郎}, \text{次郎}\}, \{\text{ボール}\}, \{\text{投げる}\})$$

3.2 複数動作に対する動作モデル

アニメーションの動作は複数の動作からなることがある。そこで本研究では、1 動作の動作モデルを拡張する。

(1) 複文の場合

述部が複数ある複文では動作を複数表す。複文の場合は拡張の仕方は動作ごとに 1 動作モデルを当てはめ、動作主、動作対象を継承する。複数動作の動作モデルを以下のように表すことにする。

$$M_S = \{m_{s,1}, m_{s,2}, \dots, m_{s,n}\} = \{(\alpha_{S,1}, \gamma_{S,1}, \delta_{S,1}), (\alpha_{S,2}, \gamma_{S,2}, \delta_{S,2}), \dots\}$$

例文 S_4 : 太郎はボールを蹴って受けた。

例文 S_4 の場合は、「蹴る」と「受けた」の 2 つの動作

があるので 2 つの 1 動作モデルがあるとし、主格の「太郎」と動作対象の「ボール」を継承する。例文 S_4 に対する動作モデルは、次のようになる。

$$M_{S_4} = \{(\{\text{太郎}\}, \{\text{ボール}\}, \{\text{蹴る}\}), (\{\text{太郎}\}, \{\text{ボール}\}, \{\text{受ける}\})\}$$

(2) 複数の文からなる場合

動作の説明が複数の文からなる場合、それぞれの単文の中での動作モデルを上記の (1) の方法で動作主と動作対象を継承した動作モデルに分割する。文間については、動作主がある場合は継承せず (例文 S_5, M_{S_5} , 図 3)、動作主がない場合は前の文の動作主を継承する (例文 S_6, M_{S_6} , 図 4)。

例文 S_5 : 太郎はボールを花子に蹴った。花子は走った。

$$M_{S_5} = \{(\{\text{太郎}\}, \{\text{花子}\}, \{\text{ボール}\}, \{\text{蹴る}\}), (\{\text{花子}\}, \{\text{走る}\})\}$$

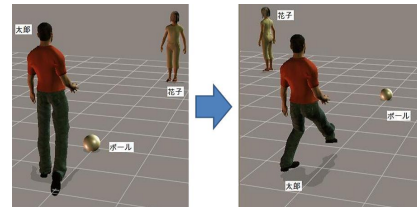


図 3: 例文 S_5 のアニメーションの一部

例文 S_6 : 太郎はボールを花子に蹴った。そして走った。

$$M_{S_6} = \{(\{\text{太郎}\}, \{\text{花子}\}, \{\text{ボール}\}, \{\text{蹴る}\}), (\{\text{太郎}\}, \{\text{走る}\})\}$$

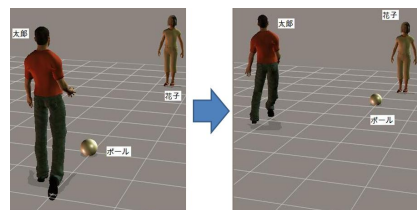


図 4: 例文 S_6 のアニメーションの一部

4 単文間の類時度 [4]

動画説明文 T と検索文 S の動作モデル $m_T = (\alpha_T, \gamma_T, \delta_T)$, $m_S = (\alpha_S, \gamma_S, \delta_S)$ で表すことにすると、動作主の格カテゴリーの適合率 P と再現率 R を次に定義する。

$$P_{\alpha}(T, S) = \frac{\alpha_S \text{と} \alpha_T \text{の一致している要素数}}{|\alpha_T|}$$

$$R_{\alpha}(T, S) = \frac{\alpha_S \text{と} \alpha_T \text{の一致している要素数}}{|\alpha_S|}$$

なお, $|\alpha_S|, |\alpha_T|$ は α_S, α_T の要素数.

また, 他の格カテゴリーの適合率 $P_{\beta}(T, S), P_{\gamma}(T, S), P_{\delta}(T, S)$ も $P_{\alpha}(T, S)$ と同様に, 再現率 $R_{\beta}(T, S), R_{\gamma}(T, S), R_{\delta}(T, S)$ も $R_{\alpha}(T, S)$ と同様に定義する.

次に検索文 S と動画説明文 T の格カテゴリーが空でない数を用いて, T に対する S の文の類似度 $C_P(T, S)$, S に対する T の文の類似度 $C_R(T, S)$ を次のように定義する.

$$C_P(T, S) = \frac{\sum_{i \in \{\alpha, \beta, \gamma, \delta\}} P_i(T, S)}{T \text{ の空でない格カテゴリー数}}$$

$$C_R(T, S) = \frac{\sum_{i \in \{\alpha, \beta, \gamma, \delta\}} R_i(T, S)}{S \text{ の空でない格カテゴリー数}}$$

格カテゴリーごとの類似度に対して, 単文間の類似具合を類似度 $sim_1(T, S)$ を, $C_P(T, S), C_R(T, S)$ の調和平均を用いて次のように定義する. なお, 類似度 $sim_1(T, S)$ は動作モデル間の類似度を計算するのと同等なので, $sim_1(m_{S_7}, m_{S_8})$ と表すものとする.

$$sim_1(T, S) = \frac{2C_P(T, S) \cdot C_R(T, S)}{C_P(T, S) + C_R(T, S)}$$

例えば, 動画説明文 T を例文 S_7 , 検索文 S を例文 S_8 とした類似度 $sim_1(S_7, S_8)$ は, まず適合率 $P_{\alpha}(S_7, S_8) = 1, P_{\beta}(S_7, S_8) = 0, P_{\gamma}(S_7, S_8) = 1, P_{\delta}(S_7, S_8) = 1$, 再現率 $R_{\alpha}(S_7, S_8) = 1, R_{\beta}(S_7, S_8) = 0, R_{\gamma}(S_7, S_8) = 1, R_{\delta}(S_7, S_8) = 1$, となり, 例文 S_7 の動作対象 が空なので S_5 の空でない格カテゴリー数は 3, 例文 S_8 は空の格カテゴリーがないので, 格カテゴリー数は 4 となる.

例文 S_7 : 太郎がボールを蹴る.

例文 S_8 : 太郎が花子にボールを蹴る.

$$C_P(S_7, S_8) = \frac{1 + 0 + 1 + 1}{3} = 1$$

$$C_R(S_7, S_8) = \frac{1 + 0 + 1 + 1}{4} = \frac{3}{4}$$

$$\begin{aligned} sim_1(S_7, S_8) &= sim_1(m_{S_7}, m_{S_8}) \\ &= \frac{2 \cdot 1 \cdot \frac{3}{4}}{1 + \frac{3}{4}} = \frac{6}{7} \end{aligned}$$

5 複数動作モデル間の類似度

動画説明文 T は述語が 1 つしかない単文, 複文, 複数文がある可能性がある. また, 検索文 S は述語が 1 つしかない単文または複文を想定している. そこで動

画説明文 T , 検索文 S に対する動作モデルは 1 動作であっても 3.2 節の複数動作のモデルを用い以下のように表す.

$$M_T = (m_{T,1}, m_{T,2}, \dots)$$

$$M_S = (m_{S,1}, m_{S,2}, \dots)$$

本研究の検索は, 検索文から類似度の高い動画説明文を探すことになるが, 結果は 1 つの動画の説明文だけではなく複数あってよい. そこで, 検索文の動作モデルの数と説明文の動作モデルの数によって類似度の計算方式を sim_2, sim_3, sim_4 の 3 つに分けて求める.

(1) 動画説明文の動作モデル数 $m =$ 検索文 S の動作モデル数 n の場合

検索文と動画説明文の 1 動作モデルの出現順ごとに 1 動作モデルの類似度 sim_1 を用いて, 類似度 sim_2 を計算する.

$$\begin{aligned} sim_2(T_1 T_2 \dots, S) \\ &= sim_2((m_{T,1}, m_{T,2}, \dots, m_{T,m}), (m_{S,1}, m_{S,2}, \dots, m_{S,n})) \\ &= \{sim_1(m_{S,1}, m_{T,1}) + sim_1(m_{S,2}, m_{T,2}) + \dots + sim_1(m_{S,n}, m_{T,n})\} / n \end{aligned}$$

(2) 動画説明文の動作モデル数 $m >$ 検索文 S の動作モデル数 n の場合

検索文 S の動作モデル $(m_{S,1}, m_{S,2}, \dots, m_{S,n})$ のほうが, 動画説明文の動作モデル $(m_{T,1}, m_{T,2}, \dots, m_{T,m})$ の数より少ないので, $(m_{T,1}, m_{T,2}, \dots, m_{T,m})$ の中から重ならない n 個を選び, 検索文の動作モデル n 個との間に (1) の類似度 sim_2 を計算する. 選んだ動作モデルの組み合わせは ${}_m C_n$ 個あり, その sim_2 の最大値を計算し, その値に $\frac{n}{m}$ を乗じたものを類似度 sim_3 とする.

$$\begin{aligned} sim_3(T_1 T_2 \dots, S) \\ &= sim_3((m_{T,1}, m_{T,2}, \dots, m_{T,m}), (m_{S,1}, m_{S,2}, \dots, m_{S,n})) \\ &= \frac{n}{m} \cdot \max(\\ &\quad sim_2((m_{T,1}^1, m_{T,2}^1, \dots, m_{T,n}^1), (m_{S,1}, m_{S,2}, \dots, m_{S,n})), \\ &\quad sim_2((m_{T,1}^2, m_{T,2}^2, \dots, m_{T,n}^2), (m_{S,1}, m_{S,2}, \dots, m_{S,n})), \\ &\quad \dots) \end{aligned}$$

(3) 動画説明文の動作モデル数 $m <$ 検索文 S の動作モデル数 n の場合

検索文 S の動作モデル $(m_{S,1}, m_{S,2}, \dots, m_{S,n})$ のほうが, 動画説明文の動作モデル $(m_{T,1}, m_{T,2}, \dots, m_{T,m})$ の数より多いので, $(m_{S,1}, m_{S,2}, \dots, m_{S,n})$ の中から重ならない m 個を選び, 動画説明の動作モデル m 個との間に (1) の類似度 sim_2 を計算する. 選んだ動作モデルの組み合わせは ${}_n C_m$ 個あり, その sim_2 の最大値を計算し, その値に $\frac{m}{n}$ を乗じたものを類似度 sim_4 とする.

$$\begin{aligned} sim_4(T_1 T_2 \dots, S) \\ &= sim_4((m_{T,1}, m_{T,2}, \dots, m_{T,m}), (m_{S,1}, m_{S,2}, \dots, m_{S,n})) \end{aligned}$$

$$= \frac{m}{n} \cdot \max(\begin{aligned} &sim_2((m_{T,1}, m_{T,2}, \dots, m_{T,m}), (m_{S,1}^1, m_{S,2}^1, \dots, m_{S,m}^1)), \\ &sim_2((m_{T,1}, m_{T,2}, \dots, m_{T,m}), (m_{S,1}^2, m_{S,2}^2, \dots, m_{S,m}^2)), \\ &\dots \end{aligned})$$

それぞれの場合分けごとの類似度 sim_2, sim_3, sim_4 をまとめて複数動作モデル間の類似度を sim_5 とする。

5.1 複数動作モデル間の類似度の計算例

例文として次の例文 $S_9 \sim S_{13}$ を組み合わせて類似度の計算例を示す。

例文 S_9 : 太郎が花子にボールを蹴って走った。

例文 S_{10} : 太郎が花子にボールを蹴った。そして走って、止まった。

例文 S_{11} : 太郎が走った。

例文 S_{12} : 花子が走った。

例文 S_{13} : 太郎が止まった。

検索文を S_9 とすると、その動作モデルは $M_{S_9} = \{ (\{ \text{太郎} \}, \{ \text{花子} \}, \{ \text{ボール} \}, \{ \text{蹴る} \}), (\{ \text{太郎} \}, \{ \}, \{ \text{走る} \}), (\{ \text{ボール} \}, \{ \text{走る} \}) \}$ となる。

また、動画の説明文を 1 個または複数並べたものを (A) S_{10} , (B) $S_7 S_{11}$, (C) $S_7 S_{12}$, (D) $S_7 S_{11} S_{13}$, (E) S_7 とすると、動作モデルは次のようになる

(A) $M_{S_{10}} = \{ (\{ \text{太郎} \}, \{ \text{花子} \}, \{ \text{ボール} \}, \{ \text{蹴る} \}), (\{ \text{太郎} \}, \{ \}, \{ \}, \{ \text{走る} \}), (\{ \text{太郎} \}, \{ \}, \{ \}, \{ \text{止まる} \}) \}$,

(B) $M_{S_7 S_{11}} = \{ (\{ \text{太郎} \}, \{ \text{花子} \}, \{ \text{ボール} \}, \{ \text{蹴る} \}), (\{ \text{太郎} \}, \{ \}, \{ \}, \{ \text{走る} \}) \}$,

(C) $M_{S_7 S_{12}} = \{ (\{ \text{太郎} \}, \{ \text{花子} \}, \{ \text{ボール} \}, \{ \text{蹴る} \}), (\{ \text{花子} \}, \{ \}, \{ \}, \{ \text{走る} \}) \}$,

(D) $M_{S_7 S_{11} S_{12}} = \{ (\{ \text{太郎} \}, \{ \text{花子} \}, \{ \text{ボール} \}, \{ \text{蹴る} \}), (\{ \text{太郎} \}, \{ \}, \{ \}, \{ \text{走る} \}), (\{ \text{太郎} \}, \{ \}, \{ \}, \{ \text{止まる} \}) \}$,

(E) $M_{S_7} = \{ (\{ \text{太郎} \}, \{ \text{花子} \}, \{ \text{ボール} \}, \{ \text{蹴る} \}) \}$

以上の動作モデルに対して類似度 sim_5 の計算例を次に示す (表 1)。

$$(A) \quad sim_5(S_{10}, S_9) = sim_3(S_{10}, S_9)$$

表 1: 検索文 S =例文 S_9 に対する複数動作モデルの類似度

	(A)	(B)	(C)	(D)	(E)
T	S_{10}	$S_7 S_{11}$	$S_7 S_{12}$	$S_7 S_{11} S_{13}$	S_7
sim_5	$\frac{9}{15}$	$\frac{9}{10}$	$\frac{7}{10}$	$\frac{9}{15}$	$\frac{1}{2}$
順位	3	1	2	3	5

$$= \frac{2}{3} \cdot \max(\frac{1 + \frac{4}{5}}{2}, \frac{1 + 0}{2}, \frac{1 + 0}{2}) = \frac{9}{15}$$

$$(B) \quad sim_5(S_7 S_{11}, S_9) = sim_2(S_7 S_{11}, S_9)$$

$$= \frac{1 + \frac{4}{5}}{2} = \frac{9}{10}$$

$$(C) \quad sim_5(S_7 S_{12}, S_9) = sim_2(S_7 S_{12}, S_9)$$

$$= \frac{1 + \frac{2}{5}}{2} = \frac{7}{10}$$

$$(D) \quad sim_5(S_7 S_{11} S_{13}, S_9) = sim_3(S_7 S_{11} S_{13}, S_9)$$

$$= \frac{2}{3} \cdot \max(\frac{1 + \frac{4}{5}}{2}, \frac{1 + 0}{2}, \frac{1 + 0}{2}) = \frac{9}{15}$$

$$(E) \quad sim_5(S_7, S_9) = sim_4(S_7, S_9)$$

$$= \frac{1}{2} \cdot \max(1, \frac{4}{7}) = \frac{1}{2}$$

(A) と (D) は動画説明文の動作モデルの数が多い場合、(B) と (C) は双方の動作モデルの数が多い場合である。このうち (A) と (D) は同等で、1 つの動画への説明が (A) の S_{10} 、3 つの動画への説明が (D) の $S_7 S_{11} S_{13}$ である。

6 おわりに

本研究では、検索文と動画説明文の文間の類似計算を複数動作がある文に拡張する手法を提案した。提案手法では、複数動作がある文を動作ごとに分け、主格、動作対象を継承させた上で各文を動作モデルに当てはめ、類似度の計算をした。また、動作の数が複数ある場合、それらの間の類似度を計算する手法を提案した。今後はより複雑な文に対応する必要がある。

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] 奥村, 難波, テキスト自動要約, オーム社, 2005.
- [4] 白井, 椎名, 説明文を付したアニメーションに対する文間の格構造ごとの類似度と格構造間の類似度を用いた検索, WebDB Forum2008, B5-2, pp1-7, 2008.