

形式概念分析を用いた推薦理由を明示する 動画推薦手法

滝沢 光介

富山県立大学 電子情報工学科

May 14, 2021

YouTube の現状

近年, YouTube などの動画閲覧サイトには大量の動画が投稿されている。その中から自分が見たい動画を効率よく探す方法として動画推薦システムが適用されている。しかし, YouTube では、動画の推薦理由が明示されておらず、ユーザーがタイトルやサムネイルから動画内容を推測しなければならない。また, YouTube では同じような観点からの動画ばかりが推薦され、推薦の偏りが生じている。

本研究の目的

ユーザーが動画内容を推測する労力を減らつつ、推薦の偏りを解消するために、動画に付与されたタグ情報を利用した形式概念分析に基づく明示的な動画推薦システムを提案する。

形式概念分析について

形式概念分析とは

オブジェクトとそれらが持つ属性の間で二項関係を定義し、オブジェクト同士の潜在的な関係を発見する手法。

	スポーツ	決勝	高校
動画 1	×	×	
動画 2	×	×	
動画 3			×
動画 4		×	×
動画 5	×		

Table 1: コンテクスト表

形式概念分析の定義(1)

コンテクスト

オブジェクト集合 G と属性集合 M が二項関係 $I \subseteq G \times M$ で表せる集合をコンテクスト K

$$\text{コンテクスト } K = (G, M, I)$$

オブジェクト g が属性 m を保有している時, gIm と表す.

$$g \subseteq G, m \subseteq M$$

コンセプト

$A \subseteq G$ と $B \subseteq M$ について A' と B' を定義する.

$$A' = \{m \subseteq M \mid gIm \text{ for all } g \subseteq A\}$$
$$B' = \{g \subseteq G \mid gIm \text{ for all } m \subseteq B\}$$

このとき, $A \subseteq G, B \subseteq M$ かつ $A' = B, B' = A$ となる (A, B) の組をコンテキスト K のコンセプトと呼ぶ.

外延と内包

コンセプト (A, B) の A を外延, B を内包と呼び, 全てのコンセプトの集合を $\mathfrak{P}(G, M, I)$ とし, 外延と内包の集合をそれぞれ $\mathfrak{P}_G, \mathfrak{P}_M$ とする.

$$\mathfrak{P}_G = \{A \subseteq G \mid A'' = A\}$$
$$\mathfrak{P}_M = \{B \subseteq M \mid B'' = B\}$$

形式概念分析の定義(3)

コンセプトランク

コンテキスト表から得られたコンセプトに階層構造を持たせるために以下のよう順序関係を適用する。得られたコンセプトの束をコンセプトランクと呼ぶ。

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_1 \supseteq B_2$$

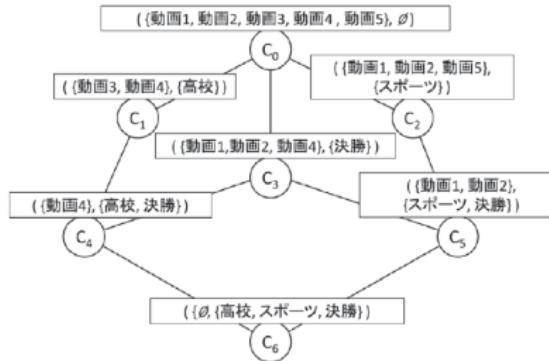


図 1: コンセプトランク

提案システムの流れ

step1: コンテキスト表作成

ユーザーが視聴した動画を基準とし、タグをたどって動画を獲得していく、コンテキスト表を作成する。

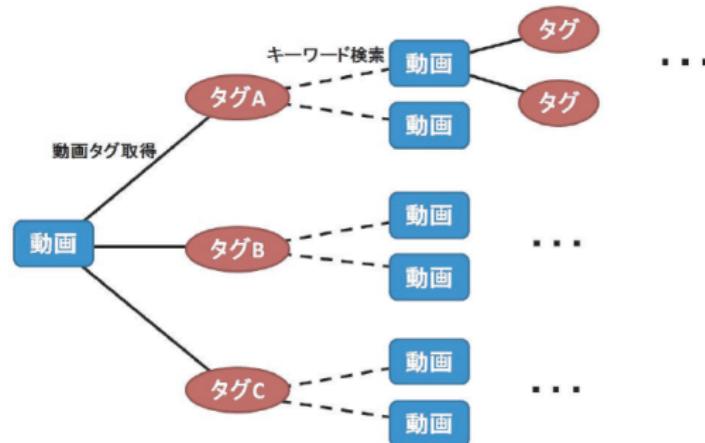


図 2: コンテキスト表作成イメージ

提案システムの流れ(2)

step2:コンセプトライティス作成

step1 で得られたコンテクスト表から形式概念分析によりコンセプトを決定し, それらからコンセプトライティスを作成する. コンセプトの決定には *S.Andrews* の *In – Close* アルゴリズムを用いる.

step3:推薦理由・関連度決定

step2 で得られたコンセプトライティスを利用して推薦理由を決定する. 推薦理由決定の際に, 以下の 2 パターンの状況に応じた処理を行う.

- ・コンセプトの内包の共通要素が存在する場合
⇒ 共通しているタグを推薦理由とする.
- ・コンセプトの内包の共通要素が存在しない場合
⇒ 基点コンセプトと対象コンセプトの内包要素が共存するコンセプトを探索し, その共通要素を推薦理由とする.

提案システムの流れ(3)

```

Require:  $\mathfrak{P}(G, M, I)$ 
Ensure:  $RC = \{RC_0, RC_1, \dots, RC_n\}$ 
 $U = \mathfrak{P}(G, M, I) = \{c_0, c_1, c_2, \dots, c_n\}$ 
 $U \leftarrow U \setminus \min(U)$ 
 $Rel = \{Rel_0, Rel_1, \dots, Rel_n\}$ 
if  $A_{\max(U)} = \phi$  then
     $U \leftarrow U \setminus \max(U)$ 
end if
 $Cov = \{c\}$ 
while  $Cov \neq \phi$  and  $U \neq \phi$  do
    if  $\uparrow Cov_0 \cap U \neq \phi$  then
        if  $B_{Cov_0} \cap B_c \neq \phi$  then
             $RC_{\uparrow Cov_0 \cap U} \leftarrow [B_{Cov_0} \cap B_c]$ 
             $Rel_{\uparrow Cov_0 \cap U} \leftarrow |B_{Cov_0} \cap B_c|$ 
        else
             $NC \leftarrow \bar{U} \cap (lc(Cov_0) \cup uc(Cov_0))$ 
             $RC_{\uparrow Cov_0 \cap U} \leftarrow [RC_{NC} \rightarrow B_{NC} \cap B_{Cov_0}]$ 
             $Rel_{\uparrow Cov_0 \cap U} \leftarrow Rel_{NC} - 1$ 
        end if
    end if
     $Cov \leftarrow Cov \cup ((lc(Cov_0) \cup uc(Cov_0)) \cap R) \setminus Cov_0$ 
     $U \leftarrow U \setminus \{\uparrow Cov_0 \cap U\}$ 
end while
for  $i \leftarrow 0$  to  $n$  do
     $Rel_{c_i} \leftarrow Rel_{c_i} + \alpha$ 
end for

```

図 3: 推薦理由決定アルゴリズム

提案システムの流れ(4)

アルゴリズムの変数

Rel:基点コンセプトと推薦コンセプトの関連度

RC:コンセプトの推薦理由集合

Cov:コンセプトを格納する変数.*Cov*₀ は基点コンセプト

NC:推薦理由が付与されておらず,かつ *Cov*₀ と被覆関係にある
コンセプト集合

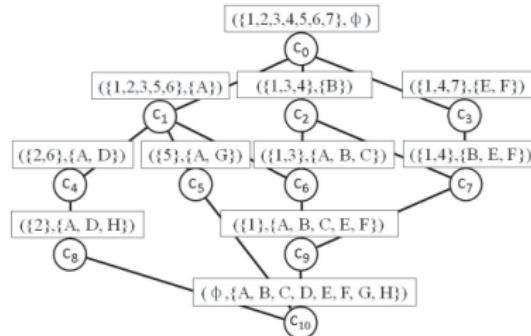


図 4: アルゴリズムを適用するコンセプトランク

提案システムの流れ (5)

11/1

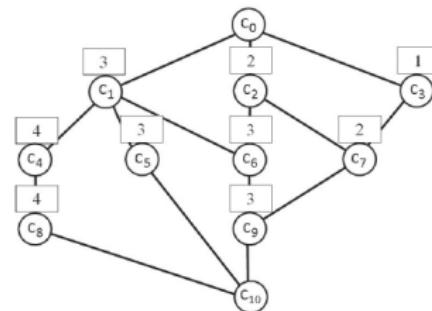


図 5: アルゴリズム適用後の Rel

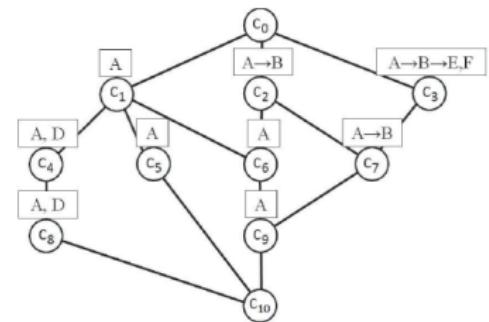


図 6: アルゴリズム適用後の RC

step4:推薦動画決定

推薦する動画を決定するにあたり, 推薦するコンセプトを決定する.
推薦するコンセプトは Rel の値に基づいて決定される. コンセプト c_i が選択される確率 p_i は, 以下のようになる.

$$pc_i = \frac{Rel_{c_i}}{\sum_{k=0} n Rel_{c_k}}$$

次にコンセプトの外延から動画を選択する. このとき外延に複数の動画が存在する場合は, 再生回数や評価が高い動画を選択するようになる.

提案システムの流れ(7)

step5:推薦動画提示

推薦する動画を決定したら視覚的にユーザーに提示する。画面の中心に視聴動画を置き、その周りに推薦動画を配置する。

$$d = a \log \frac{b}{Rel}$$

中心の動画からの距離 d は上記の式で決まり、距離が近いほど関連が高い動画と言える。

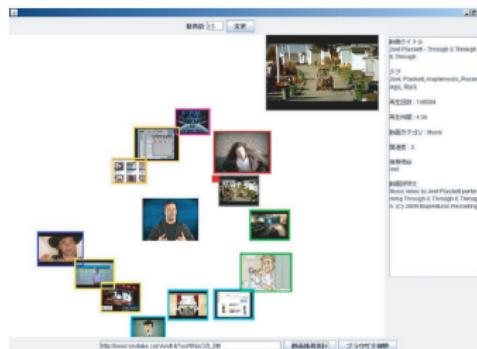


図 7: 推薦動画表示画面

評価実験 1-1

実験 1-1：平均動画選択数の比較

YouTube と提案手法の推薦動画をそれぞれ 5 つ提示し、自分がみたいと思う動画を選んでもらい平均動画数を算出する。その試行を 20 回繰り返す。

評価項目	YouTube	提案手法
平均動画選択数	2.279	1.426

Table 2: 実験 1-1 結果

評価実験 1-2

実験 1-2：適合率の比較

YouTube と提案手法の推薦動画それぞれを 20 個提示し, その動画を最後まで見たいかどうかを選んでもらい適合率を算出する. その試行を 20 回繰り返す.

評価項目	YouTube	提案手法
適合率	0.625	0.720

Table 3: 実験 1-2 結果

アンケート調査

16/1

アンケート調査

- ① 両システムにおいて、推薦動画の概要把握はどの程度できたか。
- ② 両システムにおいて、好みの動画を容易に探索できたか。
- ③ 提案システムで提示される推薦理由にどの程度納得できたか。
- ④ 提案システム利用時に、どの動画情報に注目していたのか。

	YouTube	提案手法
質問 1	3.3	3.6
質問 2	3.6	3.2
質問 3	-	3.4

図 8: 質問 1 から 3 の結果

動画情報	人数(人)
タイトル	10
推薦理由	7
閲速度	5
タグ	5
説明文	5
再生時間	4
再生回数	2
カテゴリ	2

図 9: 質問 4 の結果

提案システムの問題

提案システムでは、様々な関連度を持った動画を推薦したり、一度タグに英語が含まれるとそこから海外の動画が多く推薦されるようになり、視聴者が本来求める動画と少し異なったものが推薦されていた。

解決策

あるタグよりも上位に属するタグを親タグと呼び、これを使って動画推薦を行う。

親タグの定義

今回、親タグを取得する手法として IR (IntersectionRate) 値を用いる。

$$IR_{ij} = \frac{n(t_i t_j)}{n(t_i)}$$

$n(t_i)$: YouTube 内のタグ t_i にのキーワード検索数

改良システム

改良手法では提案手法の 5 つのステップの一番最初に親タグを取得するステップを追加する。

親タグの取得

- ① 視聴動画からタグ t_i を取得
- ② t_i のキーワード検索から上位 1000 件の動画を取得
- ③ 2 の動画に出現する動画のタグの中で t_i と共に頻度が高いタグ上位 n 個を取得
- ④ t_i と 3 のタグの IR 値を計算し、条件を満たすタグを親タグとする
- ⑤ 1 から 4 を全ての視聴動画タグに適用

提案システムでの問題点

提案手法では、視聴動画と無関係なノイズタグが原因でコンセプトラティスの巨大化が起こっていた。そこで改良手法では、まれにしか出現しないタグをノイズタグとして除去する。

ノイズタグ除去の流れ

- ① 視聴動画と親タグを用いてキーワード検索を行い、それぞれのタグを持つ動画を取得
- ② 1で取得したタグの出現回数を算出し、言っていいかであればそのタグを削除
- ③ タグを持たない動画 URL を削除

実験 2-1: 平均動画選択数

同一の視聴動画における YouTube, 提案手法, 改良手法それぞれの推薦動画を 5 つ提示し, 視聴した動画を選択してもらう. この試行を 20 回繰り返す.

	平均動画選択数
YouTube	3.411
提案手法	1.975
改良手法	2.730

Table 4: 実験 2-1 結果

実験 2-2:推薦理由の評価

- ① なぜこの動画が提示されたのかを理解できた
- ② この動画の視聴可否を決定する補助となったか

の二つの質問をユーザーに 5 満点で応えてもらいその平均を算出する。

評価項目	提案手法	改良手法
質問 1	3.475	3.912
質問 2	0.192	0.146

Table 5: 実験 2-2 結果

実装したこと

- ① 推薦動画が推薦された理由を明示的に表現
- ② 推荐情報を明示することで、実際の動画の内容と推測した内容の差異を小さくした
- ③ 多様なコンセプトから推薦を行うことで推薦の偏りを無くした
これらを実装した動画推薦システムを実装した。

今後の課題

課題としては推薦時間の短縮が挙げられる。