

原著論文

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†

北村 祐太郎 *1・澤勢 一史 *2・延原 肇 *3

動画共有サイトYouTubeの動画推薦手法における、推薦の偏り、視聴動画と推薦動画間の関連性が不明瞭、および推薦画面内の情報提示量の少なさを、解決する手法を提案する。具体的には、YouTubeの動画に付与されたタグの概念が持つ広狭に着目し、これらに対して形式概念分析を適用することで、さまざまな概念階層から動画を推薦することで、推薦の偏りを解消する。また、形式概念分析によって得られる束構造を利用することで動画の推薦理由を構成し、視聴動画と推薦動画間の関係性を明示する。さらに適切な動画情報量を持つYouTubeの動画推薦アプリケーションを構築し、ユーザの選択における負担を軽減する。提案システムの有用性を確認するため、20代の男女10人のユーザに対して、YouTubeと提案手法のシステムにおける平均動画選択数や適合率、推薦理由の有用性を調査する。さらにアンケート調査により、推薦動画を提示する際の適切な情報量について考察する。

キーワード：形式概念分析、情報推薦、動画推薦、情報爆発、グラニューラーコンピューティング

1. はじめに

近年、インターネットの急激な普及により、我々が触れることのできる情報は爆発的に増大している[1]。利用者が動画を投稿する動画共有サイトでは、毎日多くの動画が投稿されており、世界中のユーザが多種多様な動画を視聴している。動画共有サイトであるYouTubeには、2011年の時点で1分間に約48時間分の動画が投稿され、1日あたり30億回動画が再生されている[2, 7]。このような巨大なサイトでは、ユーザが目的の動画を短時間で探し出すことは困難である。

適切な動画を発見する手段として情報推薦があり[3, 5, 6, 8, 9]、YouTubeをはじめとする各種動画サイトでも導入されている。しかしYouTubeでは、推薦の根拠となる情報が明示されておらず、ユーザが推薦動画の内容を推測しながら次の視聴動画を選択しなければならず、負担が生じる。また、同じような観点から動画が多く推薦される「推薦の偏り」が発生してしまう。

本研究では、ユーザが動画選択を行う労力を軽減するとともに、推薦の偏りを解消するために、動画に付与されているタグを用いた、形式概念分析[4]に基づく明示的な動画推薦システムを提案する。形式概念分析を利用することで、各動画に付与されているタグの集合の包含関係に着目し、動画間の概念階層構造を明らかにすることが可能になる。さらに提案システムでは、形式概念分析から得られた概念構造を用い、推薦動画に対して明示的な推薦理由を付与するアルゴリズムを実装することで、動画選択の労力を軽減できる。さらに、様々な概念階層から動画が選択可能になり、多視点からの関連を考慮した動画を提示可能となる。

提案手法の有効性を確認するためにシステムを構築し、評価実験やアンケートに基づいて従来システムであるYouTubeとの比較を行う。その結果を分析し、ユーザの負担軽減についての有効性を示す。

2. 関連研究

YouTubeでは、大量に投稿されている動画の中から、ユーザに対して推薦が行われているが、次のような拡張の余地が残されている。

- 視聴動画と推薦動画間の関係性が非明示的
- 動画情報が限定的
- 推薦の偏りが発生

1点目は、現在のYouTubeでは視聴動画と推薦動画間の関係性が明示されていないことである。YouTubeの推薦動画はリスト型であり、縦一列に羅

† Formal Concept Analysis Based Movie Recommendation
Yutaro KITAMURA, Kazuhito SAWASE and Hajime NOBUHARA

*1 中部電力株式会社

Chubu Electric Power Co., Inc.

*2 筑波大学 大学院システム情報工学研究科 知能機能システム専攻
Department of Intelligent Interaction Technologies, Graduate School of Systems and Information Engineering, University of Tsukuba

*3 筑波大学 システム情報系 知能機能工学域
Faculty of Engineering, Information and Systems, University of Tsukuba

列して表示している．しかしユーザは、なぜその順番で動画を羅列しているのかを知ることができない．動画の推薦理由が不明な場合、ユーザは自ら動画間の関係性を推測する必要があり、ユーザへの負担が大きくなる．

2点目は、YouTubeの推薦画面での情報が限定的ということである．推薦画面から得られる情報は、サムネイル・タイトル・投稿ユーザ・再生回数および再生時間のみであり、視聴時に確認できる情報の一部しか表示していない．そのため、ユーザが推薦画面で推測した動画内容と、視聴した動画内容の間に差異が発生する可能性が高い．

3点目は推薦の偏りである．YouTubeの動画推薦は、視聴動画と関連度の高い動画を優先的に提示していると推測される．しかしユーザは、必ずしも高い関連度の動画ばかりを選択するとは限らない．

本研究では、推薦動画と視聴動画間の関係性を明示でき、動画概要を容易に把握できるシステムを目指す．

3. 提案システム

3.1 形式概念分析

形式概念分析とは、対象(オブジェクト)とそれらが持つ性質(属性)との間で二項関係を定義し、対象同士の潜在的な関係を発見する手法のことである[4]．コンテキスト $K=(G, M, I)$ とは、オブジェクト集合 G と属性集合 M が二項関係 $I \subseteq G \times M$ で表されている組の集合であり、どのオブジェクトがどの属性を保有しているかを表している．オブジェクト $g \in G$ が属性 $m \in M$ を保有している場合、 gIm と表す．表1にコンテキスト表の例を示す．オブジェクトが属性を保有していれば×を付与する．

例えば表1において、オブジェクト $G=\{\text{動画1, 動画2, 動画3, 動画4, 動画5}\}$ であり、属性 $M=\{\text{スポーツ, 決勝, 高校}\}$ である．また、オブジェクトの部分集合 $\{\text{動画1}\}$ は属性 $\{\text{スポーツ, 決勝}\}$ を保有しており、 $\{\text{動画1, 動画2, 動画5}\}$ は属性 $\{\text{スポーツ}\}$ を保有している．オブジェクトの部分集合 $A \subseteq G$ と属性の

表1 コンテキスト表

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

部分集合 $B \subseteq M$ について、 A' と B' を

$$A' = \{m \in M \mid gIm \text{ for all } g \in A\}$$

$$B' = \{g \in G \mid gIm \text{ for all } m \in B\}$$

と定義する．このとき A' は、 A に含まれるオブジェクトに共通する属性集合を意味する．また B' は、 B に含まれる属性に共通するオブジェクト集合を意味する．例えば表1において、オブジェクトの部分集合 $A=\{\text{動画1, 動画2}\}$ に共通する属性集合 A' は $\{\text{スポーツ, 決勝}\}$ である

$A \subseteq G, B \subseteq M$ 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．例えば $A=\{\text{動画3, 動画4}\}$ のとき、 $A'=\{\text{高校}\}$ である．また、 $B=\{\text{高校}\}$ のとき、 $B'=\{\text{動画3, 動画4}\}$ である．従って $A'=B, B'=A$ が成立し、 (A, B) はコンセプトとなる．このときの A を外延、 B を内包と呼ぶ．すべてのコンセプトの集合を $\mathfrak{P}(G, M, I)$ と表し、外延の集合 \mathfrak{P}_G と内包の集合 \mathfrak{P}_M を、

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

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

と定義する．表1において、 $\mathfrak{P}_G=\{\emptyset, \{\text{動画4}\}, \{\text{動画1, 動画2}\}, \{\text{動画3, 動画4}\}, \{\text{動画1, 動画2, 動画4}\}, \{\text{動画1, 動画2, 動画5}\}, \{\text{動画1, 動画2, 動画3, 動画4, 動画5}\}\}$ であり、 $\mathfrak{P}_M=\{\emptyset, \{\text{高校}\}, \{\text{スポーツ}\}, \{\text{決勝}\}, \{\text{スポーツ, 決勝}\}, \{\text{決勝, 高校}\}, \{\text{スポーツ, 決勝, 高校}\}\}$ である．

コンセプトに、概念の広狭の観点からの階層構造を持たせるため、

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_1 \supseteq B_2 \quad (1)$$

という順序を導入する．例えば、表3において、 $A_1=\{\text{動画4}\}$ 、 $A_2=\{\text{動画3, 動画4}\}$ 、 $B_1=\{\text{高校, 決勝}\}$ 、 B_2

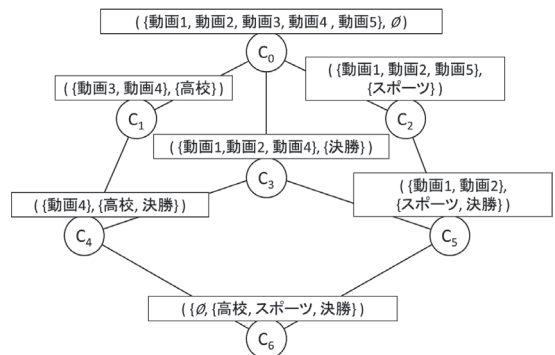


図1 コンセプトラティス

$=\{\text{高校}\}$ とする． (A_1, B_1) と (A_2, B_2) はそれぞれ概念である．このとき， $A_1 \subset A_2$ であり，かつ $B_1 \supset B_2$ となっている．したがって， $(A_1, B_1) < (A_2, B_2)$ と言える．

図1のように，すべての概念に対して順序関係についての式(1)を適用し得られる束を概念トラティスと呼ぶ．各概念の外延は，上位の概念ほど要素数が多く，下位の概念ほど要素数が少ない．また，各概念の内包は，上位の概念ほど要素数が少なく，下位の概念ほど要素数が多い．つまり，下位概念ほどタグによる動画の限定が行われているといえる．

3.2 提案システムの流れ

提案システムでは，ユーザーは視聴中の動画を基点に，タグを辿って次に閲覧する動画にたどり着くというアクションを行うことを大前提とする．このユーザーアクションを先読み予測した上で，この動画にたどり着くために必要なタグ遷移を推薦理由とする動画推薦を行うことが提案システムの本質となる．

Step 1：コンテキスト表生成

ユーザーが視聴中の動画を基点として，他の動画やタグをYouTube内のキーワード検索により取得し，コンテキスト表を生成する．図2は動画とタグを取得するイメージであり，システムは以下の操作を行う．

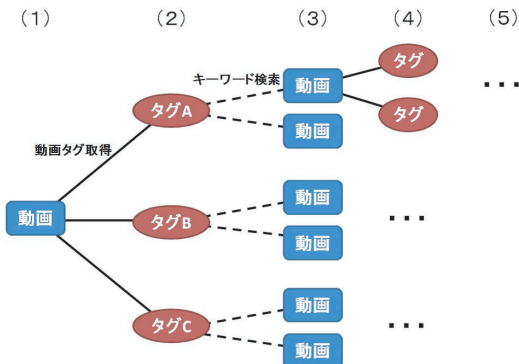


図2 コンテキスト表取得のイメージ

表2 提案手法のコンテキスト表

	宇宙	ビッグバン	科学	toys
tvQW-Yiib38	1	1	1	0
x6pzZ-IGQAg	1	0	0	1
ApVIuFE6Qbc	1	0	0	0
lk_gtC7z-NI	1	1	1	0

- (1) 視聴動画を取得
- (2) 視聴動画に付与されている視聴動画タグを取得
- (3) (2)のタグそれぞれに対して，YouTube内のキーワード検索を用い，動画を取得
- (4) (3)の動画に付与されているタグを取得
- (5) 動画を一定数取得するまで，(3)と(4)を繰り返す

ここで(3)の操作では，タグの取得順にキーワード検索を行う．例えば視聴動画タグが「A, B, C」であるとき，最初にタグAのキーワード検索結果の動画を取得し，次にタグBのキーワード検索結果の動画，最後にタグCのキーワード検索結果の動画という順で取得する．動画取得中において，あらかじめ定めた最大動画数に達した場合，動画の取得を打ち切る．

以上の操作を行ったのち，動画とタグの関係を用いてコンテキスト表を生成する．表4に動画に関するコンテキスト表の一部を示す．オブジェクト G は動画URL，属性 M はタグである．また二項関係 I は，タグが動画に付与されているという関係を表している．

Step 2：概念トラティス生成

得られたコンテキスト表から形式概念分析を用いて概念を生成し，概念間関係から概念トラティスを生成する．概念生成には，高速なS.AndrewsのIn-Closeアルゴリズム[10]を用いる．

Step 3：推薦理由・関連度決定

Step 2により得られた概念同士の間関係を利用し，推薦理由を決定する．図3に示す概念トラティスを用いて，推薦理由の決定方法を説明する．各概念の上に表示しているアルファベットは外延を，単語は内包を表す．推薦理由の決定方法は，基点概念と対象概念の内包の共通要素の有無によって，操作が異なる．ここで基点概念とは，視聴動画に付与されている全タグ集合と内包が一致する概念である．図3では c_2 を基点概念とする．対象概念とは，推薦理由を付与す

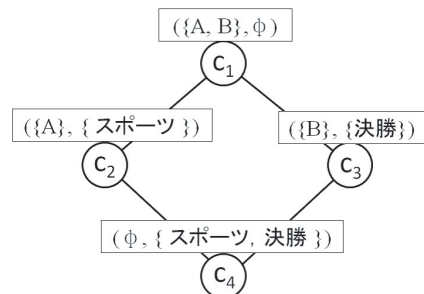


図3 共通の内包がない概念

るコンセプトである。

共通要素が存在する場合

内包の共通要素が存在する場合、その共通要素を推薦理由とする。例えば図3の c_2 から c_4 を推薦する場合を考える。 c_2 と c_4 の内包の共通要素はスポーツなので、推薦理由は「スポーツ」となる。ここで記号 $[\]$ によって囲まれた文字は、対象コンセプトにおける基点コンセプトからの推薦理由を表し、「スポーツ」とは、 c_2 の「スポーツ」タグから c_4 が推薦されたという意味である。

共通要素が存在しない場合

内包の共通要素がない場合、基点コンセプトと対象コンセプトの内包の要素が共存するコンセプトを探索する。例えば c_2 から c_3 を推薦する場合を考える。 c_2 と c_3 の内包の共通要素は存在しない。そこで先ほど推薦理由を決定した c_4 に注目すると、 c_2 の内包の要素である「スポーツ」と、 c_3 の内包の要素である「決勝」が共存している。よって推薦理由は「スポーツ→決勝」となる。これは、 c_2 の「スポーツ」タグから「決勝」タグが導出され、 c_3 が推薦されたという意味である。以上のように、推薦理由は基点コンセプトと対象コンセプトの共通要素に注目して決定する。その結果、構築したコンセプトラティスの各コンセプトに対して、推薦理由を付与することが可能となる。次にアルゴリズムの定義と変数について、図4を用いて説明する。

図4では、 c_4 を基点コンセプトとしている。各コンセプトの上に表示している数字は外延、アルファベットは内包を表す。まず、アルゴリズムの定義について説明する。任意の内包 $X \in \mathfrak{P}_M$ に対して、

$$\uparrow X = \{Y \in \mathfrak{P}_M \mid Y \supseteq X\}$$

と表す。例えば $\uparrow c_4 = \{c_4, c_8, c_{10}\}$ であり、 $\uparrow c_2 = \{c_2,$

$c_6, c_7, c_9, c_{10}\}$ である。また、 $X, Y, Z \in \mathfrak{P}_M$ において、 $X \subset Z \subset Y$ を満たす Z が存在しない場合、 $X \prec Y$ と表し、「 Y は X を被覆する」という。このとき、 X を被覆する内包の集合 $uc(X)$ および Y に被覆される内包の集合 $lc(Y)$ を、

$$uc(X) = \{Y \in \mathfrak{P}_M \mid X \prec Y\},$$

$$lc(Y) = \{X \in \mathfrak{P}_M \mid X \prec Y\}$$

と表す。例えば図4において $uc(c_7) = \{c_9\}$ 、 $lc(c_7) = \{c_2, c_3\}$ である。

次に、アルゴリズムの変数について説明する。 U の初期集合は $\mathfrak{P}(G, M, I)$ 、 Rel は基点コンセプトと推薦コンセプトの関連度を表す。 RC はコンセプトの推薦理由集合であり、属性の部分集合、または、属性の包含関係をたどることによって得られる因果関係を表す。 Rel は、基点コンセプトから近いコンセプトほど大きく、遠いコンセプトほど小さい。 Cov はコンセプトを格納し、その先頭要素を Cov_0 と表す。 Cov の初期集合には、基点コンセプト c が含まれている。 NC は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を表す。また、 $\alpha = 1 - \min(Rel)$ である。

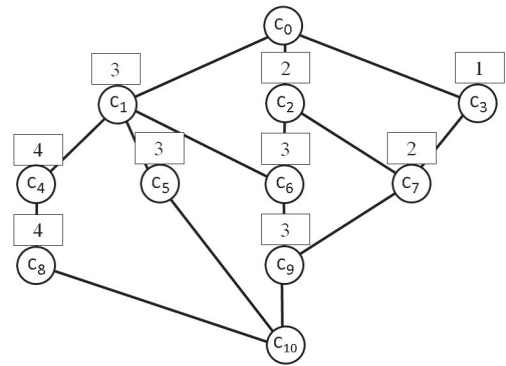


図5 アルゴリズム適用後の Rel

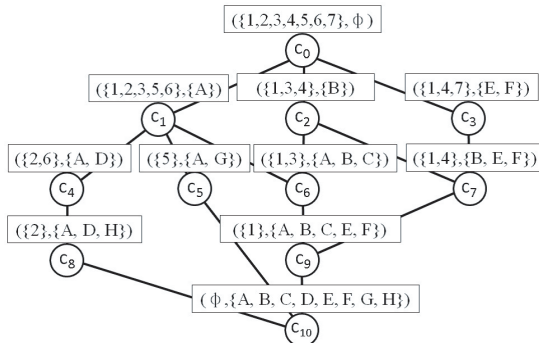


図4 アルゴリズムを適用するコンセプトラティス

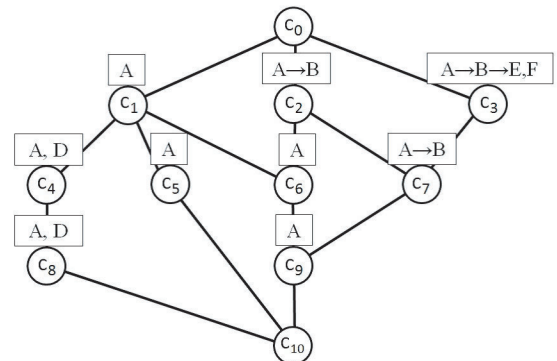


図6 アルゴリズム適用後の RC

Algorithm 7.1 推薦理由決定**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**

アルゴリズム7.1は、コンセプトに推薦理由を付与するアルゴリズムである。入力コンセプト $\mathfrak{P}(G, M, I)$ 、出力は各コンセプトへの推薦理由集合 RC であり、 RC の要素数はコンセプト数 $n+1$ である。

このアルゴリズムでは、最小元である $\min(U)$ を除去する。最小元の内包は ϕ であるため、推薦理由を決定できないからである。また、最大元 $\max(U)$ の外延 $A_{\max(U)}$ が空の場合、推薦する動画が存在しないので除去する。図4の最小元は c_0 、最大元はコンセプト c_{10} である。

アルゴリズム7.1を適用した結果、 RC は図6のように、それぞれ $RC_{c_1} = [A]$, $RC_{c_2} = [A \rightarrow B]$, $RC_{c_3} = [A \rightarrow B \rightarrow E, F]$, $RC_{c_4} = [A, D]$, $RC_{c_5} = [A]$, $RC_{c_6} = [A]$, $RC_{c_7} = [A \rightarrow B]$, $RC_{c_8} = [A, D]$, $RC_{c_9} = [A]$ となる。

図5は、図4に対してアルゴリズム7.1を適用させた後の関連度 Rel である。関連度 Rel は、最小値が0

以下の値をとらないように、整数 α を加算する。

Step 4 : 推薦動画決定

推薦動画を決定するために、推薦コンセプトを決定する。推薦コンセプトは、 Rel の値に基づいてルーレット選択を行い、決定する。このとき、コンセプト c_i が選択される確率 p_{c_i} は、

$$p_{c_i} = \frac{Rel_{c_i}}{\sum_{k=0}^n Rel_{c_k}} \quad (\in [0, 1])$$

である。表9は、図4における、推薦理由が付与されているコンセプトの Rel と p_{c_i} である。 Rel_{c_i} の値が大きい、つまり基点コンセプトから近いコンセプトが選択されやすく、結果として視聴動画に似た動画が推薦されやすい。また、基点コンセプトから遠いコンセプトが推薦される可能性も十分にあり、推薦の偏りを解消することが期待できる。

次に推薦コンセプトの各外延から1つずつ推薦動画



図7 提案手法の推薦提示画面

を選択する。外延に複数の動画が存在する場合、再生回数や評価が高い動画を選択されやすくなる。このとき、あるコンセプトにおいて、外延のすべての要素が既に推薦動画となっている場合が考えられる。その場合は代わりとして、既に推薦を行ったコンセプトから動画の選択を行う。以上の操作を行い、推薦動画を決定する。

Step 5：推薦動画提示

推薦動画を選択した後は、視覚的にユーザへ提示する。図7は作成したシステムであり、以下でその機能を説明する。

画面の中心に視聴動画、その周りに推薦動画のサムネイルを表示し、その大きさは再生回数に対応している。視聴動画と推薦動画の距離 d は、

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

で決定する。 a, b は定数であり、関連度 Rel の値が大きいほど中心に配置する。推薦動画は、表10に示される15カテゴリでまとめられている。カテゴリの違いは、動画の外枠の色によって認識できるようになっている。推薦動画のサムネイルをクリックすると、右側にタイトル、動画タグ、推薦理由、関連度などの動画情報が表示される。これらの情報を提示することにより、動画情報から推測した内容と実際の動画内容との差異を小さくする。

提案手法では様々なコンセプトから推薦を行っている。これによって推薦の偏りが解消され、多様な関連度の動画が提示される。するとユーザは多くの選択肢から動画選択を行えるようになる。また、視聴動画と推薦動画の関連度を視覚的に明示することで、ユーザが次の視聴動画を決定する際の負担を軽減することが期待できる。

4. 評価実験1

本章では、YouTubeの推薦システムと提案システムとの比較を、主観評価実験に基づき行うことで、提案システムの有効性を確認する。対象としたのは20代の男女10名であり、ここでは2つの実験とアンケート調査を行う。

4.1 実験1-1：平均動画選択数の比較

YouTubeの問題点として推薦の偏りが挙げられ、提案システムでは解決するために、形式概念分析を用いた多様な推薦を行っている。しかし多様な推薦を行うあまり、ユーザに興味のある動画が推薦に提示されない可能性がある。この実験では、両システムで提示した推薦動画のうち、視聴したい動画がどれだけ存在するのか調査する。この実験では、同一の視聴動画に対するYouTubeと提案手法の推薦動画をそれぞれ5つ提示し、視聴したい動画を選択してもらい、平均動画選択数を算出する。試行回数は20回であり、20個の視聴動画は、被験者が実験中に選択する。ユーザはYouTubeの推薦動画グループまたは提案手法の推薦動画グループに対して、それぞれ視聴したい動画を選択する。このときユーザは、どちらのグループが提示されているかを知ることができない。実験1-1における平均動画選択数は、

$$\text{平均動画選択数} = \frac{\text{視聴したい動画数}}{\text{試行回数}} \quad (\in [0, 5])$$

で表す。表3に実験結果を示す。実験1-1の結果から、平均動画選択数はYouTubeの方が多いという結果が得られた。その理由として、YouTubeでは視聴動画と関連の高い動画が推薦されることが挙げられる。提案手法では推薦の偏りを解消するために、様々な関連度を持った動画を推薦する。視聴動画を多様な観点から捉えて推薦するため、ユーザごとに選択する動画が異なる。さらに提案手法では海外の動画を推薦する割合が高い。ユーザは英語表記で提示される動画を避けるという傾向があった。これらの要素が、提案手法の平均動画選択数を低くする原因だと考えられる。

表3 実験1-1結果

評価項目	YouTube	提案手法
平均動画選択数	2.279	1.426

表4 実験1-2結果

評価項目	YouTube	提案手法
適合率	0.625	0.720

4.2 実験1-2：適合率の比較

YouTubeでは視聴動画と推薦動画間の関係性が不明であり、動画情報は限定的であるという問題を解決するために、提案システムでは新たなインタフェースを用いて、視聴動画と推薦動画間の関係性や動画情報を提示を行った。この実験では、両システムの推薦画面で興味を持った動画が、実際の視聴時に好みの動画であったかを調査する。YouTubeと提案手法の推薦動画それぞれ20個を実際に視聴し、最後まで視聴したかを判断してもらい、適合率を算出する。ユーザは推薦動画の中から、次に視聴したい動画を選択する。その動画を実際に視聴し、最後まで視聴したかどうかを判断する。これをYouTubeと提案手法のシステムでそれぞれ20回行い、適合率を算出する。実験1-2における適合率は、

$$\text{適合率} = \frac{\text{最後まで視聴したい動画数}}{\text{試行回数}} \quad (\in [0, 1])$$

で表す。

表4に実験結果を示す。実験1-2の結果より、適合率は提案手法の方が高いことが判明した。YouTubeの推薦動画では、動画情報の一部しか提示されていない。その結果、推薦動画から推測した動画内容と実際に視聴した動画内容で差異が発生したため、適合率が低くなったと考えられる。提案システムでは、視覚的に推薦動画を提示し、推薦理由や関連度を用いて明示的に推薦を行っている。ユーザは提示された推薦動画を、動画内容や推薦理由に納得してから視聴するため、YouTubeより高い適合率を得たと考えられる。また、有意水準のt検定を行った結果、有意差が認められた。

4.3 アンケート調査

最後にアンケート調査を行う。調査項目は次の4項目である。

質問1 従来システムと提案システムにおいて、推薦動画の概要把握はどの程度できたか。

質問2 従来システムと提案システムにおいて、好みの動画を容易に探索できたか。

質問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。

表5 質問1から質問3の結果

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

表6 質問4の結果

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

質問4 提案システム利用時は、どの動画情報に注目していたのか。（複数回答可）

表5と表6に結果を示す。

表5での各質問は、1から5までの5段階評価で回答し、点数が高い方が良い評価である。

質問1の調査結果より、提案システムの動画概要が把握しやすいことがわかった。質問4の調査結果より、タグや推薦理由、関連度といった情報がユーザに注目されていることが明らかになった。よってYouTubeの推薦画面で確認できない情報が、提案システムで採用しているGUIにおいて、好みの動画を決定するのに役立っていると考えられる。これらの結果は、実験1-2での提案手法の適合率が高い理由を裏付けることとなった。

質問2の調査結果より、YouTubeが提案手法に比べ、好みの動画を見つけやすいという結果が得られた。提案手法の値が低い理由として、システムの使いにくさが挙げられる。具体的には、推薦実行から動画提示までの待ち時間、右側に提示している動画情報の過多などである。これらはユーザに労力をかける原因となっており、解決策として、新たな提示方法の考案や、質問4を参考とした提示情報の再検討が挙げられる。また、質問1および質問2の調査結果に対して、有意水準のt検定を行ったが、残念ながら明確な有意差は認められず、有意傾向がある、という結果が得られた。

質問3と質問4の調査結果より、ユーザは推薦理由にある程度納得しており、推薦理由が動画選択の補助となっていることも判明した。しかし推薦動画に対して、より適切な推薦理由を提示するために、推薦理由の精度を改善する手法を考案する必要がある。

5. 提案システムの拡張

前節の実験結果において、YouTubeと提案手法の平均動画選択数に大きな差があることがわかった。その原因として、提案手法では様々な関連度を持った動画を推薦すること、海外の動画を多く推薦することを挙げた。提案手法では、一度英語のタグが出現すると、そこから得られる動画やタグのほとんどが英語になってしまう。このようなコンテキスト表生成方法が原因で、視聴動画と関連性の薄い動画が多く提示され、平均動画選択数が少なくなったと考えられる。

この問題を解決するため、親タグという概念を用いて動画を取得する。親タグとは、あるタグよりも上位に属する、より一般的な概念を持ったタグのことを指す。例えば「読売ジャイアンツ」の親タグは「野球」、「スポーツ」などが該当する。このように視聴動画と関連し、かつ上位に属するタグを取得することができれば、様々な観点を持った動画を推薦可能であると考えられる。また、表記ゆれによって、適切に動画を取得できなかったタグについても、上位のタグを得ることにより、問題解決できると予想する。本研究では、親タグを取得する手法として、IR(Intersection Rate)値を用いる。YouTubeに存在するすべてのタグ集合 t の要素であるタグ t_i, t_j のIR値 IR_{ij} は次式によって算出する。

$$IR_{ij} = \frac{n(t_i t_j)}{n(t_i)} \quad (\in [0, 1]) \quad (2)$$

ここで $n(t_i)$ は、YouTube内におけるタグ t_i のキーワード検索結果数、 $n(t_i t_j)$ は、タグ t_i とタグ t_j でYouTubeによるAND検索を行ったときのキーワード検索結果数である。

また、次式を満たすとき、タグ j はタグ i の親タグであると定義する。

$$n(t_i) < n(t_j) \quad \text{and} \quad IR_{ij} \geq \alpha \quad (3)$$

ここで α は閾値を表し、これより大きいIR値を持ったとき、タグ t_j はタグ t_i の親タグである。

例えば、タグ t_i = “読売ジャイアンツ”の検索結果数 $n(t_i)$ が600件、タグ t_j = “野球”の検索結果数 $n(t_j)$ が1500件、 $t_i t_j$ = “読売ジャイアンツ野球”の検索結果数 $n(t_i t_j)$ が300件であるとする。このとき IR_{ij} は、

$$IR_{ij} = \frac{n(t_i t_j)}{n(t_i)} = \frac{300}{600} = 0.5 \quad (4)$$

となり、仮に α の値を0.3と設定する場合、「野球」は「読売ジャイアンツ」の親タグであると判断される。

視聴動画のタグより検索結果数が多く、かつIR値が α より大きいタグを親タグとすることで、視聴動画と関連した動画を取得し、多様な推薦を行うことを目指す。

5.1 改良システムの流れ

改良システムは次の6ステップから構成される。

- Step 1 : 親タグを取得
- Step 2 : コンテキスト表生成
- Step 3 : コンセプトラティス生成
- Step 4 : 推薦理由・関連度決定
- Step 5 : 推薦動画決定
- Step 6 : 推薦動画提示

ここでは提案システムと手法が異なるStep 1とStep 2について説明する。

Step 1 : 親タグを取得

ここでは、視聴動画に付与されている視聴動画タグそれぞれに対して親タグを求める。親タグを取得するために以下の操作を行う。

- (1) 視聴動画からタグ t_i を取得
- (2) t_i のキーワード検索結果から、上位1000件の動画を取得
- (3) (2)の動画に出現するタグの中で、 t_i との共起頻度が上位 n 個であるタグを取得
- (4) t_i と(3)のタグにおけるIR値をそれぞれ計算し、条件を満たすタグを親タグと判断
- (5) (1)から(4)の操作をすべての視聴動画タグについて適用

Step 2 : コンテキスト表生成

親タグ取得後にコンテキスト表を生成する。提案手法では、視聴動画と無関係なタグであるノイズタグが原因でコンテキスト表が巨大化し、コンセプトラティスが複雑化する。そのため、視聴動画から推測困難な推薦理由が提示されたり、視聴動画と無関係のタグを含むコンセプトが選択されるという問題が生じる。

そこで改良手法では、稀にしか出現しないタグを除く。これによりコンパクトなコンセプトラティス

となり、適切な推薦理由の提示が期待できる。

コンテキスト表生成のために、以下の操作を行う。

- (1) 視聴動画タグ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、一定の値以下であれば削除
- (3) タグを持たないURLを削除

本研究では主観評価実験により、出現回数が5回以下のタグをノイズタグとした。

6. 改良システム評価実験

6.1 実験2-1：平均動画選択数の比較と考察

同一の視聴動画に対するYouTube、提案手法、改良手法の推薦動画をそれぞれ5つ提示し、視聴したい動画を選択してもらい、平均動画選択数を算出する。この実験から、親タグを用いる改良手法の推薦動画が、YouTubeや提案手法の動画と比較してどれだけ選択されるのかを調査する。試行回数は20回であり、視聴動画は音楽、ペット・動物、映画・アニメ、スポーツの4ジャンルからそれぞれ5つの動画をあらかじめ用意してもらい、その後実験を行った。ユーザは3つの動画グループに対して、視聴したい動画を選択する。これを20回繰り返し、平均動画選択数を算出する。表7に実験結果を示す。

実験1の結果から、平均動画選択数はYouTubeが一番多く、次に改良手法、提案手法という順になった。改良手法は、視聴動画タグの親タグを用いることで、視聴動画と関連のある動画を多く取得可能となり、平均動画選択数が提案手法を上回ったといえる。さらにノイズタグを除去したことにより、コンセプトラティスが簡単化された。その結果、基点コンセプトと関連度の高いコンセプトを多く推薦することになり、これも提案手法より平均動画選択数が増加した原因だと考えられる。

また、実験1よりも今回の実験の方が平均動画選択数が高い。その理由として、被験者の動画選択方法の違いが挙げられる。前者の実験ではジャンルを指定せ

ず、20個の視聴動画を自由に選択してもらった。このとき被験者は視聴動画を、実験で提示した推薦動画から選択する傾向が見られた。これを繰り返していった結果、視聴動画が興味の対象から徐々に外れていったと考えられる。そのため、推薦された動画の中で、被験者の興味を持つ動画が少なかったことが予想される。後者の実験ではジャンルを指定し、あらかじめ20個の視聴動画を用意してもらった。被験者はキーワード検索画面からクエリを入力し、その検索結果から視聴動画を吟味して選択する傾向が見られた。その結果、被験者の関心が高い動画が視聴動画に選ばれ、推薦された動画中にも興味を持つ動画が多く含まれたことが考えられる。

6.2 実験2-1：推薦理由の評価

提案システムと改良システムでは、推薦理由の提示を実装している。そこで実験2-2では、推薦理由が理解できるものであるか、推薦理由がユーザの動画選択を助けることができるかを調査する。両システムの推薦動画における推薦理由を評価することで、適切な推薦理由について考察を行う。実験2-1で選択した20個の視聴動画から、提案手法と改良手法の推薦動画を提示した。その際、推薦理由を動画とともに提示し、全ての推薦動画に対して以下の2つの質問を行った。

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

質問1では、視聴動画と推薦動画の関係性が、推薦理由によって明示できているかを評価する。評価は1～5までの5段階で回答してもらい、点数が高いほど良い評価である。質問2では、提示された推薦理由が、次の視聴動画を選択する補助となっているかを評価する。提示された推薦理由が動画の視聴可否を決定する補助となった場合はチェックをしてもらう。質問2の値 P_s は、

$$P_s = \frac{\text{推薦理由が補助となった動画数}}{\text{試行回数} \times 5} \quad (\in [0, 1])$$

で表す。ユーザはどちらかのグループに属する動画1

表7 実験2-1 結果

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

表8 実験2-2 結果

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

つ1つに対して、2つの質問に回答してもらう。これを20回繰り返して、評価を行う。表8に実験結果を示す。

質問1では改良手法の値が高いという結果になった。改良手法ではノイズタグを出現回数によって除去し、一般的なタグが推薦理由として提示される。そのため、動画の内容をよく表すタグが推薦理由として表示された。また、推薦理由が簡潔なものが多く、直感的に把握可能であった。これらの要素がユーザの理解を助け、高い評価につながったと考えられる。

質問2では、改良手法の値が提案手法より低くなっている。改良手法では出現回数の多いタグが残るため、一般的な概念を持つタグが多く提示される。一方提案手法ではタグ除去の処理を行っていないため、ノイズタグは多く発生するが、具体的な概念を持つタグが多く提示される。そのため、動画の内容に踏み込んだ推薦理由を提示できる頻度が多くなり、動画選択の

補助になると回答した割合が高くなったと推測する。

表11は、動画タイトル「サッカー日本対タジキスタン8-0で歴史的優勝」からの、両手法を用いた動画推薦例である。提案手法1は、提案手法において、質問1の評価が1かつ質問2でチェックが付与されなかった動画である。ここではサッカーの動画から歌手の動画が推薦されており、タグ同士の繋がりもわかりにくい。そのため、ユーザにとって理解しにくい推薦になったと考えられる。提案手法2は、提案手法において、質問1の評価が5かつ質問2でチェックが付与された動画である。推薦理由の「前園」は、動画中の試合でキャプテンを務めている。また、「オリンピック」や「ブラジル」は、この試合がオリンピックの対ブラジル戦であることを表している。動画タイトルからはわからず、かつ動画の内容を表している推薦理由が提示されると、ユーザの視聴動画選択を補助することが可能

表9 Rel_{c_i} と p_{c_i}

c_i	Rel_{c_i}	p_{c_i}	c_i	Rel_{c_i}	p_{c_i}
c_1	3	$\frac{3}{25}$	c_6	3	$\frac{3}{25}$
c_2	2	$\frac{2}{25}$	c_7	2	$\frac{2}{25}$
c_3	1	$\frac{1}{25}$	c_8	4	$\frac{4}{25}$
c_4	4	$\frac{4}{25}$	c_9	3	$\frac{3}{25}$
c_5	3	$\frac{3}{25}$			

表10 推薦動画の15カテゴリ

1:エンターテインメント	2:ゲーム	3:コメディ	4:スポーツ	5:旅行とイベント
6:ニュースと政治	7:教育	8:ブログと人	9:ペットと動物	10:ハウツーとスタイル
11:映画とアニメ	12:音楽	13:自転車と乗り物	14:科学と技術	15:非営利団体と活動

表11 動画タイトル「サッカー日本対タジキスタン8-0で歴史的優勝」の推薦例

	タイトル	推薦理由
提案手法1	David Bowie - Let's Dance	サッカー, 日本, 対 → extreme, action → Let's
提案手法2	マイアミの奇跡	サッカー → 前園 → オリンピック, ブラジル
改良手法	中田英寿とザッケローニの対談	サッカー → 日本代表, ザッケローニ

となる。改良手法は、改良手法で推薦された動画と推薦理由である。推薦理由で提示しているのは一般的な概念を持ったタグであり、視聴動画と改良動画の関係がわかりやすいが、動画タイトルから容易に推測可能であり、視聴動画選択の補助とはならない。これはある動画における例であるが、両手法においてこのような傾向が見られることがわかった。

7. おわりに

動画共有サイトであるYouTubeには、視聴動画と推薦動画の関係性が明かではない、動画情報が限定的、推薦の偏りという問題があった。そこで本研究では、形式概念分析を用いたYouTubeの動画推薦アプリケーションを構築した。

提案システムでは、コンセプト間の関係を用いて推薦理由を明示し、視聴動画と推薦動画の関係性を示した。また、構築したアプリケーション上で推薦動画の情報を提示することで、あらかじめ推測した内容と実際に視聴した内容との差異を小さくした。さらに様々なコンセプトから動画を選択することで、推薦の偏りを解消し、多角的な推薦を行った。

20代の男女10名の被験者に対して実験を行った結果、提案システムでは動画概要の把握が容易であることがわかった。また、推薦動画も多様な種類が提示できていることも確認した。改良システムでは親タグを導入してノイズタグを除去し、性能を確認するために被験者実験を行った。その結果改良システムは、改良前システムと比較し、平均動画選択数や推薦理由の理解度が高いことが判明した。今後は推薦までの待ち時間を短縮させ、ユーザに提示する情報を精選する必要がある。また、多角的な推薦を行いつつ、ユーザの興味がある動画を提示するアルゴリズムの改良も課題である。

参考文献

- [1] 総務省情報通信政策研究所, “我が国の情報流通量の指標体系と計量手法に関する報告書”, 2011.
- [2] <http://youtube-global.blogspot.com/2011/05/thanks-youtube-community-for-two-big.html>
- [3] 土方嘉徳, “情報推薦・情報フィルタリングのためのユーザプロファイリング技術”, 人工知能学会誌, Vol.19, No.3, p.365-371, 2004.
- [4] 鈴木治, “解説:形式概念分析とその応用”, 日本知能情報ファジィ学会, 2006.
- [5] 角田雅照, 秋永知宏, 引地一将, 大杉直樹, 柿元健, 門田曉人, 松本健一, “協調フィルタリングを用いたソフトウェア技術者向け開発技術推薦の試み”, 教育システム情報学会誌, Vol.27, No.4, pp.316-326, 2010.
- [6] 吉井和佳, 後藤真孝, “音楽推薦システム”, 情報処理(情報処理学会誌), Vol.50, No.8, pp.751-755, 2009.
- [7] 小島博, 山田亜樹, 仲秋洋, “浸透するタイムシフト広がる動画視聴〜「デジタル放送調査2010」から・パートI〜”, NHK放送文化研究所, 2011.
- [8] J.L.Herlocker, J.A.Konstan, and J.Riedl, “Explaining Collaborative Filtering Recommendations”, Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, pp.241-250, 2000.
- [9] R. and Swearingen, K, “The Role of Transparency in Recommender Systems”, in Proc. of the SIGCHI Conf. on Human Factors in Computing Systems, pp.830-831, 2002.
- [10] S. Andrews, “In-Close, a Fast Algorithm for Computing Formal Concepts”, International Conference on Conceptual Structures, 2009.

(2012年6月1日 受付)

(2012年10月2日 採録)

[問い合わせ先]

〒305-8573 茨城県つくば市1-1-1

筑波大学 システム情報系 知能機能工学域

延原 肇

TEL: 029-853-5153

FAX: 029-853-5153

E-mail: nobuhara@iit.tsukuba.ac.jp

著者紹介



きたむら ゆうたろう [非会員]
北村 祐太郎

1987年生。2010年 釧路工業高等専門学校専攻科 修了。2012年 筑波大学大学院 システム情報工学研究科 博士前期課程修了。修士(工学)。同年より中部電力株式会社入社。



さわ せ かずひと [非会員]
澤勢 一史

2008年筑波大学第三学群工学システム学類卒業。2010年筑波大学大学院システム情報工学研究科知能機能システム専攻博士前期課程修了。現在、筑波大学大学院システム情報工学研究科知能機能システム専攻博士後期課程在学中。2010年より日本学術振興会特別研究員(DC1)。形式概念分析に基づく、大規模情報の可視化システムに関する研究に従事。IEEE、情報処理学会、電子情報通信学会各会員。



のぶはら はじめ [正会員]
延原 肇

2002年3月東京工業大学大学院博士課程修了。博士(工学)。2002年4月～9月カナダアルバータ大学博士研究員。2002年10月～2006年3月東京工業大学大学院総合理工学研究科助手。2006年4月～現在 筑波大学大学院システム情報工学研究科 講師。

Formal Concept Analysis Based Movie Recommendation

by

Yutaro KITAMURA, Kazuhito SAWASE and Hajime NOBUHARA

Abstract :

YouTube as the most popular video hosting service is growing rapidly, making it difficult for users to find video that fit their particular interest. Implementation of video recommendation method is one way to solve this problem. However, several difficulties remain, one of them is no clear relation between the original and recommended video. Moreover, there are many cases that the recommended videos mostly have similar contents to the original videos. In this paper, we try to solve these problems by concept of words and to explicitly state the reason for recommendation. We build video recommendation application for YouTube using Formal Concept Analysis, and we performed the experiments to evaluate our system. By analyzing the results of experiments, we look at the effect of our system on user's load.

Keywords : Formal Concept Analysis, Recommendation, Movie Recommendation, Information Explosion, Granular Computing

Contact Address : **Hajime NOBUHARA**

Faculty of Engineering, Information and Systems, University of Tsukuba
1-1-1, Tsukuba, Ibaraki, 305-8573, JAPAN
TEL : 029-853-5153
FAX : 029-853-5153
E-mail : nobuhara@iit.tsukuba.ac.jp