

低次画像特徴量を用いた動画の視聴者層推定

山元浩平^{1,a)} 中山英樹^{1,b)}

1. はじめに

近年、ユーザがインターネット上で所望のコンテンツを見つける手段として、また逆に、プロバイダがユーザに対して有効なターゲット広告を行う手段として、レコメンデーションやコンテンツマッチング等の技術の需要が高まってきている。これらの技術は、ソーシャルネットワーキングサイトやエレクトロニックコマースサイトなど、特定の分野で一定の成果を上げている。しかし、上記サイト等と異なり、動画共有サイトにおいては、ユーザがログインを行わずに動画を視聴するケースや、新しくアップロードされた動画を視聴するケースなど、視聴履歴データのないユーザや視聴者統計データの乏しい動画に対して、有効なレコメンデーション・ターゲット広告を行えていないのが現状である。そこで本研究では、その一助として、動画の低次画像特徴量から視聴者層を推定する手法を提案する。動画のサムネイル画像から得られる SIFT[1]などの低次画像特徴量を入力とし、ロジスティック回帰を行うことで、どの視聴者層が一番その動画を見ている可能性が高いかを推定するモデルを考えた。また、動画共有サイト YouTube[2]の動画と視聴者統計データを利用した評価実験を行ない、提案手法による動画の視聴者層推定の有効性を検証するとともに、各種低次画像特徴量で学習・識別に用いた識別モデルを Precision-Recall-Curve で比較し、性能評価を行った。

2. 提案手法

2.1 手順

本研究の提案手法の手順を図1に示す。本手法は、以下のように大きく分けて2つのステップから構成される。

Step 1. 視聴者統計データを保持する各動画のサムネイル画像から各種低次画像特徴量を抽出し、画像一枚ごとの特徴ベクトルを作成する。

Step 2. ステップ1で得られた特徴ベクトル集合を訓練データの入力、視聴者層の各クラスを出力として学習を行う。学習にはロジスティック回帰を用い、年齢層、性別等の視聴者層ごとに区切られた各クラスに識別し、ターゲット視聴者層を推定する。

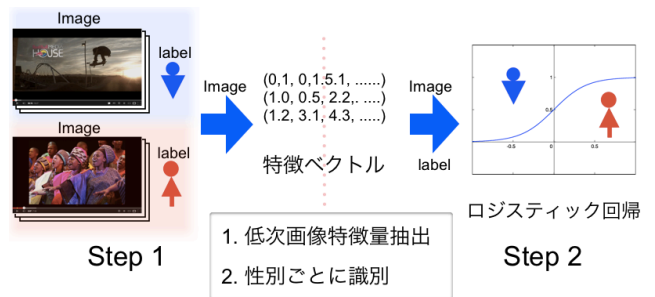


図1 提案手法の手順

2.2 低次画像特徴量抽出

本研究で用いる低次画像特徴量の種類とその抽出方法を表1にまとめる。SIFT等の局所特徴量は Fisher vector に変換し、画像一枚ごとの特徴ベクトルとして表現している。

表1 画像特徴量とその抽出方法

画像特徴量	抽出方法
Fisher vector[3] (SIFT)	SIFT, C-SIFT, Opponent-SIFT, RGB-SIFTをそれぞれ抽出した後、PCAで64次元に圧縮
Fisher vector (C-SIFT[4])	し、Fisher kernelを用いてFisher vectorに変換する。GMMの混合数は64個。
Fisher vector (Opponent SIFT[5])	
Fisher vector (RGB-SIFT[5])	
RGBヒストグラム	RGBカラー空間のR, G, Bの値のヒストグラム
輝度ヒストグラム	HSLカラー空間のHの値のヒストグラム
彩度ヒストグラム	HSLカラー空間のLの値のヒストグラム

2.3 識別モデル

動画共有サイトの動画から視聴者層を推定する研究として、Ulgesら[6]の研究がある。Ulgesらは各動画のセマンティックなコンテキスト（動画の内容。例えばサッカーや歌手など）を推定し、そのコンテキストによって視聴者層の推定を行っており、動画の視聴者層推定に、コンテキストを利用することの有効性を示した。しかし、あらゆるコンテキストに対してラベル付けを行っていくことは、現実的に困難である。また、コンテキストのラベル付けを行うことで、本来推定に重要な情報が落ちてしまう可能性も考える。そこで本研究では、動画のコンテキストというラベル付けを介することなく、直接的に低次画像特徴量と高次のユーザデータ（ここでは各性別）をロジスティック回帰で結び付けることを検討した。

3. 実験

3.1 実験概要

YouTube Trends Map[7]から得られる性別ごとの人気動画のサムネイル画像を用いて、低次画像特徴量による視聴者層推定の精度評価実験を行った。YouTube Trends Map

1 東京大学大学院情報理工学系研究科 〒 113-8656 東京都文京区弥生 1-1-1
 a) yamamoto@nlab.ci.i.u-tokyo.ac.jp
 b) nakayama@ci.i.u-tokyo.ac.jp

では、過去 24 時間で最も多く視聴されたトップ 10 の動画画像を、地域、性別、年齢層ごとに検索することができる。今回、このサイトから地域を日本に固定して動画画像を検索・収集し、性別ごとによく視聴されている動画画像のサムネイル画像のデータセットを得た。それらの画像から各種低次特徴量を抽出し、それをロジスティック回帰によって、男性と女性に識別することで、本手法の有効性を検証した。また、識別に使ったモデルの性能を Precision-Recall-Curve で比較した。

3.2 実験条件

【データセット】

YouTube Trends Map より、性別ごとの人気動画画像のサムネイル画像を収集し、重複を含む男性 3029 枚、女性 3252 枚、計 6281 枚を訓練用のデータセットとし、重複を含まない男性 354 枚、女性 251 枚を検証用のデータセットとして用いた。

【画像特徴量】

2 章で示したように、識別に使う低次画像特徴量として、大域特徴量である RGB ヒストグラム、輝度ヒストグラム、彩度ヒストグラムと、局所特徴量の SIFT, C-SIFT, opponent-SIFT, RGB-SIFT をそれぞれ Fisher vector に変換し、画像一枚の大域的な特徴ベクトルとしたものを用いた。

【モデルパラメータ】

識別は、コストパラメータ $C=1000$ の L2 正則化ロジスティック回帰で行った。

3.3 実験結果

ロジスティック回帰によって得られた各特徴量の識別モデルの Precision-Recall-Curve を図 2 に、適合率、再現率、F 値を表 2 に示す。また、最も性能が良かった Fisher vector (C-SIFT) を入力としてロジスティック回帰を行った際、事後確率が高かったサムネイル画像の例を、性別ごとに図 3 に示す。

図 2, 表 2 より、画像の局所的な形の特徴を含む各種 SIFT 特徴量を Fisher Vector に変換した特徴量の性能が良いことが分かる。また図 3 より、性別ごとに動画画像の嗜好の特徴がよく表れていることが確認できる。

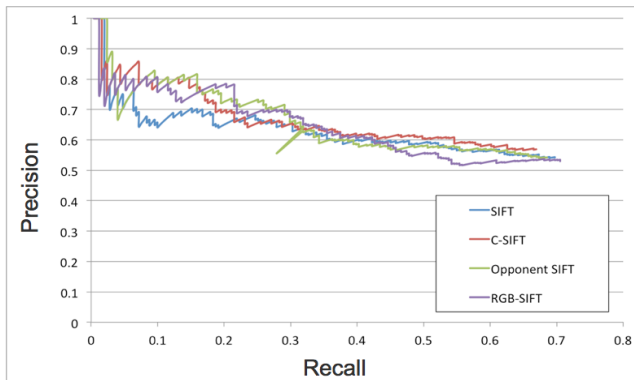


図 2 特徴量ごとの Precision-Recall-Curve 比較

表 2 特徴量ごとの適合率, 再現率, F 値比較

	適合率	再現率	F 値
Fisher vector (SIFT)	65	63	63
Fisher vector (C-SIFT)	66	65	66
Fisher vector (Opponent SIFT)	65	63	63
Fisher vector (RGB-SIFT)	65	62	62
RGBヒストグラム	58	58	58
輝度ヒストグラム	51	51	51
彩度ヒストグラム	56	55	55

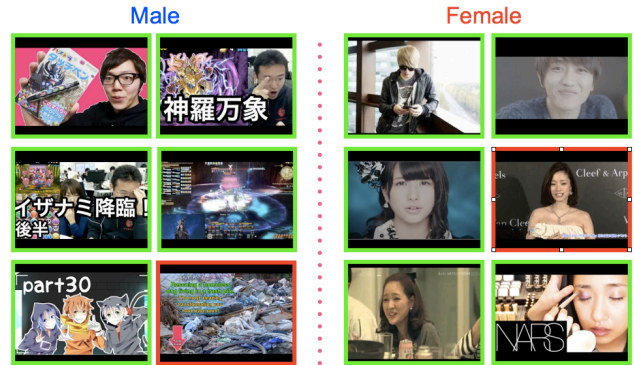


図 3 テストに用いたサムネイル画像のうち、性別ごとのロジスティック回帰の事後確率が高かった画像の例

4. まとめ

本稿では、低次画像特徴量を用いて動画画像の視聴者層を推定する手法を提案した。また、評価実験により、提案手法が動画画像の視聴者層推定に有効であることを確認した。

今後は、サムネイル画像の低次画像特徴量だけではなく、他のモダリティの特徴量等も利用し、それらを組み合わせることで、視聴者層推定のためのモデルの改良を検討する。また、識別においも、性別だけではなく年齢層も考慮することで、よりきめ細かいターゲティングが可能になるよう改良を行う。

謝辞 本研究は、公益財団法人放送文化基金、JST CREST の研究課題「複雑データからのディープナレッジ発見と価値化」の支援を得て実施された。

参考文献

- [1] Lowe, D. G.: Object recognition from local scale-invariant features. In Proc. IEEE ICCV (1999)
- [2] YouTube, <https://www.youtube.com/>
- [3] Perronnin, F., Sánchez, J., Mensink, T.: Improving the Fisher kernel for large-scale image classification. In: Proc. ECCV (2010)
- [4] Burghouts, G. J., Geusebroek, J.M.: Performance evaluation of local color invariants. Computer Vision and Image Understanding 113(1), 48-62 (2009)
- [5] van de Sande, K. E. A., Gevers, T., and Snoek, C. G. M.: Evaluating color descriptors for object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9), 1582-96 (2010)
- [6] Ulges, A., Koch, M., Borth, D.: Linking Visual Concept Detection with Viewer Demographics. In: Proc. ICMR (2012)
- [7] YouTube Trends Map, <http://www.youtube.com/trendsmap/>