

E-043

インターネット動画コンテンツへのテレビジャンルの自動付与 Automatic Classification of the Internet Videos into the Genres of TV Programs

水岡 良彰†
Yoshiaki Mizuoka

鈴木 優†
Masaru Suzuki

1. はじめに

近年、インターネットサービスの発展やテレビの多チャンネル化により、さまざまな動画コンテンツが閲覧可能となっている。そのため、ユーザは多くのコンテンツから見たいコンテンツを探す必要がある。探す手段としては、例えばテレビの場合、番組表から探す、番組名から探す、出演者から探す、ジャンルなどの分類から探すなどが考えられる。

特に、多くのコンテンツから目的のコンテンツを探す場合には、大まかに区分けされたジャンルあるいはカテゴリといった分類が、候補の絞り込みのための重要な手掛かりになるだろう。しかし動画配信サービスによって動画の分類体系は異なっている。そのため普段慣れ親しんでおらず、分類体系を把握していない動画配信サービス上では、分類を手掛かりに目的の動画コンテンツを探すことができないと考えられる。

そこで本研究では、普段テレビを視聴しているユーザが目的の動画コンテンツを探せるよう、動画コンテンツにテレビと同じ体系のジャンルを付与することを提案する。本稿では YouTube[1]をインターネット動画コンテンツの例として、ジャンル自動付与実験を行い、精度を評価する。

以下、2章では具体的な動画コンテンツの分類について述べ、3章ではテレビジャンル付与の事前検証として人手によるジャンル付与の実施と考察を述べる。4章で提案するジャンル自動付与手法を説明し、5章で実験と考察、6章でまとめと今後の課題を述べる。

2. 動画コンテンツの分類

本稿で付与を行うテレビジャンルと、ジャンル付与対象である YouTube のカテゴリそれぞれについて説明する。

テレビジャンルはデジタル TV 放送向けとして、社団法人電波産業会の ARIB STD B10 にて規定されている。ジャンルはまず大分類に分けられ、各分類がさらに細かく中分類として分けられており、各放送局によって付与されたジャンルを電子番組表(EPG: Electronic Program Guide)から取得できる。

一方、YouTube カテゴリは YouTube にて決められた分類である。アップロードされた動画は、YouTube カテゴリのいずれかに属しており、ランキングやキーワード検索に対して、YouTube カテゴリを指定して目的の動画を探すことができる。

テレビ番組と YouTube 動画の分類を表 1 に示す。表 1 の左列にはテレビジャンルの大分類、右列には国や地域の設定を日本としたときの YouTube カテゴリの一覧を示してい

表 1 テレビと YouTube の分類の違い

テレビジャンル(大分類)	YouTubeカテゴリ
・ニュース/報道	・ニュースと政治
・スポーツ	・スポーツ
・音楽	・音楽
・趣味/教育	・教育
・バラエティ	・コメディ
・映画	・映画とアニメ
・アニメ/特撮	エンターテイメント
情報/ワイドショー	自動車と乗り物
ドキュメンタリー/教養	ペットと動物
ドラマ	旅行とイベント
劇場/公演	科学と技術
福祉	ゲーム
その他	ハウツーとスタイル
	ブログと人

る。この表より、表の上の方(・の付いた項目)に示した分類は、分類名からテレビジャンルと YouTube カテゴリ間である程度対応がつかうと考えられるものの、下の方に示した分類では互いに対応のつく分類が存在しないことが分かる。

ここではテレビジャンルと YouTube カテゴリを例に挙げたが、他のコンテンツでも同様のことがいえ、一般に同じ動画コンテンツであっても、動画配信サービスによってコンテンツの分類が異なっている。

3. テレビジャンル付与の事前検証

テレビ番組と YouTube 動画は、動画配信サービスや動画コンテンツの分類や質、配信方法などが異なる。そこで、YouTube 動画にテレビジャンルを手で付与することで、そもそも本質的にテレビジャンルを YouTube 動画に付与可能なか検証および考察を行った。

3.1 人手によるジャンル付与の対象と方法

本検証ではテレビジャンルの大分類-中分類の組を、YouTube 動画へ人手で付与した。なお、ジャンル付与作業は筆者 1 人で実施した。

ジャンル付与対象の YouTube 動画は、2009 年 6 月 15 日に取得した、全 14 の YouTube カテゴリにおける各再生回数上位 10 件の計 140 件、再生回数上位 100 件、おすすめ動画 100 件とした。取得にあたっては、リージョンを日本、再生回数の期間は全期間と設定し、YouTube API により取得した。

各 YouTube 動画への付与ジャンル数は 1~3 ジャンルとし、付与するジャンルが 1 つも無いと判断された場合のみ、大分類と中分類ともに「その他」を 1 つ付与した。すなわち、「その他」が付与された動画数は、他のどのテレビジャンルも付与されなかった動画数を表す。

3.2 人手によるジャンル付与の結果および考察

各 YouTube カテゴリ再生回数上位の動画 10 件について、各テレビジャンルが付与された動画数を表 2 に示す。なお、1 動画に複数のテレビジャンルが付与されることがあるこ

†株式会社東芝 研究開発センター 知識メディアラボ
ラトリー

とから、各 YouTube カテゴリの付与ジャンル数は 10 以上となっている。

表 2 より、YouTube カテゴリからテレビジャンルへの 1 対 1 対応が付かないものが多く見受けられる。表 1 で対応が付くと考えられるとして挙げた YouTube カテゴリについても、実際には対応が付いていないものがあることが分かる。実際にジャンル付与を実施してみたところ、中には簡単な編集やコメントによって既存の動画を別の視点で見せている動画や、そもそも意図の分からない動画などが存在し、適切と感じるテレビジャンルが受け手によって変わってしまいそうな動画が存在していた。このような動画には複数のジャンルを付与することが望ましい。これらの点から、動画の説明や中身を考慮せずに YouTube カテゴリを単純にテレビジャンルへマッピングする方法では、ジャンル付与は難しいといえる。

次に、「その他」以外のテレビジャンルが付与された YouTube 動画の割合を表 3 に示す。表 3 より、ほとんどの動画に「その他」以外が付与されていることが分かる。すなわちテレビジャンルが付与できない動画は少ないので、YouTube 動画へのテレビジャンル付与は可能といえる。

以上のことから、YouTube 動画にテレビジャンルを付与することは多くが可能といえる。しかしテレビジャンルを YouTube 動画に付与する際、基本的に YouTube カテゴリからのマッピングによって一意に決めることはできない。また、本稿では YouTube 動画にテレビジャンルを付与するが、分類の異なる他のインターネット動画についても適用可能な技術が望ましい。したがって、ジャンル自動付与ではジャンル付与対象の分類を用いるのではなく、動画の説明や中身を用いた技術が望ましいといえる。

4. テレビジャンル自動付与手法

本章では以下のようなテレビジャンル自動付与手法を提案する。本手法は、大きくジャンル行列生成とジャンル付与に分けることができる。ジャンル行列生成では、学習コーパスから各ジャンルを表すジャンルベクトルの集合であるジャンル行列を獲得する。ジャンル付与では、付与対象から付与対象を表すコンテンツベクトルを獲得し、学習で得られた各ジャンルベクトルとの類似度に基づいてジャンルが付与される。以下で、各手順について説明する。

さらにジャンルベクトルおよびコンテンツ行列生成およびジャンル付与のそれぞれについて、LSA (Latent Semantic Analysis)[2]の適用による改良を説明する。

4.1 ジャンル行列生成

ジャンル行列生成は次の手順で行う。

1. 学習コーパスであるテレビ番組 EPG データの、タイトル・説明・詳細説明の各テキストを抽出し、形態素解析を行い、出現単語を獲得する。
2. 必要に応じて、出現単語をフィルタリングする。
3. 出現単語から、ジャンルに対する単語の重み行列であるジャンル単語行列を生成する。この行列の各行は、ジャンルを表すジャンルベクトルに対応する。

重み行列の生成には、TF-ICF (Term Frequency - Inverse Category Frequency)[3]を用いる。これは情報検索等でよく用いられる TF-IDF (Term Frequency - Inverse Document Frequency)[4]を拡張したものである。TF-IDF は単語出現頻度(TF)と文書頻度の逆数(IDF)との積からなる。一方 TF-

表 2 付与テレビジャンルと YouTube カテゴリの共起数

テレビジャンル 大分類	YouTubeカテゴリ											合計		
	映画とアニメ	自動車と乗り物	ペットと動物	スポーツ	旅行とイベント	ゲーム	コメディ	ブログと人	ニュースと政治	エンターテイメント	教育		ハウツーとスタイル	科学と技術
ニュース/報道								1	5			1		7
スポーツ		1		8			1	1		1				12
情報/ワイドショー				3	2		1	1		1		2		10
ドラマ							1							2
音楽	5	1	10				1		1		3	1	1	23
バラエティ				2	3		5	3	1	3	2	5	2	26
映画	3	1					1							6
アニメ/特撮	5						3	1		1	2		1	13
ドキュメンタリー/教養		1	1			3		1	3		2		7	18
劇場/公演														0
趣味/教育	1	6	2	10		1	10	1		2	1	3	4	44
福祉														0
その他	1	1			4			2	2	2	2		1	15
合計	15	11	13	10	13	13	11	10	14	11	14	13	15	176

表 3 「その他」以外が付与された YouTube 動画の割合

各カテゴリ再生回数 上位10件合計140件	再生回数 上位100件	おすすめ 動画100件
89.3%	92%	99%

ICF は IDF のかわりにカテゴリ頻度の逆数(ICF)を用いることで、カテゴリに対する重み付けが算出できる。

4.2 ジャンル付与

ジャンル付与の手順は次の通りである。

1. テレビ番組 EPG データに対しては、タイトル・説明・詳細説明の各テキストを抽出し、形態素解析を行い、出現単語を獲得する。YouTube メタデータに対しては、タイトル・タグ一覧・説明の各テキストを抽出し、形態素解析を行い、出現単語を獲得する。
2. ジャンル行列生成のときと同様に、必要に応じて、出現単語をフィルタリングする。
3. 付与対象の出現単語のうち、学習コーパスに出現した語について、各要素の値に TF-ICF を用いたジャンル単語行列を生成する。なお ICF 値については学習コーパスにより得られた値を用いる。このジャンル単語行列の各行は、コンテンツを表すコンテンツベクトルとなっている。
4. 各コンテンツベクトルと、ジャンル行列生成で得られた各ジャンルベクトルとのコサイン類似度を調べ、閾値 t 以上のジャンルをコンテンツに付与する。

4.3 LSA の適用

ジャンルベクトル生成やジャンル付与の過程で得られるジャンル単語ベクトルは、出現単語の種類が多いため、非常にスパースとなると考えられる。そこで語の類似性を考慮して行列のスパース性を解消する効果のある LSA を適用することで、上記の問題の解決を図る。

LSA には、語の類似性を考慮して行列のスパース性を解消する効果があり、さらにデータ中のノイズを除去する効果や、行列を良い精度で近似する効果がある。

LSA の具体的な適用方法は次の通りである。まずジャンル行列生成で得られるジャンル単語行列 X を特異値分解により $X = U \Sigma V^T$ と分解し、大きい順に k 番目までの特異値を用いて $\hat{X} = U_k \Sigma_k V_k^T$ と変換する。このときの U_k を圧縮したジャンル単語行列として用いる。さらに、ジャンル付与でのジャンル単語行列 Y を $U'_k = Y V_k \Sigma_k^{-1}$ と変換した

ときの U'_k を圧縮したジャンル単語行列として用いる。なお本稿では、元の行列サイズに比べてランク数は十分小さいため、ランク数の低減は行わずに $k = \text{ランク数}$ とする。

5. テレビジャンル自動付与実験

YouTube 動画およびテレビ番組にテレビジャンルを自動付与する実験を行う。テレビ番組への付与は比較のためである。

ジャンル付与の手掛かりには動画コンテンツに付与された説明文を使用する。テレビ番組はテレビ局により付与された EPG のテレビジャンルを正解とし、YouTube 動画は 3 章にて人手で付与したテレビジャンルを正解とした。評価指標として適合率、再現率、F 値を算出する。

なおテレビジャンル付与は中分類で行い、付与された中分類から対応する大分類に置き換えることで大分類の付与を行う。

5.1 実験条件

学習コーパスは、2009 年 3 月 16 日から 4 月 27 日までの毎週月曜日に取得した 7 週間分の EPG データについての、地上波放送分の EPG データ (22877 番組) と、地上波・BS・CS 放送分の EPG データ (142291 番組) の 2 通りとした。

ジャンル付与対象は、2009 年 6 月 15 日に取得したおすすめ動画 100 件 (100 動画) と、2009 年 7 月 13 日に取得した当日の地上波放送分の EPG データ (420 番組) の 2 通りとした。なお今回の実験で付与対象とする YouTube 動画におすすめ動画を選択したのは、動画の説明文が比較的しつかりと書かれており扱いやすいと判断したためである。

また、単語のフィルタリングは、全形態素を使用する場合、名詞のみを使用する場合、名詞および英字とカタカナの未知語を使用する場合、の 3 通りとした。

LSA の適用については、LSA なしと LSA ありの 2 通りとした。

実験条件は以上の計 24 通りの組み合わせとした。

5.2 評価方法

評価は、付与されたテレビジャンルの大分類について、再現率と適合率、およびこれらの調和平均である F 値を用いる。これらの値はコサイン類似度の閾値で変化するため、閾値を 0.01 刻みで動かして調べる。

なお、付与対象が多く、問題を検索タスクと見なした場合には、評価指標として適合率を重視すべきと考える。しかし今回の実験では、付与対象数に限りがあり、適合率のみを見ると絞り込み過ぎた結果を良いとして評価してしまう恐れがある。そこで今回は F 値を中心にすることとする。

5.3 結果

表 4 に各条件の各閾値設定での評価値を示す。まず、F 値が最大となる閾値設定での結果については、YouTube 動画へのジャンル付与では、学習コーパスが地上波・BS・CS 放送 EPG データの方が高い F 値を示している。一方、テレビ番組へのジャンル付与では、学習コーパスに地上波放送 EPG データのみを使用した方が F 値は高い。また全体的に、テレビ番組よりも YouTube 動画へのジャンル付与の方が F 値は低い。さらに、単語のフィルタリングについては、使用する語が名詞+英字+カタカナの場合が最も F 値が高く、次に名詞のみが高く、全形態素が最も低い。また、LSA を用いた方が高い F 値を示している。

次に、YouTube 動画へのジャンル付与について、F 値が最大となる閾値設定での評価値と、テレビ番組へのジャンル付与で F 値が最大となる設定と同じ閾値設定での評価値を比較する。名詞のみや名詞+英字+カタカナを使用する条件で、LSA を適用していない場合は F 値が大きく下がっているが、LSA を適用すると下げ幅が小さくなっている。

5.4 考察

F 値最大の閾値設定の結果について：

YouTube 動画へのジャンル付与では、学習コーパスが地上波・BS・CS 放送 EPG データの方が高い F 値を示している。これは、より多くの学習コーパスを扱う方が、多くの出現語を扱って有利なためと考えられる。

一方、テレビ番組へのジャンル付与では、学習コーパスに地上波放送 EPG データのみを使用した方が高い F 値を示している。これは、地上波放送と BS・CS 放送では、放送される番組の傾向や、EPG に出現する語彙の傾向に違いがあるため、付与対象である地上波放送と同じ EPG データのみを学習コーパスとした方がうまく付与できたのだと考えられる。

また、YouTube 動画へのジャンル付与の方がテレビ番組よりも低い F 値を示している。これは、学習コーパスがテレビ番組 EPG であるので、YouTube 動画の説明文で使用される語彙の傾向に違いがあるためと考えられる。これを改善するには、テレビ番組 EPG には出現しないが YouTube 動画メタデータには出現する語を、ジャンル付与の手掛かりにできるようにする必要があると考えられる。

単語フィルタリングについては、ジャンル付与に使用する語が、名詞+英字+カタカナの場合が最も F 値が高く、次に名詞のみが高く、全形態素が最も低い。まず全形態素の場合は、ジャンル付与の手掛かりとならないノイズが多かったために精度が上がらなかったと考えられる。名詞のみの場合には、テレビジャンル付与と関係の無い形態素を除去できたため全形態素の場合に比べて精度は向上したが、有用な手掛かりであるのにも関わらず形態素解析器にて未知語と判定された語を使用できていないため名詞+英字+カタカナに比べて精度が低いと考えられる。これに対し、名詞+英字+カタカナは、未知語と判定されるもののうち、語の区切りが正しく処理されやすいと思われる英字とカタカナの追加を考えた条件である。実際の実験結果で F 値は高くなっており、想定した通り、有用な手掛かりが増えることで F 値が改善したのだと考えられる。なお、今回実施した実験条件では、名詞+英字+カタカナの場合が最も良かったが、英字やカタカナには意味を成さない文字列も多く含まれていることが確認できており、改善の余地がある。

LSA の有無では、LSA ありの方が高い F 値を示している。今回、テレビ番組 EPG に対して特異値分解を行ったので、テレビ番組の分類にとって良いように語の類似性の考慮やノイズ除去が行われたことになるが、YouTube 動画へのジャンル付与でも精度が向上している。これは、テレビ番組の分類にとってのノイズが YouTube 動画の分類にとってもノイズであり、またテレビ番組 EPG 上での語の類似性は YouTube 動画メタデータ上でもいえたために、YouTube 動画の分類でも効果があったと考えられる。

また、LSA を適用することで、高い F 値を取る閾値の値が大きくなっている。これは、LSA によるスパース性の解消によって、表層語の表現が異なっても意味的に類似

表4 各条件での各閾値設定の評価値

			F値最大の閾値設定								TVと同じ閾値設定				
			TV				YouTube				YouTube				
			閾値	適合率	再現率	F値	閾値	適合率	再現率	F値	閾値	適合率	再現率	F値	
LSAなし	地上波	全形態素	t=0.30	0.3922	0.3702	0.3809	t=0.19	0.139	0.6479	0.2289	t=0.30	0.1438	0.3028	0.195	
		名詞	t=0.11	0.6078	0.6066	0.6072	t=0.04	0.381	0.3944	0.3875	t=0.11	0.5789	0.0775	0.1366	
	地上波+BS+CS	全形態素	t=0.10	0.6333	0.626	0.6296	t=0.04	0.3831	0.4155	0.3986	t=0.10	0.5556	0.0704	0.125	
		名詞	t=0.26	0.2242	0.4922	0.3081	t=0.18	0.1327	0.7394	0.2251	t=0.26	0.1316	0.493	0.2077	
	LSAあり	地上波	全形態素	t=0.12	0.6537	0.5194	0.5788	t=0.04	0.3856	0.4155	0.4	t=0.12	0.7059	0.0845	0.1509
			名詞	t=0.09	0.5812	0.624	0.6019	t=0.04	0.4049	0.4648	0.4328	t=0.09	0.5714	0.0845	0.1472
地上波+BS+CS		全形態素	t=0.42	0.8005	0.6376	0.7098	t=0.17	0.1616	0.7042	0.2628	t=0.42	0.2838	0.1479	0.1944	
		名詞	t=0.42	0.7526	0.7074	0.7293	t=0.40	0.4409	0.3944	0.4164	t=0.42	0.4561	0.3662	0.4063	
地上波		全形態素	t=0.50	0.8345	0.6841	0.7519	t=0.42	0.4911	0.3873	0.4331	t=0.50	0.5352	0.2676	0.3568	
		名詞	t=0.39	0.7103	0.5891	0.6441	t=0.27	0.2208	0.493	0.305	t=0.39	0.2925	0.2183	0.25	
地上波+BS+CS	全形態素	t=0.44	0.7358	0.6802	0.7069	t=0.43	0.5248	0.3732	0.4362	t=0.44	0.5368	0.3592	0.4304		
	名詞	t=0.44	0.7495	0.6899	0.7185	t=0.37	0.493	0.493	0.493	t=0.44	0.5429	0.4014	0.4615		

している語であれば類似度が算出されるようになるため、類似度の値が全体的に大きくなったためと考えられる。

TVと同じ閾値設定の結果について：

前述したように、F値はコサイン類似度の閾値で変化する。ジャンル付与に使用する単語が名詞のみや名詞+英字+カタカナとする条件下では、F値が最大となる閾値設定での結果に比べ、LSAを適用していない場合はF値が大きく下がっているが、LSAを適用すると下げ幅が小さくなっている。この現象について、より詳細に観察するために、学習コーパスに地上波・BS・CS放送のEPGを使用し、使用する単語を名詞+英字+カタカナとした条件での、閾値を0.01で刻んだ際のF値の推移を図1に示す。LSAを適用した場合は、高いF値をとる閾値の範囲が広く、ジャンル付与対象の違いでその閾値の範囲が大きく変わらない。一方でLSAを適用しない場合は、高いF値をとる閾値の範囲が狭く、またジャンル付与対象によって高いF値をとる閾値は異なってしまうことが分かる。

これはLSAの近似やノイズ除去の効果により、ジャンル付与の主要ではない手掛かりをそぎ落とし、ジャンル付与対象に関わらず適用できる手掛かりが残ったためであると考えられる。

今回の評価実験では、正解のテレビジャンルを人手で付与したYouTube動画の用意があったが、実際には存在しない。そのため、適切な閾値設定が分からないので、テレビ番組にジャンル付与した際に良かった閾値設定を用いることができる。以上からLSAによって適切な閾値設定の範囲が近くなるのは好ましい性質といえる。

6. まとめと今後の課題

本稿では、YouTube動画を例に、動画コンテンツにテレビと同じ体系のジャンルを自動付与する手法を提案し、実データを使った実験により、品詞による適切な単語のフィルタリングやLSAの適用が良好な効果をもたらすことを示した。

今後、動画コンテンツへのジャンル付与の精度を向上させるには、テレビ番組のEPGと動画コンテンツのメタデータで使われる語彙集合の違いを埋める必要がある。そのために、外部のコーパスから異なる語での意味の近さなどを得たり、他の関連したコンテンツのメタデータを利用したりする方法が考えられる。さらに同じ表記の単語でも動画配信サービスによって異なった意味で使われることも考えられるため、対策の必要がある。

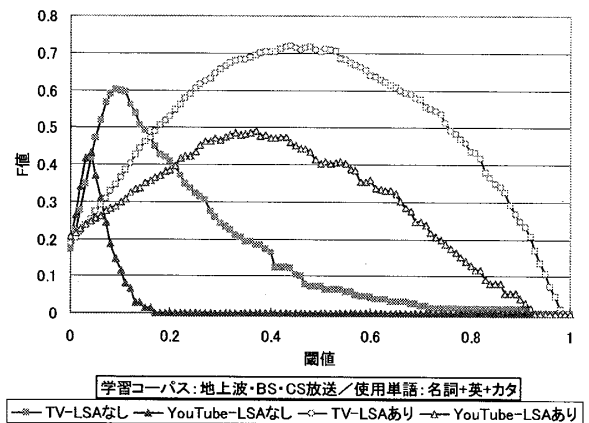


図1 閾値に対するF値の推移

本稿ではインターネット動画の例としてYouTubeを扱ったが、今後は別のコンテンツへの適用も検討していく。加えて、本研究で提案したジャンル自動付与手法は学習コーパスが存在すれば、テレビジャンルに限らず別の既存の分類を付与することが可能であるので、今後はテレビジャンル以外の付与も検討していく。

参考文献

- [1] YouTube, <http://www.youtube.com/>
- [2] T. K. Landauer, P. W. Foltz, and D. Laham: An introduction to latent semantic analysis, *Discourse Processes*, Vol. 25 (2&3), pp. 259-284, 1998.
- [3] Cho K. and Kim J.: Categorization on Hierarchical Category Structure by using ICF(Inverted Category Frequency) Weighting, *Proc. of KISS Conference*, pp. 507-510, 1997.
- [4] Salton, G.: *Automatic Text Processing: The Transformation Analysis and Retrieval of Information by Computer*, Addison Wesley, 1988.