

ニュース特性と個人の嗜好に合わせた動画ニュース推薦システムの提案

今泉 徹†

小川 祐樹†

諏訪 博彦†

太田 敏澄†

電気通信大学†

1. 背景と目的

多くの動画ニュースが毎日放送されている。動画ニュースには、多様なジャンルの情報があり、また映像や音声によって出来事を伝えることができる。このような特徴からテレビニュースは重要な情報源である。

しかしながら、一日に大量のニュースが放送されるため、視聴者は情報を探し出す必要があり、視聴者にとって欲しい情報や有益な情報を入手することは困難である。また、多くのニュースは時事情報であるため、それに対する個人のニーズは時々刻々と変化していく。

このような、課題を解決するために本研究では、ニュース特性と個人の嗜好に合わせた動画ニュース推薦システムを提案することを目的とする。この提案手法により、複数の放送局で 1 日に放送されたニュースから、視聴者が自分のニーズに合わせ短時間で必要な情報を取得できることを解決すべき課題とする。

2. 関連研究

Kamal Ali ら (2004) は、協調フィルタリングとコンテンツフィルタリングによるハイブリッド推薦システムによって、予測評価値の高いアイテムから、ハードディスクに自動録画する TiVo を構築した。これら既存のシステムは、番組単位での推薦を行っている。しかしながら、我々は視聴者に対して短時間での情報取得を支援することを目的としており、番組単位の推薦では必要なニュース情報を的確にとらえられない。

馬ら (2000) はインターネット上で配信されるニュース記事の特性から、3 つの特徴量 (流行度、新鮮度、緊急度) を定義し、あらかじめ入力した嗜好情報を基にした視聴者プロフィールと併用するフィルタリング手法を提案した。しかしながら、適合フィードバックを用いておらず、個人の嗜好の時間変化は考慮していないため、本研究と異なる。

そこで、本研究では、ニュースシーンを対象に、個人嗜好性とニュース特性の指標を組み合わせた推薦システムを提案する。

3. 提案手法

本節では、提案システムの仕組みを説明する。

3.1. 提案手法による推薦システムの概要

本システムの概要を図 1 に示す。本システムでは、まずニュースシーンのメタデータからシーン ID や、ヘッドライン情報、シーンの内容に関するメモ情報を入力データとして扱う。このシーンメタデータから、内容情報を抽出し、またニュース特性に基づいた指標により各指標のスコアを算出する。

次に、視聴者の推薦システムの利用履歴から視聴者プロフィールを構築する。得られた視聴者プロフィールとシーンメタデータからトピック嗜好性とシーン嗜好性の

スコアを算出する。

本提案手法では、トピックに対するスコアを算出し、それにより選択されたトピック内シーンのスコアを算出する。視聴者は、トピック中で興味のあるトピックを選択しその中から見たいシーンを選択し視聴する。

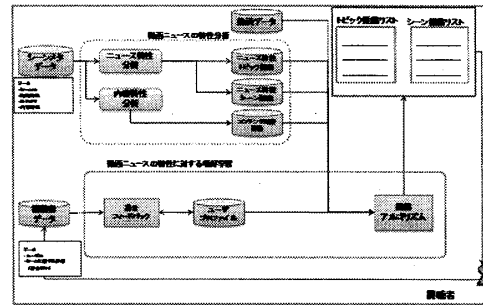


図 1. 提案システムの概要

3.2. 嗜好の抽出：シーン嗜好性とトピック嗜好性

3.2.1 視聴者プロフィール

視聴者プロフィールを作成する手法として、Rocchio (1971) の式による適合フィードバックを用いることで 1 日毎に視聴者プロフィールを更新する。

3.2.2. シーン嗜好性およびトピック嗜好性スコア

コンテンツベースによるスコアは、各視聴者の作成した視聴者プロフィールと各ニュースシーンの特徴量とのコサイン類似度を算出し決定する。

なお、トピックの特徴量は、トピック内のシーンの特徴量より算出する。

トピック A の持つタグの $Tfidf$ は、

$$Tfidf_A = \frac{\sum \text{トピック内のタグ } A \text{ の } Tfidf}{\text{トピック内の合計シーン数}}$$

として算出し、これをトピックの特徴量として扱った。

3.3. 動画ニュースの特性

本節では動画ニュースの特性について述べる。

我々は推薦を行うために考慮すべき動画ニュースの特徴として、ニュース特性を定義する。

各ニュース特性は中村 (2010) の指標を用い、トピック特性を「流行性」と「新規性」で表現し、またニュースシーン特性を「要約性」と「進展性」で表現する。

3.4. ニュース特性と嗜好性のハイブリッド

ニュース特性と嗜好性を組み合わせる手法として、重み付き線形和を用いることにより、トピックのスコアおよび、シーンのスコアを算出する。この手法は、Claypool (1999) の手法を基にした。なお、4 つのニュース特性に関する指標は、中村の手法により算出する。

各指標に対して、初期の重みは各指標とも等価とし計算を行う。

提案手法では、コンテンツベースフィルタリングの

スコアと各ニュース特性指標によるスコアの重み付き線形和を推薦スコアとして用いる。

4. 実験

本節では、提案手法の有効性を評価するために行った従来手法との比較実験について、概要と結果を示す。

本実験では、ニュース指標を推薦に用いた提案手法の有効性を検証することを目的とする。今回の実験では、提案システムとしてトピック嗜好性と流行性指標を用いたトピック推薦を行い、それに基づきシーンの推薦を行った。シーン推薦のスコアはシーンの嗜好性と要約性指標を用いることで算出した。

4.1. データセット

分析には、2009年3月23日から26日までの期間TVで放送されたTV動画シーンデータを用い、アンケートを作成した。アンケートは1日の全シーンからランダムで200シーン選択し3日分作成した。被験者は10代～50代までの男性、女性合計20人である。評価方法は見たい/どちらでもない/見たくない、の三段階とした。

4.2. 検証方法

本実験では、提案手法の推薦リスト上位20シーン作成し、既存手法との推薦結果の違いや、精度比較により行う。推薦リストへの視聴者の評価方法は、アンケート結果と比較することで評価とした。実験には、23日分のアンケート結果を用いて初期の視聴者プロフィールを作成した。初期視聴者プロフィールを用いて24日から26日までの3日間の推薦リストを評価した。

4.3. 結果

表2は、被験者1に対する26日の提案手法による推薦リストである。提案手法では、まず上位20のトピックを算出した。その後各トピック内のシーンの推薦スコアを算出し、最もスコアの高いシーンを各トピックの推薦シーンとして20シーン推薦した。

図2に3日間の二つの手法の精度を比較したものを示す。精度は、

$$\text{精度} = \frac{\sum \text{推薦リスト中の評価値“見たい”}}{\sum \text{評価値数}}$$

により算出した。

表1. 被験者1に対する推薦トピック例【提案手法】

シーンID	推薦順位	ヘッドライン
10006113	1	トヨタ自動車・新型「プリウス」初公開
10006919	2	馬根県出雲市・国内最大級の風力発電施設がくめざましニュースON>北朝鮮・長距離弾道ミサイル「テポドン」発射準備
10005477	3	道ミサイル「テポドン」発射準備
10007536	4	自民党・情報通信活用の経済対策
10006923	5	【野球】WBC日本代表・優勝

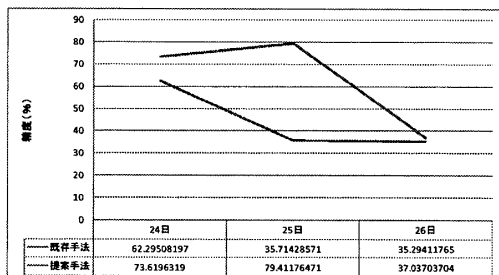


図2. 3日間における提案手法の精度の変化

5. 考察

今泉(2010)の結果より、視聴者が流行性のある話題に対して興味を好む傾向があることを示したが、既存手法では流行性のないニュースばかり推薦される被験者もいた。また、既存手法では複数の同じ内容のニュースが推薦されてしまうという結果になった。これは、複数の番組のシーンメタデータのヘッドライン情報やメモ情報が同様のものであり、それによって推薦されるシーンに偏りが出たと考えられる。しかしながら、今泉の結果より、視聴者は複数のトピックに興味があると言える。すなわち、視聴者はニュースを視聴する際に複数の話題を効率よく視聴する必要がある。

表1に示すように提案手法では、多様な話題のニュースシーンを推薦することができた。また、流行性の最も高かったWBCの話題などが推薦されている。

次に精度について考えると、図2から複数のトピックのシーンを推薦した提案手法の方が既存手法より精度が高いことが示された。

図2の結果から、既存の手法では、二日目以降から精度の減少がみられた。これは、同一内容のシーンの推薦が多く、その結果は視聴者のニーズとは一致しないためと考えられる。それに対して提案手法においては、2日目に精度が上昇し、3日目に精度が低下した。提案手法による2日目の結果は流行性の高いこれらのシーンを推薦したことによって精度が向上したと考えられる。一番内容の盛り上がりつつあるトピックを推薦できたことが25日の推薦精度向上に影響したと考えられる。また、提案手法3日目精度の低下は、前日の流行性の高いニュースシーンを視聴者が見たいのに対して、26日には視聴者が流行性のあるニュースシーンを見たいと思わなかったことによると考えられる。

6. 結論

本研究では、提案手法により多様な情報を推薦することを実現した。それにより推薦リストに対する精度を上昇させることができた。アンケート結果より視聴者は複数の話題を見たいと考えており、そのニーズに対応する手法として、提案手法ではコンテンツベースフィルタリングよりも有効であるという結果を得た。

7. 今後の課題

本実験ではニュース特性の4指標のうち流行性、と要約性のみを用いた検証だったため、残りの新規性、および進展性を組み合わせた推薦により、さらに各視聴者のニーズに合わせた推薦システムを構築する。また、実際に推薦リストに対して評価をしてもらうことで、提案手法のより正確な検証を行う。

参考文献

- [1]Kamal Ali, and Wijnand van Stam. (2004).TiVo: Making Show Recommendations Using Distributed Collaborative Filtering Architecture, Proc.10th Int'l Conf. on Knowledge Discovery and Data Mining, pp.394-401.
- [2]馬強, 角谷和俊, 田中克己(2000), 放送型情報配信システムのための時系列を考慮した情報フィルタリング, 情報処理学会論文誌, vol.41, No. SIG 6 (TOD7), pp.46-57.
- [3]Rocchio, J. (1971) „Relevance Feedback in Information Retrieval, The SMART Retrieval System Experiments in Automatic Document Processing, Prentice Hall Inc.
- [4]中村浩之, 小川祐樹, 諏訪博彦, 太田敏彦 (2010), 動画ニュース選択のためのニュース評価指標の構築, 第16回社会情報システム学シンポジウム
- [5]Mark Claypool, Anuja Gokhale, Tim Miranda, Pavel Murnikov, Dmitry Netes, Matthew Sartin(1999),Combining Content-Based and Collaborative Filters in an Online Newspaper,ACM SIGIR Workshop on Recommender Systems,Algorithms and Evaluation.
- [7]今泉徹, 小川祐樹, 諏訪博彦, 太田敏彦(2010), ニュース特性と個人の嗜好に合わせた動画ニュース推薦システムの提案, 第16回社会情報システム学シンポジウム