

話題性を考慮した動画ニュース推薦システムの提案 A proposal of recommender system for Animation news by extraction of topic

今泉 徹† 小川 祐樹† 諏訪 博彦† 太田 敏澄†
Toru IMAIZUMI Yuki OGAWA Hirohiko SUWA Toshizumi OHTA

1. はじめに

動画ニュースには多様なジャンルの情報があり、記事のみのニュースに比べて、映像や音声を記録しており重要な情報源である。しかしながら、一日に大量の動画ニュースが放送されるために、ユーザは情報を探し出す必要があり、ユーザにとって欲しい情報や有益な情報を入力するのが困難である。また、ニュースは時事情報であるので、個人のニュースに対するニーズは時々刻々と変化していく。

ニュースというコンテンツを対象とした場合、ユーザが本質的に求める情報とは、コンテンツ自身が持つ内容的価値に対するものだけではなく、ニュースという特性が持つ社会の時事情報としての社会的価値を持った情報である。例えば、現在最も放映されている話題のニュース、あるいは、類似した話題のニュース群において新しいニュース、他のニュースに比べてよくまとまっているニュースなどが挙げられる。既存のレコメンドエンジンは、コンテンツの価値をコンテンツ自身で閉じた内容特性しか把握していないため、ユーザがニュースというコンテンツに本質的に望む社会的価値への嗜好をすくいだすことはできていない。

そこで本研究では、ニュースの価値として、コンテンツ自身が示す内容的価値だけでなく、ニュースコンテンツのもつ社会的価値という特性を考慮した推薦システムについて検討することを目的とする。

2. 関連研究

TiVo(2004)は、協調フィルタリングと内容ベースフィルタリングによって予測された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システムである。このシステムにおいて cold-start 問題を解決するため内容ベースフィルタリングを補完的に用いている。また、折原ら(2006)は電子番組表(EPG)と AV 視聴履歴を用いることにより TV 番組推薦システムを構築している。しかしながら、これらのシステムは番組ごとの推薦をすることを目的としており、シーンごとに推薦するシステムは存在しない。ニュースの内容は多岐に及び、番組単位の推薦では、個人の嗜好を的確にとらえられず、有益な情報を提供できないため、本研究ではシーン単位での推薦を行う。

馬ら(2000)は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、ユーザプロフィールと併用したフィルタリング手法を提案している。しかし、適合フィードバックを用いていないため、嗜好性の変化は取り入れられていない。

角谷ら(2002)はマルチチャンネル型ニュース配信システムを提案し、時間情報に基づくクラスタリング方式により情報自身の内容から重要度を計算することを目的として研究を行っており、個人化推薦は行っていない。

推薦システムには大きく分けて内容ベースフィルタリングによる手法と強調フィルタリングによる手法の2種類がある。内容ベースフィルタリングは、内容そのものの属性群からユーザの嗜好を反映させる特徴を抽出し、ユーザの好む内容と類似した内容を推薦する手法である。動画ニュースというフローコンテンツの特性を考えたとき、評価情報が必要な協調フィルタリングよりも内容ベースフィルタリングが適していると考えられる。

本論文では、毎日配信されるニュースの内容を考慮し、社会に対する話題性を考慮した推薦システムを提案するために、内容ベースフィルタリングをベースとした推薦方式を検討する。

3. 動画ニュースの特性

動画ニュースの特性とユーザのニーズおよび社会の関係を図1に示す。社会で発生した様々な事件や出来事は、動画ニュースとして編成され報道される。ユーザは、社会について知りたいというニーズを持っているが、すべての社会を自分自身で把握することは不可能なため、動画ニュースを視聴することで社会について知ることとなる。動画ニュースは日々配信され、その情報は時間とともに変化する。時系列データであるニュースの特性として、発生する内容が予測不可能であることや、時系列的につながっていることが考えられる。視聴者はこれらのニュースに対して自身のもつ嗜好性のみではなく、社会的話題性から観たい情報を選択する。

我々は、ニュースの持つ多面的な価値を表現するために、コンテンツ自身が持つ情報による内容的価値と、ニュースコンテンツとしての社会的価値という観点から動画ニュースを考える。特に社会的価値を「流行性」、「新規性」、「要約性」の3つの視点により表現する。

「流行性」とは、最も放送されているニュースシーン数が多い話題は、社会的に関心度の高い情報であることを示す価値である。「新規性」とは、類似したニュース群において最も付加価値の高いニュースであることを示す価値である。すなわち、同様のニュースの中で新しい

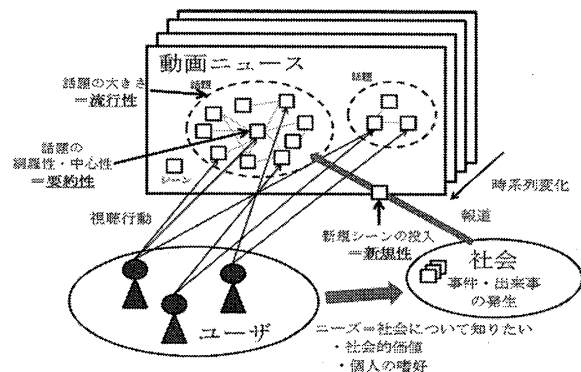


図1. 動画ニュースの特性とユーザのニーズの概念図

†電気通信大学 大学院情報システム学研究科
Graduate School of Information Systems
University of Electro-Communications

内容を含むニュースのことである。「要約性」とは、一つの話題の中で時系列的にそれまでの情報を網羅したニュースであることを表現する価値である。

4. 動画ニュースと社会的価値の関係性の検証

本分析の目的は、視聴者の観たいシーンが嗜好性だけでなく、社会的価値に影響されることを検証することである。本稿では、アンケートによって話題性について分析する。

4.1. データセット

分析には、2009年3月23日から29日までの期間TVキー局で放送されたTV動画のシーンデータを用い、アンケートを作成している。動画のシーン数は一日平均約1,600シーンであり、その中から200シーンをランダムで選択し一日分のアンケートシートとして合計7日分を作成している。被験者は、10代から50代までの男性、女性の合計20人である。評価方法としては各シーンに対して、見たい/どちらでもない/嫌い、の三段階で評価してもらっている。

4.2. 分析結果

アンケート期間中に発生したニュースの中で、最も放映数の多かった5つの話題の中からWBC日本代表と北朝鮮問題に関するアンケート結果と被験者からのコメントを基に流行性、新規性、要約性について述べる。

WBC日本代表に関する分析結果を図2に示す。図2は曜日毎に、全シーンに対してWBC日本代表に関するシーンが放送されたかを示す。火曜日の内容として、日本代表がアメリカを破り決勝進出を伝えるニュースが放送された。また水曜日は韓国を破りWBC二連覇を達成したニュースが放送されている。これにより、火曜日と水曜日のWBC日本代表に関するシーンが他の曜日に比べ多く放送されている。アンケート結果からWBC日本代表のニュースを観たいと評価した結果が火曜日33%、水曜日27%と月曜日の15%に対して増加し、木曜日に再び減少している。これは、図3の火曜日に準決勝でアメリカに勝利したニュース、水曜日に韓国を破りWBC二連覇を達成したニュースが放映された日と一致している。このことから、視聴者がよく報道されるニュースを好んで視聴していることがわかる。

次に、北朝鮮問題に対する分析結果について述べる。ユーザの北朝鮮に関するシーンに対して観たいと評価した数の割合は、木曜23%、金曜27%と月曜から水曜までよりも観たいと回答した結果よりも多かった。これは、木曜日にテポドンの発射準備に関するニュースに対しての報道、金曜日の政府が破壊措置命令を発令したこと、およびイージス護衛艦を日本海に配備したニュースの放映された日と一致している。このことから、新しい情報が動画ニュースとして投入されると視聴者がそのニュースを好んで視聴していることがわかる。

週末のシーン数は平日のシーン数よりも少ない。また、週末と平日の1シーンあたりの放送時間を比較すると、週末のほうが長いことが確認されている。これは、週末にはその一週間にあった出来事をまとめて報道する番組が増加するためと考えられる。被験者からも土日においては今までの放送されたニュースの内容をまとめた動画ニュースが多くなっているとコメントが得られている。こ

のことから、視聴者は平日と週末で異なる視聴行動を行っていることがわかる。

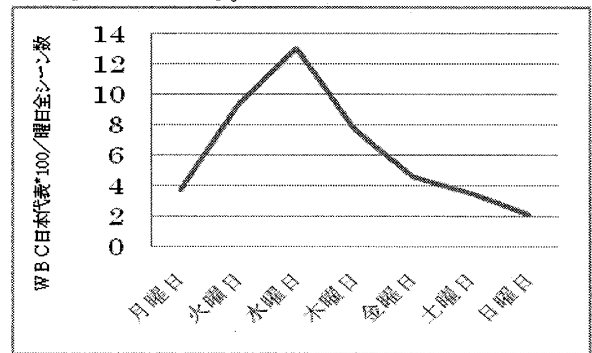


図2. WBC日本代表放送数の全シーンに対する割合

5. 社会的価値をふまえた動画ニュース推薦

4節の検証により、視聴者がよく報道されるニュースを好んで視聴していること、視聴者が新たに投入されたニュースを好んで視聴していること、週末と平日では視聴行動に差があり週末にはまとめられたニュースを希望していることが確認されている。

これまでの推薦システムは、個人の嗜好に合った推薦をすることを目的に構築されている。しかし、動画ニュースという日々刻々と変わり、個人の嗜好とは関係なく、社会的価値が重視されるコンテンツには、不十分であると考えられる。例えば、野球や北朝鮮について興味がない人であっても、周りの人々が注目しているWBCに関する話題や、自身の危険に関係するミサイル情報などには関心を持つことは十分に考えられ、それらの動画ニュースを推薦できる推薦システムが必要となると考える。

以上の結果から、ニュースの社会的価値を表す流行性と新規性を指標として加えた新たなユーザプロファイルを提案する。提案式は、

$$q' = \eta \left(\alpha q + \frac{\beta}{|D_R|} \sum_{d_i \in D_R} d_i - \frac{\gamma}{|D_N|} \sum_{d_i \in D_N} d_i \right) + \phi * \max(\sigma * fresh, \varepsilon * trend * abst)$$

となる。ユーザプロファイルを作成する式として、Rocchioの式に流行性、新規性、要約性の指標であるfresh, trend, abstを付加する。具体的な指標の定義については今後の課題である。

参考文献

- [1] Kamal Ali, and Wijnand van Stam. "TiVo: Making Show Recommendations Using Distributed Collaborative Filtering Architecture", Proc.10th Int'l Conf. on Knowledge Discovery and Data Mining, pp.394-401. (2004)
- [2] 折原良平, 村上知子, 坂本典哉, 堀口健生: 『コンテンツの高精度推薦技術によるデジタル機器の価値向上』 東芝レビュー, Vol. 61, No.12, pp.13-17. (2006)
- [3] 馬強, 角谷和俊, 田中克己: 『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(2000).
- [4] 角谷和俊, 松本好市, 高橋美乃梨, 上原邦昭: 『マルチチャンネル型ニュース配信システムのための時系列クラスタリング』 情報処理学会論文誌, vol.43, No.SIG 5(TOD 14), pp.87-97. (2002)
- [5] Rocchio, J.: "Relevance Feedback in Information Retrieval," The SMART Retrieval System Experiments in Automatic Document Processing, Prentice Hall Inc(1971).
- [6] 折原良平, 村上知子, 坂本典哉, 堀口健生: "コンテンツの高精度推薦技術によるデジタル機器の価値向上," 東芝レビュー, Vol. 61, No.12, pp.13-17(2006)