

自然言語処理的アプローチによる テレビ視聴データの解析

土橋 諒太, 陳 晨, 三浦 真和, 中田 和秀

1. はじめに

全録画レコーダー, オンライン配信, スマートテレビなどの普及により, ユーザーは時間的, 空間的制限を超えたテレビ視聴が可能となった. 一方, 番組製作側は視聴デバイスやインターネットから収集したユーザーの基本情報や視聴履歴などのデータの分析が容易になっている. このような状況下で, 本研究は視聴履歴やユーザーと番組の属性情報を活用して, ユーザーの層別, 番組推薦, CM 最適化などさまざまなタスクにおいて役立つ, ユーザーと番組の特徴を抽出する汎用的な手法を提案する. このとき, 自然言語処理のテクニックを複数導入している.

本研究では, まずテレビ視聴の時系列性を利用するため, Word2Vec [1] と Doc2Vec [2] を適用してユーザーと番組の特徴ベクトルを生成した. 次に, 正解ラベルを用いた学習によって先に生成した特徴ベクトルの改善を行った. 改善には, ニューラルネットワークを利用し, さらに注意機構 [3] を導入した. また, ユーザーと番組固有の属性情報の埋め込みを行った. 実データを用いた検証により, 提案手法で作成した特徴量は, クラシクな特徴抽出手法である Probabilistic Latent Semantic Analysis (以下 PLSA) [4] や Non-negative Matrix Factorization (以下 NMF) [5] より優れていることが確認できた. また, ユーザー特徴量のクラスタリングや番組特徴量の類似度分析などを通して, 実用につながる知見が得られた.

2. データの概要と特性

本研究では, 経営科学系研究部会連合協議会主催, 平成 30 年度データ解析コンペティションで提供された

データを使用する. まず, 提供されたデータの内, 放送時間に沿った視聴 (リアルタイム視聴) と録画再生 (タイムシフト視聴) の視聴履歴の特性について考察する.

2.1 リアルタイム視聴履歴の時系列性

ユーザーの視聴行動は生活習慣やライフスタイルを反映しており, それらの関係性についてさまざまな研究がなされてきた. たとえば, 木村 [6] では視聴時間について, 年齢層別の視聴習慣の相違について分析している. また, 岡崎と井上 [7] では, ユーザーの視聴履歴から, 視聴パターンなどの時系列特徴を抽出し, 隠れマルコフモデルを用いた分析手法を提案している.

本研究では, リアルタイム視聴履歴の局所的な時系列性と中長期に渡る視聴系列について着目する. 番組の局所的な時系列性とは, リアルタイムで視聴した番組とその近くで視聴した番組との関連性のことであり, 番組を特徴付けるものと考えられる. 一方, 月や年などの中長期的な時間幅におけるユーザーの視聴系列に関しては, ライフスタイルや趣味嗜好などの深層的な要因により, 個人それぞれの時系列的な視聴パターンが顕在化し, ユーザーの特徴が表現されていると考えられる.

2.2 タイムシフト視聴とユーザー嗜好の関連性

前述のリアルタイム視聴履歴はユーザーの趣味嗜好を含んだ大量のデータである. しかし, ユーザーの内的・外的要因と思われる理由によって起こるザッピングや視聴停止などにより, ノイズデータとすべきデータが多いことも確認できた. 一方, タイムシフトで視聴する番組は, ユーザーが興味関心をもつ番組のみを録画・視聴するため, 相対的に数量は少ないが, ノイズが少ないデータと考えられる. したがって, リアルタイム視聴履歴とは別に, タイムシフト視聴履歴を活用することで, 分析精度の向上が期待できる.

表 1 はリアルタイム視聴とタイムシフト視聴におけるカテゴリ別視聴割合を表している. ユーザーの嗜好が表れやすいと考えられるドラマ, アニメ, 芸能カテゴリで, タイムシフト視聴内での視聴割合が多くなっている. 一方, ながら見が多いと考えられる報道カテ

どばし りょうた, ちん しん, みうら まさかず,
なかた かずひで
東京工業大学工学院経営工学系
〒 152-8552 東京都目黒区大岡山 2-12-1
受付 19.7.25 採択 19.11.9

表 1 番組カテゴリ別視聴割合

	リアルタイム	タイムシフト
報道	0.18	0.02
教育	0.34	0.11
音楽	0.02	0.02
ドラマ	0.07	0.34
アニメ	0.02	0.07
映画	0.01	0.04
スポーツ	0.06	0.02
芸能	0.31	0.38

表 2 リアルタイム・タイムシフト視聴時間割合

	リアルタイム	タイムシフト
視聴時間割合	0.42	0.58

ゴリの視聴割合は、リアルタイム視聴と比べてタイムシフト視聴では極端に少なくなっている。

表 2 はリアルタイム視聴とタイムシフト視聴に対して、1分に満たない視聴時間のデータを削除したうえでユーザーが番組を視聴した場合の視聴時間割合（番組視聴時間/番組放映時間）を集計した平均値である。タイムシフト視聴時間割合はリアルタイム視聴時間割合に比べて約 0.16 高いことから、ユーザーはタイムシフト視聴番組により強い関心をもってしていると解釈できる。

本研究では、タイムシフト視聴がより正確にユーザーの嗜好を反映しているという推測に基づき、抽出された特徴量の良さを評価すると同時に、特徴量の改善も行うタイムシフト予測タスクを設計した。

3. 関連研究

本研究の関連研究として、時系列性をもったデータに対して低次元空間への埋め込みを応用した研究について説明する。

自然言語処理における単語の埋め込みとは、単語を低次元ベクトル空間の元として表現する手法である。2013年に提案された Word2Vec による単語の埋め込みは単語の局所的な時系列性を考慮しており、大きな成功を収めている。その後、Word2Vec のアイデアを他のドメインに適用し、特徴抽出を行う研究も盛んに行われている。たとえば、Barkan and Noam [8] は、ユーザーの注文履歴から商品の特徴ベクトルを得る Item2Vec を提案した。Grbovic and Haibin [9] は、Airbnb のユーザーの予約に関する閲覧履歴を利用して、ユーザータイプと物件タイプを同じ空間上に埋め込む手法を提案した。これらの手法は正解ラベルなしのデータから

特徴ベクトルを得ることができ、データの収集、前処理などのコストが少ないという利点をもつ。しかし、正解ラベルなしのデータで学習した単語の埋め込みを用いて実際にダウンストリームのタスクを解くときには、ファインチューニングが必要となることが多い [10]。

それに対して、正解ラベルを用いた学習による埋め込みは、ファインチューニングなしでドメインに特化した特徴ベクトルが得られる。たとえば、Ni et al. [11] は Taobao という中国のネットショッピングサービスの全ユーザーの行動履歴を対して、正解ラベルありマルチタスクによってユーザーと商品の特徴ベクトルの生成に成功した。それらの特徴ベクトルは Taobao 内部の推薦システムに利用されている。ただし、正解ラベルを用いた学習による埋め込みは、大規模なラベル付き教師データと長い学習時間が必要になるという問題が生じる [12]。

4. 提案手法

本研究は、正解ラベルなしデータを用いた学習と正解ラベルを用いた学習の二種類の埋め込みの利点を融合し、ユーザーと番組の特徴ベクトルを構築する手法を提案する。全体の枠組みは次のようなものである。まず Word2Vec によって番組の特徴ベクトル、Doc2Vec によってユーザーの特徴ベクトルを生成する。次に、タイムシフト予測タスクを通して、生成した特徴ベクトルを改善する。このとき、予測器としてニューラルネットワークを用いており、よりよい特徴ベクトルを得るためいくつかの工夫を加えた。

4.1 節では特徴ベクトルの生成、4.2 節では特徴ベクトルの改善について説明を行う。

4.1 特徴ベクトルの生成

Word2Vec と Doc2Vec は自然言語処理の分野で近年さかんに利用されている手法である。Word2Vec は大量のテキストデータに対し、分布仮説のもと、局所的な共起関係から単語の特徴抽出を行う。分布仮説とは、似たような文脈を共有する単語が似ているはずという仮説である。たとえば、「私は___を飲むのが好きだ。」という文脈では、コーヒー、ビール、梅酒、ウイスキーなど飲み物の単語が入りやすい。このような局所的な共起関係に近い単語は、似た特徴をもっていると捉える。今回対象とするテレビ番組の視聴系列においても、局所的な視聴番組の共起関係から番組の特徴が抽出できる可能性がある。たとえば、図 1 のような二つの視聴系列があった場合、ドラマ B とドラマ D は似た特徴をもっている可能性が高い。

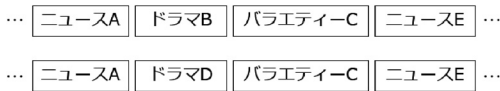


図1 視聴系列の例

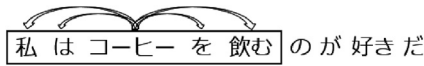


図2 Skip-Gram のイメージ図

このため、各番組を単語、視聴順に記録された各ユーザーの視聴系列を文章とみなし、Word2Vecを適用することで番組の特徴抽出を試みた。

Word2Vecには、Skip-GramとCBOW (Continuous Bag of Words) という二種類の学習アルゴリズムが存在する。本研究では先行研究 [8, 9] に従い Skip-Gramを用いた。Skip-Gramでは、たとえば「私はコーヒーを飲むのが好きだ」という文において、「コーヒー」という中心語に対し、文脈語（ここでは前後2単語）となる「私」「は」「を」「飲む」が出現する確率を考える。（図2を参照）。

中心語 w_t に対し文脈語 w_{t+j} が出現する確率を $\mathbb{P}(w_{t+j}|w_t)$ と表記する。Skip-Gramでは、単語 $w \in W$ (W は全語彙の集合) の表現を、特徴ベクトル \mathbf{v}_w と補助ベクトル \mathbf{u}_w で行う。そして、 $\mathbb{P}(w_{t+j}|w_t)$ は softmax 関数を利用し、次のように表せると仮定する。

$$\mathbb{P}(w_{t+j}|w_t) = \frac{\exp(\mathbf{u}_{w_{t+j}}^T \mathbf{v}_{w_t})}{\sum_{w \in W} \exp(\mathbf{u}_w^T \mathbf{v}_{w_t})} \quad (1)$$

単語の系列 $\{w_1, w_2, \dots, w_{T-1}, w_T\}$ による学習は、次のような対数尤度が最大となるような $\mathbf{v}_w, \mathbf{u}_w$ ($w \in W$) を見つけることになる。

$$\max \frac{1}{T} \sum_{t=1}^T \sum_{-m \leq j \leq m, j \neq 0} \log \mathbb{P}(w_{t+j}|w_t) \quad (2)$$

そして得られた \mathbf{v}_w ($w \in W$) を各単語の特徴ベクトルとする。本研究では、全ユーザーのリアルタイム視聴系列を使って Word2Vec で学習させることにより、番組ごとの特徴ベクトルを得ている。

一方、中長期にわたるリアルタイムの視聴系列はユーザーの嗜好趣味に強く依存している。よって、そこから時系列性をもつ視聴パターンを利用し、ユーザーの趣味嗜好を反映した特徴量の抽出が可能であると期待できる。そこで、Doc2Vec を利用したユーザーの特徴抽出が考えられる。Doc2Vec は Word2Vec を文章単位に拡張したモデルであり、可変長の文章を入力し

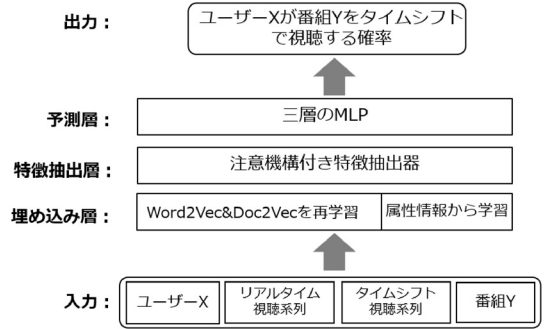


図3 タイムシフト視聴予測のスキーム

固定長の文章特徴ベクトルを生成することができる。PV-DM (Distributed Memory Model of Paragraph Vectors), PV-DBOW (Distributed Bag of Words version of Paragraph Vector) という二種類の学習アルゴリズムが存在するが、本研究では内部の文脈構造情報を捉えつつ、文章の意味ベクトルを生成できる PV-DM を利用する [2]。PV-DM は Word2Vec と違い、文章の ID と前の文脈を利用して、次に来る単語を予測するというタスクを解く。このため、文章内部の単語間の関係を考慮でき、長期的な時系列情報を取り入れられる。学習終了後、文章の ID に対応するベクトルを文章の特徴ベクトルとして用いる。Word2Vec 適用時と同様に、各番組を単語、各ユーザーの視聴系列を文章とみなし、Doc2Vec を適用することによって、各ユーザーに対する特徴ベクトルを得ることができる。

4.2 特徴ベクトルの改善

本節では、Word2Vec と Doc2Vec によって生成した特徴ベクトルを正解ラベルを用いた学習であるタイムシフト予測タスクによって改善する手法を説明する。

タイムシフト予測タスクとは、[ユーザー X, ユーザー X のリアルタイム視聴系列, ユーザー X のタイムシフト視聴系列, 番組 Y] を入力とし、ユーザー X がターゲット番組 Y をタイムシフトで視聴するかどうかを予測するタスクである。予測器としてニューラルネットワークを用いた。

今回使用するニューラルネットワークのスキームを図3で示した。埋め込み層では、入力したユーザー、リアルタイム視聴系列の各番組、タイムシフト視聴系列の各番組、ターゲット番組を、対応した特徴ベクトルに変換する。このとき、再学習と属性情報の埋め込みを行う。次の特徴抽出層では、注意機構を用いて、リアルタイム視聴系列の特徴ベクトルをまとめたリアルタイム特徴ベクトル、タイムシフト視聴系列の特徴ベクトルをまとめたタイムシフト特徴ベクトルを生成する。

表 3 属性情報

属性情報	
ユーザー	世帯主・主婦コード, 性別, 未既婚, 年齢, 職業コード
番組	局コード, 曜日コード, 大分類コード, 中分類コード, 分類コード, 放送形式

最後の予測層では, ユーザーの特徴ベクトル, リアルタイム特徴ベクトル, タイムシフト特徴ベクトル, ターゲット番組の特徴ベクトルを使い, 3層の全結合ニューラルネットワークにより, ユーザー X がターゲット番組 Y を視聴する確率を出力する (学習時間を考慮して再帰型ニューラルネットワークは使用しない).

タイムシフト予測タスクの学習終了後に, 埋め込み層に保存されている重みをユーザー特徴ベクトルと番組特徴ベクトルとする.

埋め込み層での「再学習」と「属性情報の追加」, 特徴抽出層での「注意機構の導入」について, 順に説明を行う.

4.2.1 再学習

2.2節で述べたように, タイムシフトで視聴された番組はユーザーの嗜好を強く反映していると考えられる. よって, タイムシフト視聴を正確に予測できる特徴ベクトルは, ユーザーの嗜好を反映した良質な特徴ベクトルである可能性が高い. その性質を利用して, 特徴ベクトルの改善を行う.

まずニューラルネットワークの埋め込み層の初期値として, Word2Vec と Doc2Vec による学習で得られた特徴ベクトルを代入する. そして, タイムシフト予測タスクを通じて学習させることにより, 誤差逆伝播によって勾配情報が最初の埋め込み層まで遡り, ユーザーと番組の特徴ベクトルを修正することができる. これを再学習と呼ぶ. Bidirectional Encoder Representations from Transformers (以下 BERT) [13] などの正解ラベルがないデータから学習を行う手法では, 特定のタスクを精度よく解くためにファインチューニングを行うが, 本研究ではタイムシフト予測タスクを通して, より良質の特徴ベクトルを得るところが新しい.

4.2.2 属性情報の追加

提供データの中で, ユーザーと番組の属性情報に関連する項目を表 3 にまとめた.

これらの属性情報はユーザーと番組の特徴を含むため, 視聴履歴とは別種の特徴が抽出可能であると考えられる. よって, これらの属性情報に対し, ユーザーや番組の埋め込みと同様にして低次元空間への埋め込

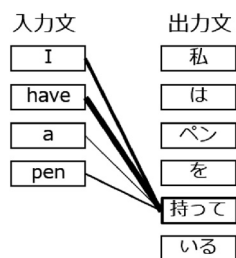


図 4 機械翻訳における注意機構のイメージ図. 線の太さは関連度を表す.

みを行った. 埋め込みを行う際, まずユーザーと番組の属性情報をダミー変数またはカテゴリ変数で表して One-hot のベクトルに変換し, ニューラルネットワークの埋め込み層により埋め込みを行った. その結果得られた特徴ベクトルは, 埋め込み層において視聴系から生成した特徴ベクトルと連結させ, 最終的なユーザーと番組の特徴ベクトルとした.

4.2.3 注意機構の導入

ここでは, 特徴抽出層で行うリアルタイム特徴ベクトルとタイムシフト特徴ベクトルの作成法について説明する. まず, あるユーザーのリアルタイム視聴系列 $\{x_1, x_2, \dots, x_n\}$ について考える. 埋め込み層によって, 各番組に対応した特徴ベクトル系列 $\{v_1, v_2, \dots, v_n\}$ が得られる. この可変長な特徴ベクトル系列から, 予測層 (全結合) の入力となる固定長の特徴ベクトルを作成する必要がある. 最もシンプルな方法として, 下記のように特徴ベクトル系列の平均を取ることが考えられる.

$$\tilde{v} = \frac{1}{n} \sum_{i=1}^n v_i \quad (3)$$

しかし, この手法では視聴系列中の番組を均等に扱っている. 実際は, 番組によってターゲット番組との関連性は異なる. したがって, 特徴ベクトル系列の平均を取るという方針は重要な情報を減少させている可能性がある.

そこで, われわれが取り上げたのは, 近年機械翻訳分野で注目を浴びている注意機構 [3] という手法である. 注意機構とは, 図 4 のように入力系列の各単語に対してアテンション重みを計算し, 重要な単語を自動的に抽出できる仕組みである.

本研究では, いくつか提案されている注意機構の中でも, グローバルソースターゲット型注意機構 [3] を使用する. この注意機構を機械翻訳で使われる Encoder-Decoder 構造を例として説明する. Encoder-Decoder

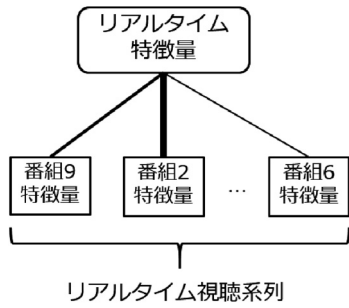


図5 注意機構を用いて、リアルタイム視聴系列から系列特徴量を生成するイメージ図。線の太さは番組の重みを表す。

構造は、まず Encoder によって入力文の単語を逐次解析する。次に、Decoder は Encoder から得られた各単語の情報を用いて、出力文の単語を逐次生成する [14]。Encoder における各時刻の隠れ状態の系列（可変長）を $\bar{H} = \{\bar{h}_1, \bar{h}_2, \dots, \bar{h}_n\}$ としよう。Decoder のある時刻の隠れ状態 h_t を用いて、Encoder のどの時刻をどれくらい注目するかを決める重み α_t を計算する。たとえば、時刻 s における重み $\alpha_t(s)$ は、softmax 関数を使い下記のように計算する。

$$\alpha_t(s) = \frac{\exp(\text{score}(h_t, \bar{h}_s))}{\sum_{i=1}^n \exp(\text{score}(h_t, \bar{h}_i))} \quad (4)$$

ここで、 $\text{score}(\cdot)$ は、二つのベクトル間の関連性を評価する関数であり、大きい方がより関連していることを意味する。なお、計算効率のため、スコア関数として内積を使用することが多い。すると、Encoder の隠れ状態の系列 \bar{H} に対して、加重平均 c_t を計算することができる。

$$c_t = \sum_{s=1}^n \alpha_t(s) \bar{h}_s \quad (5)$$

この機構により、 h_t と \bar{H} の各要素の関連性を自動的に組み込むことが可能となる。

本研究では、このアイデアを流用することにより、ユーザーのリアルタイム視聴系列に対し、ターゲット番組との関連性を自動的に組み入れたリアルタイム特徴量を作る（図5のイメージ）。

具体的には、特徴ベクトル系列 $V = \{v_1, v_2, \dots, v_n\}$ とターゲット番組の特徴ベクトル q に対して、各 $s \in \{1, \dots, n\}$ に対応する重みを以下のように与える。

$$\alpha_q(s) = \frac{\exp(q^T v_s)}{\sum_{i=1}^n \exp(q^T v_i)} \quad (6)$$

そして、下記のような加重平均によってリアルタイム特徴ベクトルを作成する。

$$\tilde{v} = \sum_{s=1}^n \alpha_q(s) v_s \quad (7)$$

同様の手順によって、タイムシフトの特徴ベクトル系列から、ターゲット番組との関連性を組み入れたタイムシフト特徴ベクトルも作成することができる。また、ターゲット番組との関連性ではなく、ユーザーとの関連性を考慮してリアルタイム特徴ベクトルとタイムシフト特徴ベクトルを作成できる可能性がある。よって、特徴抽出層では4本の特徴ベクトル（ターゲット番組とユーザーそれぞれに関連したリアルタイム特徴ベクトルとタイムシフト特徴ベクトル（計4本））を計算し、予測層に送っている。

5. 実データを用いた検証

ここでは、まず4.1節で説明した Word2Vec と Doc2Vec による特徴量生成の有効性を検証する。次に4.2節で説明したニューラルネットによる再学習、属性情報の追加、注意機構の導入に関する特徴量改善の効果を測定する。

5.1 実験データ

平成30年度データ解析コンペティションで貸与された VR-CUBIC 提供データのリアルタイム・タイムシフト視聴履歴およびユーザー・番組の属性情報を用いて検証を行う。実験では、2017年4月3日～2017年12月31日を訓練期間、2018年1月1日～2018年2月28日を検証期間とした。ユーザーは、訓練期間・検証期間共に視聴履歴が残っている6,756世帯を対象とした。リアルタイム視聴履歴は、ザッピングなどのノイズを低減するため、視聴時間割合が各ユーザー各番組で2割以上のデータのみを使用し、タイムシフト視聴履歴については、すべてのデータを使用した。

5.2 タイムシフト予測タスク

2.2節で述べたように、タイムシフト視聴履歴はリアルタイム視聴履歴よりも正確にユーザーの嗜好を反映していると考えられる。よって、タイムシフトの視聴予測の精度によって、生成した特徴量の有効性を測ることにする。

タイムシフト予測タスクの実験設定について説明する。まず訓練データと検証データの視聴履歴をユーザーごとに2週間ずつ分割する。次に、分割されたデータについて、1週間目のタイムシフト視聴番組および1・2週間目のリアルタイム視聴番組を予測に用い

表4 Word2Vec・Doc2Vecの有効性の検証

	正答率	適合率	再現率	F 値
PLSA	0.61	0.46	0.87	0.60
NMF	0.63	0.47	0.83	0.60
W ¹	0.78	0.63	0.84	0.72
D ²	0.65	0.48	0.83	0.61
W・D ³	0.81	0.68	0.83	0.74

¹ Word2Vec のみで番組のベクトルを学習

² Doc2Vec から学習した番組とユーザーのベクトルを利用

³ Word2Vec から番組ベクトル, Doc2Vec からユーザーベクトルを利用

るデータとし, 2週間目のタイムシフト視聴番組を予測対象とする. ここで, 予測対象には実際にタイムシフトで視聴された番組のみが含まれているため, 2週間目に放送されており, 2週間目のタイムシフト視聴番組に含まれていない番組を負例としてサンプリングし, 正例:負例 = 1:2 になるまで予測対象に加えた.

5.3 Word2Vec・Doc2Vecの有効性

代表的な特徴抽出法であるNMF, PLSAを比較手法とし, Word2Vec・Doc2Vecによる特徴ベクトル生成の有効性を検証する. Word2Vec・Doc2Vec, NMF, PLSAを学習期間のリアルタイム視聴履歴に適用し, それぞれの手法について20次元のユーザー特徴ベクトルと20次元の番組特徴ベクトルを生成した. 次元数やパラメータは複数試したが, 事前実験でタイムシフト予測タスクの精度が良かったものを用いた. Word2Vec・Doc2Vecの学習パラメータは, エポック数を3, ウィンドウサイズを5, ネガティブサンプル数を10としている.

NMFとPLSAはユーザーごとの視聴番組行列を用いて, 番組とユーザーの特徴ベクトルを計算する手法である. まず学習期間のリアルタイム視聴履歴からユーザーごとに各番組の視聴回数をカウントすることでユーザー数×番組数のサイズとなる行列を作成した. 次に, ニュースなどの放送回数が多い番組の影響力を減らしてユーザーごとに特徴的な番組を強調するため, tf-idf¹に変換して視聴番組行列を作成している [15]. よって, 視聴履歴の時系列性は考慮されていないことに注意する.

この実験では, 予測器として単純な3層の全結合ニューラルネットを使用した(4.2節で述べた再学習・属性情報追加・注意機構は導入していない). リアルタ

¹ tf-idfはTerm Frequency(単語出現頻度)とInverse Document Frequency(逆文書頻度)の積である.

表5 特徴ベクトル改善手法の検証

再 ¹	属 ²	注 ³	正答率	適合率	再現率	F 値
			0.81	0.68	0.83	0.74
○			0.82	0.70	0.80	0.75
	○		0.82	0.69	0.85	0.76
		○	0.85	0.75	0.85	0.80
○	○		0.85	0.73	0.86	0.79
○		○	0.88	0.78	0.88	0.83
	○	○	0.91	0.83	0.90	0.87
○	○	○	0.92	0.85	0.93	0.89

¹ 再学習, ² 属性情報, ³ 注意機構

イム特徴ベクトルとタイムシフト特徴ベクトルは, それぞれ特徴ベクトル系列の平均としている. 出力層にはシグモイド関数を使用し, 検証時の予測視聴確率の閾値は0.5とした. 誤差関数はクロスエントロピー誤差を使用した.

以上の設定で, 検証期間に対するタイムシフト視聴予測の結果が表4である. Word2Vec・Doc2Vecは正答率, 適合率, F値が比較手法と比べて最も高く, F値については比較手法が0.6であるのに対して, 0.74と大きく上回っている. この結果は, Word2Vec・Doc2Vecによる時系列性を考慮した特徴量の生成が有効であることを示している.

5.4 特徴ベクトルの改善の効果

再学習, 属性情報追加, 注意機構の導入という三つの特徴ベクトルの改善手法について, それらの有効性を検証する. 5.3節と同様に, 検証はタイムシフト予測タスクにより行う. ユーザーと番組の属性情報を追加する場合, それぞれ10次元のベクトルに埋め込んでいる. 特徴抽出層で利用する際は, 視聴系列から生成した20次元のベクトルと連結させ30次元の特徴ベクトルとしている.

三つの改善手法をそれぞれ導入し, 検証期間でタイムシフト視聴予測を行った結果が表5である. 但し, 再学習なしの場合, Word2VecとDoc2Vecで得られた特徴ベクトルを改善させることはできない. したがって, 表5の数値は参考値として載せている. 再学習のみを導入するだけでは余り効果がなく, 属性情報や注意機構と併用することで改善効果が出ることが見てとれる. また, すべての改善手法を導入した場合, 正答率, 適合率, 再現率, F値のすべての評価指標で最も高い. 特にF値についてはベースラインが0.74であるのに対して0.89まで改善されており, すべての改善手法を組み込んだ提案手法の有効性が確認できる.

表 6 ユーザー特徴量のクラスタリング

世帯数	特徴
1	746 流行ドラマを好む, 中・高校生が多い
2	537 教育テレビ・アニメを好む, 子育て世帯
3	726 夜の報道番組を好む
4	591 フジテレビを好む
5	829 NHK 総合と報道番組を好む
6	680 局ジャンルに偏らない, 高年層
7	557 視聴時間が長い
8	2084 視聴時間短い

6. 特徴量の活用

本節では、ユーザー特徴ベクトルに対するクラスタリングや番組特徴ベクトルに対する類似度分析によって、ユーザーの嗜好や番組の視聴傾向を適切に反映した有用な特徴量が得られていることを確認する。また、これらの分析を踏まえて、提案手法のテレビ番組制作・CM 戦略やテレビ番組推薦への応用可能性について考察する。6 節で用いる番組とユーザー特徴ベクトルとは、5.4 節のすべての改善手法（再学習、属性情報、注意機構）を用いた場合と同様の設定で学習を行い、モデルの埋め込み層の重みとして保存されている各ユーザーと番組に対応するユーザー特徴ベクトルと番組特徴ベクトルを用いている。

6.1 ユーザー特徴ベクトルのクラスタリング

ユーザーの特徴ベクトルを使い、K-means 法でユーザー 6,750 世帯を 8 個のクラスタに分割した。クラスタ数は意味解釈の納得性から探索的に 8 個に決定した。そして、世帯情報、アンケートデータ、視聴履歴を用いて、各クラスタの該当世帯数、特徴についてまとめたものが表 6 である。視聴局や視聴ジャンルなどの視聴傾向を含む多面的なユーザー特性情報に基づいてクラスタリングできていることが確認できる。

クラスタリングの応用について考察する。現状のテレビ番組制作・CM 戦略におけるユーザーのセグメンテーションには性別および年齢（F1、M1 など）という限定的なユーザー属性が用いられることが多い。しかし、提案手法によるクラスタリングによって、視聴傾向を含む多面的なユーザー特性情報を活用できる可能性がある。

一例として、日本テレビの朝のニュース番組 ZIP! を用いて、性別および年齢による層別と提案手法による層別の違いを説明する。ZIP! に対する、性別および年齢による層別割合を表したものが表 7、提案手法による層別割合を表したものが表 8 である。

表 7 ZIP! 視聴者の性別/年齢による層別割合 (%)

F1 ¹	F2 ²	F3 ³	M1 ⁴	M2 ⁵	M3 ⁶	T ⁷
15	16	16	13	12	9	19
¹ 女性 20 歳～34 歳			² 女性 35 歳～49 歳			
³ 女性 50 歳～			⁴ 男性 20 歳～34 歳			
⁵ 男性 35 歳～49 歳			⁶ 男性 50 歳～			
⁷ 男女 13 歳～19 歳						

表 8 ZIP! 視聴者の提案手法による層別割合 (%)

1	2	6	7	その他 (3, 4, 5, 8)
57	14	8	7	14

表 9 相棒 (刑事ドラマ) の類似番組

類似番組	ジャンル	類似度
ドクター X ¹	医療ドラマ	0.957
臨場	刑事ドラマ	0.927
スーパー J チャンネル	報道	0.912
おかしな刑事 ²	刑事ドラマ	0.901

¹ ドクター X 外科医大門未知子

² おかしな刑事・居眠り刑事とエリート警視の父娘捜査

性別および年齢による層別割合では、どちらかといえば女性と若者がメインの視聴者であることがわかるが、提案手法による層別割合ではクラスタ 1、クラスタ 2 がメインであり、若く流行に敏感な視聴者や子育て世帯が多いと解釈できる。したがって、提案手法を用いると、子供に人気のタレントの起用や流行商品の CM の放送などのテレビ番組制作・CM 戦略が有効であるという示唆が得られる。さらに、テレビ番組制作を行う際の戦略として、既存の年齢と性別による区分だけではなく、提案手法によるクラスタリングを考慮して番組のターゲティングを行うことで、新規の視聴者獲得や番組内容向上につながる可能性がある。たとえば、若者向けに感度の高い層に向けて番組を作成する際に、クラスタ 1（流行ドラマを好む、中・高校生が多い）の視聴率が高い番組を参考とすることが考えられる。

6.2 番組特徴ベクトルの類似度分析

刑事ドラマである「相棒」の特徴ベクトルに対し、番組間の Cosine 類似度を計算し、類似度が上位となった 4 番組をまとめたものが表 9 である。刑事ドラマである「相棒」の類似番組として、同ジャンルの刑事ドラマが多いが、医療ドラマ・報道番組も現れる。したがって、提案手法により生成された番組特徴ベクトルは、ジャンルに収まらないユーザーの視聴傾向に基づく多面的な番組の類似性を含んでいるといえる。この

類似度が高い番組のユーザー層は一致する可能性が高い。番組特徴量の類似度分析について二つの応用を挙げたい。一つ目の応用は、番組配信サービスにおいて、類似度を新たな番組推薦の基準とすることで、ジャンルや人気ではなく、各ユーザーの視聴傾向を考慮して推薦を行うことである。協調フィルタリングや教師あり学習で視聴確率を予測し、その確率が高い番組を推薦する方法では推薦されなくても見るような人気番組が推薦されやすい。提案手法を用いて、ユーザーが好む番組と類似度が高い番組、またはあるユーザーの類似ユーザーが好む番組を推薦することで、推薦されなければ見ないような番組を推薦できる可能性がある。二つ目の応用は、CM 広告主が類似度を CM 投入の選択指標とし、類似番組に連続的に CM を投入することで、効率的に CM 認知率の向上を図ることである。

7. 終わりに

本研究はテレビ視聴データに対し、ユーザーと番組の特徴を抽出するための手法を提案した。

提案手法は、自然言語処理分野のテクニックを利用することにより、データの時系列性を考慮した分析となっている。また、正解ラベルなしデータを用いた学習の結果を正解ラベルを用いた学習で再学習させることにより、精度の高さと学習時間の少なさを両立させているところに特徴がある。

商品の購買履歴や、Web サイトの閲覧履歴など、時系列性をもった大規模データは多い。たとえば、EC サイトについて、ユーザーの閲覧データをリアルタイム、購買データはタイムシフトとみなすことで、提案手法が適用可能である。このようなデータにも適用し、その性能を調査することは今後の課題としたい。

参考文献

[1] T. Mikolov, K. Chen, G. Corrado and J. Dean, “Efficient estimation of word representations in vector space,” arXiv reprint, arXiv:1301.3781, 2013.
 [2] Q. Le and T. Mikolov, “Distributed representations of sentences and documents,” *International Conference on Machine Learning*, **32**, pp. II-1188-II-1196, 2014.
 [3] M.T. Luong, H. Pham and C.D. Manning, “Effective

approaches to attention-based neural machine translation,” In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1412-1421, 2015.

[4] T. Hofmann, “Probabilistic latent semantic analysis,” In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 289-296, Morgan Kaufmann Publishers Inc., 1999.
 [5] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” In *Advances in Neural Information Processing Systems*, pp. 535-541, 2001.
 [6] 木村義子, “テレビ視聴時間の規定要因を探る—「日本人とテレビ・2015」調査から—,” *放送研究と調査*, **66**, pp. 38-52, 2016.
 [7] 岡崎孝太郎, 井上克巳, “多デバイス接触履歴からの視聴行動モデル化と知識更新多デバイス接触履歴からの視聴行動モデル化と知識更新,” 2017 年度人工知能学会全国大会 (第 31 回), 2017.
 [8] O. Barkan and K. Noam, “Item2vec: neural item embedding for collaborative filtering,” In *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, 16, IEEE, 2016.
 [9] M. Grbovic and C. Haibin, “Real-time personalization using embeddings for search ranking at airbnb,” In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 311-320, ACM, 2018.
 [10] 堅山耀太郎, “Word Embedding モデル再訪 (特集 自然言語処理と数理モデル),” *オペレーションズ・リサーチ: 経営の科学*, **62**(11), pp. 717-724, 2017.
 [11] Y. Ni, D. Ou, S. Liu, X. Li, W. Ou, A. Zeng and L. Si, “Perceive your users in depth: Learning universal user representations from multiple e-commerce tasks,” In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 596-605, 2018.
 [12] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami and C. Dyer, “Neural architectures for named entity recognition,” In *Proceedings of NAACL-HLT*, pp. 260-270, 2016.
 [13] J. Devlin, M.W. Chang, K. Lee and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” In *Proceedings of NAACL-HLT*, pp. 4171-4186, 2019.
 [14] I. Sutskever, O. Vinyals and Q. V. Le, “Sequence to sequence learning with neural networks,” In *Advances in Neural Information Processing Systems*, **2**, pp. 3104-3112, 2014.
 [15] T. Hofmann, “Probabilistic latent semantic indexing,” In *Proceeding SIGIR'99 Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 50-57, 1999.