修 士 学 位 論 文

オンライン動画に対する印象と視聴行動の解析

平成 28 年度

東京大学大学院　情報理工学系研究科
電子情報学専攻
48-156432　福島 悠介

指導教員　山崎 俊彦 准教授

内容梗概

近年，YouTube をはじめとする動画投稿サイトや，TED のような多種多様なプレゼンが掲載されるウェブサービス，Udacity や Coursera といったインターネット上で誰もが講義を受講できるサービス（通称 MOOC）の普及により，人間が視聴者に対して話しかけたり訴えかける動画が電子的なデータとして大量にウェブ上に存在する．本稿ではこれらのプレゼンテーション動画について，これに対して視聴者が抱く印象の解析や，放送前に視聴する人数の予測を行う手法を提案した．また，類似する手法がオンラインコンテンツだけでなく，テレビ番組に対しても同様に視聴する人数，すなわち視聴率の予測を行う手法を提案し，高い精度で予測が行えることを示した．

本稿で扱ったデータと予測の対象を以下に記す．

- 大規模な講演会である TED の 1,646 本のプレゼンテーションに対し，オンラインで視聴した視聴者の抱く印象の平均精度 91.1% での予測
- 国内最大級のオンライン学習サイトである Schoo の 2,327 本の授業に対し，相関係数 $r = 0.73$ での開講前の視聴者数と $r = 0.54$ での途中退席率の予測
- 数あるテレビ番組の中でも高視聴率が見込まれるコンテンツである 678 本のテレビドラマに対する放送前の視聴率の相関係数 $r = 0.83$ での予測

これらの実験においては，各コンテンツが持つ数多の要素のうち，視聴者の抱く印象や視聴数に大きな影響を与える要因の解析も行った．

目次

表目次

図目次

第 1 章

序論

1.1　背景

ソーシャルメディアの普及により，今や誰もが写真や動画，文書をはじめとするあらゆるコンテンツの発信者となることが可能である．これらのサービスを用いることで容易に多数に向けて発信が可能となった一方で，ウェブを介して発信したコンテンツは，不特定対数の閲覧者がどのような反応を示したか，そしてどれだけの範囲に行き渡ったのかを作り手が制御することは困難である．ましてやその予測となれば，コンテンツを受け取る相手の素性は不明のため，なおさらである．

前述の閲覧者の反応や人数の指標として，ソーシャルメディアの多くは各コンテンツに対して簡潔な，例えば Like か Dislike かといった評価をユーザーが投稿できるようなシステムや，閲覧数をカウントするシステムを持っている．ソーシャルメディアの匿名性を考慮すると，これらの数値は当てにならないと一笑に付すこともできるが，その反応は作り手や話し手が目の前にいる状態での反応よりも無遠慮な，すなわち本心からの反応であるとも考えられる．我々の研究はコンテンツの中身，そしてそれに付されたメタデータを用いて，これらのソーシャルメディア特有のユーザーの反応を予測するものである．

1.2　目的

本研究の目的は，ウェブ上に存在するコンテンツのうち，特に視聴者に対して話を行う類の映像コンテンツ，すなわちプレゼンテーション動画について，視聴者の抱いた印象の予測と，その動画を視聴したり途中で離脱したりする人数を予測することである．前者については大規模な講演会である TED のプレゼンテーション群，後者については国内最大規模の MOOC である Schoo の授業群を用いて行った．また，そのコンテンツを構成する

要素のうちどれが印象や閲覧者数にどの程度影響を与えるかの解析を行うことで，プレゼンテーションの準備を行う段階，あるいは授業の告知を行う段階でより良いコンテンツを作成する支援を行える可能性を示した．加えてオンライン動画以外に対しても類似する手法が有用であるかを確認するため，動画の閲覧者数をテレビ番組における視聴率に対応させ，テレビドラマに対する視聴率の予測を行った．

1.3　構成

本稿の構成を以下に示す．第 1 章では，本研究の背景や目的について述べた．第 2 章では，TED のプレゼンテーション動画に対し，視聴者の抱く印象の予測を行う手法と実験について述べる．第 3 章では，MOOC の閲覧者数と授業からの離脱率の予測を行う手法の提案と，Schoo の授業群を用いた手法の評価を行う．第 4 章では，第 3 章で用いた手法と類似する手法をテレビドラマの視聴率予測に対して適用した結果について述べる．第 5 章では，本稿のまとめと今後の展望について述べる．

第 2 章

TED のプレゼンテーション動画に対する印象推定

2.1 はじめに

本章では大規模な講演会である TED[*1]のウェブサイトで公開されている大量のプレゼンテーション動画を解析し，自動でプレゼンテーション動画の評価を行うとともに，プレゼンテーション動画のどの要素が聴衆にどのような影響をもたらすのかを明らかにすることを目的とする．プレゼンテーションとそれに対する聴衆の印象との関係を対象とした研究 [10, 21] は近年その数を増しているが，多くは Kinect や Google Glass の普及により人の動きを容易に取得できる環境が整ったことを背景として，話者のジェスチャーを含めた特性や，聴衆のプレゼン中の所作に焦点を当てたものである．プレゼンテーションを構成する要素としてはジェスチャー以外にもその内容や話し方，スライドなどが挙げられるが，本稿で特に注目するのは話す内容と音声である．プレゼンテーションの文書と音声から抽出される特徴を用いて，各プレゼンテーション動画に視聴者がどのような印象を抱くかの推定を行い，その精度を求める．ウェブ上に存在する動画の大半にはデプスカメラなどの特殊なセンサーの情報は付与されておらず，従来のプレゼンテーションを対象とした研究では各々が用意した小規模なデータセットが用いられてきた．一方で，話し言葉を文字列へと変換する音声認識は既に実用化の段階にあり，近年は更なる発展を遂げている [27]．本章では人手で付与された字幕をプレゼンテーション動画の内容として用いるが，話す内容と音声のみから印象を高い精度で推定可能であるならば，音声認識技術と組み合わせることでプレゼンテーション動画の音声のみから印象の推定を行うことが可能となる．ユー

[*1] https://www.ted.com/

ザのプレゼンテーションを入力とする場合も，ウェブ上のプレゼンテーション動画を入力とする場合であっても，必要なデバイスの少なさは敷居の低さに繋がると考えられる．

今回データセットとして用いた TED には科学や心理学，芸術をはじめとする様々な分野のプレゼンテーション動画が存在し，2017 年現在では最大の規模を誇るデータセットといえる．印象の推定に加え，多様なプレゼンテーション動画の根底にあり，印象に影響を与える要素を明らかにすることが本稿の目的である．

2.2　関連研究

2.2.1　はじめに

印象解析や印象推定は文書 [48] や音楽 [37] など様々な対象について行われてきた．画像を対象とした研究として，Lu ら [39] は画像の美しさに注目し，画像全体の情報に画像をクロッピングした一部分のみの情報を組み合わせることで，審美性の推定精度が向上することを示した．Soleymani ら [54] は動画の視聴を対象とし，脳波や視線情報を用いて喜びや安堵といった 6 種の印象について高精度での推定を行った．ただしいずれもプレゼンテーション動画を対象としたものではない．また，既存研究の多くはコンテンツの評価に数個の指標を用いるのに対し，本章では次節で述べる 14 の印象によってプレゼンテーション動画の評価を行う．

2.2.2　文書に対する印象推定

文書と感情との結びつきを対象としたものとしては，Pang ら [49] の研究を皮切りに，そのポジティブさとネガティブさを測る，あるいはこの 2 クラスに文書を分類する研究が行われてきた．本稿の手法と近いものとして，Zhang ら [64] は Amazon[*1]における衣服に対する 1 万件のレビューを用いて，90% の精度でそれが好意的なものか，批判的なものかの分類を行った．同様の 2 クラス分類を深層学習により行ったものも近年多く見られる [9, 13, 29] が，いずれも一文が数百単語からなる比較的短い文書群の識別を数万のトレーニングデータを用いて行っている．これに対し，TED のプレゼンテーションは 2,000 個程度であり，文字起こしを行うとそれぞれは数千単語にもなる長文であるため，データの規模は小さく，各データはより複雑である．深層学習は印象に影響を与える要因の解析に適さないという問題もある．また，既存研究の多くは書き手側の印象，すなわちユーザーレビューに対してレビュワーが好意的か否かのような印象を予測の対象とするのに対し，

[*1] http://www.amazon.cn/

本研究ではプレゼンテーションを視聴する側の抱く印象を予測の対象とする．

2.2.3　プレゼンテーションを対象とする研究

プレゼンテーションの評価を行う研究として，Luzardo ら [40] は大学生のプレゼンテーションのスライドと音声を用いて，プレゼンテーションの質が高いか低いかの 2 クラス間の識別を行った．スライドの特徴量としては図の数やフォントサイズが用いられ，音声の特徴量としては音の高低や発話時間などが用いられた．その結果，スライドを用いた場合は 65%，音声を用いた場合は 69% の精度でプレゼンテーションの良し悪しが判定可能であることを示した．本研究と同様に TED の動画群を用いた研究として，Weninger ら [59] は次節で述べる BoW モデルのみを用いて，視聴者により投稿された印象の推定を行った．

実際にユーザのプレゼンテーションを対象とするシステムには [34] などがある．これはマイクとカメラから得られた話者の音声と振る舞いを分析し，話速度，アイコンタクトの度合いなどをリアルタイムに話し手にフィードバックするシステムである．本研究とは音声を用いる点で共通しているが，[34] では話速度や発話時間，声の高さなど各要素に基準値を設けるのに対し，本研究では音声全体を一つの要素として聴衆に与える影響の解析を行う．他にユーザへのフィードバックまでを目的とした研究として，Google Glass や Kinect の出現により，これらのデバイスを活用して話し手，時には聴衆のあらゆる情報を用いて解析を行う研究 [19, 22, 44] も近年見られる．例えば [22] では Kinect1 台，マイクと RGB カメラ 2 台，Google Glass3 台のデバイスを用いている．多くの情報を得ることで解析の精度や可能なフィードバックの種類は増すものの，実世界に手法を導入する敷居は向上する．

2.3　プレゼンテーションの特徴ベクトルの作成

本節ではプレゼンテーション動画を構成する多数の要素のうち，話す内容と音声に注目し，特徴量を作成する手法を示す．

2.3.1　内容特徴量

プレゼンテーションの内容を表現する特徴量を作成するため，文書分類の手法としてこれまで広く用いられてきた Bag-of-Words (BoW) [24] , Latent Semantic Indexing (LSI) [16] , Latent Dirichlet Allocation (LDA) [3] の 3 つに，Mikolov らの手法 [43] が実装された word2vec，Bojanowski らの手法 [5] が実装された fastText による特徴量を加

えた 5 つを用いる．また，単体で最も精度の良い word2vec に Spherical K-means (spk-means) [18] を適用した手法，そこにさらに Approximated locality-constrained linear coding (LLC) [57] を組み合わせた手法を提案し，計 7 つの特徴量の比較を行う．

Bag-of-Words

BoW は文書の内容を特徴ベクトル化する手法として最も簡単なもので，各文書に各単語が何回出現するかを数えて特徴量とする手法である．各文書の特徴ベクトルの次元数はデータセット内に出現する単語の種類数となる．文書の情報を欠くことはないが，次元数が膨大であること，単語間の関係が考慮されないなどの欠点がある．

データセット内に存在するユニークな単語数を V とすると，それらを辞書順に並べた際の v 番目の単語 w_v のベクトル $\boldsymbol{v}_{w_v}$ は，次元数が V で v 番目の要素のみが 1 であるベクトルで表される．データセット内の文書 d の特徴ベクトル $\boldsymbol{v}_d$ はそこに含まれる全単語のベクトルの総和を正規化したもの，すなわち以下の式で表される．

$$\boldsymbol{v}_d = \frac{\sum_{w \subset d} \boldsymbol{v}_w}{\| \sum_{w \subset d} \boldsymbol{v}_w \|} \tag{2.1}$$

BoW モデルによる各文書のベクトルは外部情報を含まず，正規化により失われる総単語数を除き，文書の情報は失われない．

LSI, LDA

LSI や LDA は大量の文書を用いた事前学習を行い，関連のある単語を 1 つの次元にまとめることで文書を低次元のベクトルで表現することを可能とする手法である．同一文書中に共起して出現しやすい単語が関係のある単語とみなされ，各単語にはその次元への寄与の大きさに応じた重みがつけられる．例えば soccer と goal は同じ次元にまとめられ，結果として soccer の頻出する文書と goal の頻出する文書では特徴ベクトルが似ることとなる．

今回は学習用の文書として Wikipedia の記事群を用いた．Wikipedia の文書数 N, 出現するユニークな単語数 V を用いて，tf-idf により重み付けされた単語文書行列を $A_{V \times N}$ と表す．これを特異値分解により 3 つの行列に分解し，$T_{V \times K}$ の k 列目までを取り出すことで $T_{V \times k}$ としたものの v 行目が，v 番目の単語 w_v の LSI による k 次元の単語ベクトルを表す．ただし $K = min(V, N)$ である．

$$A_{V \times N} = T_{V \times K} S_{K \times K} \left(D_{N \times K}\right)^T \tag{2.2}$$

LDA は LSI を拡張したモデルであり，文書が複数のトピックから構成されると考え，その構成比を離散分布により表現する．はじめにトピック数 K を与え，Wikipedia の記

事群で学習を行うとき，文書 d を構成するトピックの分布 $\boldsymbol{\theta}_d = (\theta_{d,1}, \cdots, \theta_{d,K})$ と，各トピック k における各単語の出現確率 $\boldsymbol{\phi}_k = (\phi_{k,1}, \cdots, \phi_{k,V})$ が Dirichlet 分布に基づき生成される．本稿ではこの $\boldsymbol{\phi}$ を利用し，ある v 番目の単語 w_v について，以下のように各トピック内での w_v の出現確率を用いて単語のベクトル $\boldsymbol{v}_{w_v}$ を定める．

$$\boldsymbol{v}_{w_v} = (\phi_{1,w_v}, \cdots, \phi_{K,w_v}) \tag{2.3}$$

LSI, LDA いずれの手法においても，BoW と同様に，ある文書に含まれる全単語のベクトルの総和を長さ 1 に正規化したものを文書ベクトルとして用いる．各単語のベクトルはそれが含まれる文書全体から学習されるため，文書全体の話題をうまく分類するようなベクトルが得られると考えられる．プレゼンテーションの内容の分類と，聴衆の受ける印象の分類とが等質の問題かを明らかにすることがこの特徴量の役割となる．

Word2vec，fastText によるモデル

Word2vec により実装された Mikolov らのモデルは，ニューラルネットワークを用いて単語ベクトルを得る手法である．LSI や LDA と同様に大量の文書による事前学習が必要だが，学習時に文書内での単語の共起ではなく前後の数個の単語との関係を用いる．前後の単語から中央の単語を予測するモデル（CBOW）と一つの単語から周囲の単語を予測するモデル（Skip-gram）が提案されているが，本稿では単語の意味関係を表すのに優れた Skip-gram モデルを用いる．LSI や LDA では一つの文書をあらかじめ BoW モデルに落とし込んだ後に学習を行うため，文書内での語順は結果に影響しない．一方で Mikolov らのモデルではある単語の学習の際，その周辺の単語のみを用いるため，文法的な情報が保持される点，そして文書全体の話題の情報が入りにくい点が差異となる．この手法の特徴は単語ベクトルの足し引きが意味を持つ点であり，プレゼンテーション中に出現する全単語のベクトルの和をとり長さを 1 に正規化することで文書の特徴量とする．

fastText により実装されたモデルはこれをベースとした手法であり，各単語に一つのベクトルを割り当てるのではなく，各単語を文字レベルで n-grams（[5] では n は 3 以上 6 以下）に分割し，各 n-gram について前述のモデルを用いてベクトルを割り当てるモデルである．その後，再び単語を構成する n-grams のベクトルを足し合わせたものをその単語のベクトルとする．単語の一部が意味を持つ（英語であれば接頭辞の pre-が “前” を表すなど）言語においてはこの手法は有効であると考えられ，論文中で行われている複数の実験において，ドイツ語やチェコ語については word2vec のモデルに比べて大幅に精度が向上するものの，本章で扱う英語においては精度が向上しない実験もある．[5] では [43] で行われた 2 種類（意味的，言語的）の実験について比較を行っており，州と都市（Chicago-Illinois）や性別（brother-sister）といった意味的な関係については word2vec で実装されたモデル

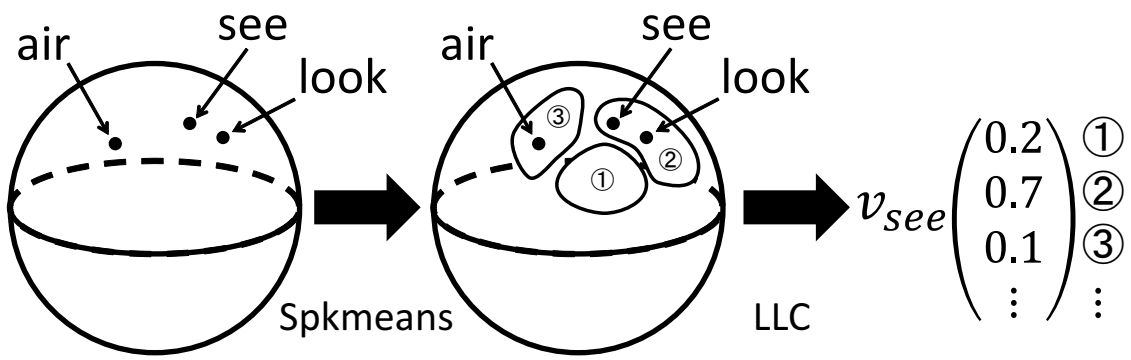


図 2.1: Spkmeans と LLC による単語ベクトルの作成

が，単語の変化（work-works, easy-easiest）のような言語的な関係については fastText で実装されたモデルが英語については良い精度を示したと報告している．

2.3.2　Word2vec のソフトクラスタリング

本手法により文書から特徴量を抽出する流れを図 2.1 に示す．Word2vec により各単語がベクトル化された空間に対し，距離計算にコサイン距離を用いる K-means である Spherical K-means を適用することで単語群を K 個のクラスタに分類することができる．Word2vec では似た意味の単語が近いベクトルを持つため，see と look のように意味の近い単語が同じクラスタへ属することとなる．この段階では各単語は，次元数が K で自分が属するクラスタの要素のみが 1 の 1-of-K ベクトルとみなすことができ，近い意味の単語群は完全に同じベクトルとなっている一方で，元の word2vec のベクトルに比べると多くの情報が失われている．そこで Spherical K-means により分割された各クラスタを用いてソフトクラスタリングを行い，各単語の各クラスタへの帰属度を求め，これを各単語の特徴ベクトルとする．今回ソフトクラスタリングの手法として用いた LLC は，与えられたベクトルを近傍の n 個の基底ベクトルの線形結合で表現する手法である．d 次元の N 個のデータを $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}] \in \mathbf{R}^{d \times N}$，$K$ 個の基底ベクトルを $\boldsymbol{B} = [\boldsymbol{b_1}, \boldsymbol{b_2}, ..., \boldsymbol{b_K}] \in \mathbf{R}^{d \times K}$ とすると，各データを基底ベクトルで表す際の係数行列 $\boldsymbol{C} = [\boldsymbol{c_1}, \boldsymbol{c_2}, ..., \boldsymbol{c_N}]$ は次の関数を最小化することで求められる．本章では各単語の word2vec によるベクトルが $\boldsymbol{X}$，Spherical K-means 後の各クラスタの中心ベクトルが $\boldsymbol{B}$，ソフトクラスタリング後の各単語のベクトルが $\boldsymbol{C}$ に対応する．

$$\min \sum_{i=1}^{N} ||\boldsymbol{x_i} - \boldsymbol{B}\boldsymbol{c_i}||^2 + \lambda ||\boldsymbol{d_i} \odot \boldsymbol{c_i}||^2 \tag{2.4}$$

$$\text{s.t. } 1^T \boldsymbol{c_i} = 1, \forall i$$

$$\boldsymbol{d_i} = \exp\left(\frac{dist\left(\boldsymbol{x_i}, \boldsymbol{B}\right)}{\sigma}\right) \tag{2.5}$$

ただし $dict\left(\boldsymbol{x_i}, \boldsymbol{B}\right)$ は $\boldsymbol{x_i}$ と $\boldsymbol{B}$ の各要素のユークリッド距離を並べたベクトルを表す．変化させるべきパラメータは $\boldsymbol{c_i}$ の要素数 n である．これは最大でいくつの基底を用いて一つのベクトルを表現するかを表し，小さいほどよりハードなクラスタリングとなる．$n = 1$ のときは各単語のベクトルが，自分の属するクラスタのみが 1 である単なる 1-of-K ベクトルとなることからわかるように，本手法においてハードなクラスタリングとは，元の word2vec による単語ベクトルの情報を失う一方で似た意味の単語がより近いベクトルとなることを意味する．この n の値により，類似する単語をどれだけ近いベクトルで表現するかを変更することが可能である．ある文書の特徴ベクトルは，文書内に出現する全単語のベクトルの和をとり長さを 1 に正規化したものとする．

2.3.3　言語的特徴量

プレゼンテーションの文法的な複雑さも印象に影響を与えると考え，これを数値化したものが言語的特徴量である．文書の読みやすさを数値化しようとする研究 [32, 41, 53] は古くから行われており，1 文あたりの単語数，1 単語あたりの文字数，1 単語あたりの音節数，難単語の割合の 4 つが頻繁に用いられる．単語の難しさの数値化については，3 音節以上の単語の割合を用いるものや，日常的な単語と非日常的な単語を定義し非日常的な単語の出現回数を数えるものなどがある．本稿ではオンライン学習サイトである BigIQkids[*1]の学年別英単語リストを用い，プレゼンテーションに各学年で習う単語が何度出現するかのヒストグラムを特徴量とした．これはアメリカの初等教育にあたる 1 年生から 8 年生までの各学年，そして大学進学適性試験 (SAT) レベルの単語の出現回数の 9 次元からなる．これに単語の総数などを加え，表 2.1 に示す 44 次元の特徴量を作成する．

2.3.4　音声特徴量

たとえ原稿が同じでもプレゼンテーションの出来は話し手の技量に大きく左右される．話の中身とは無関係に音声のみから抽出されるものが音声特徴量である．我々は

[*1] http://www.bigiqkids.com/

表 2.1: 言語的特徴量

特徴量	次元	詳細
1 文あたりの単語数の平均	1	
1 単語あたりの文字数の平均	1	
1 単語あたりの音節数の平均	1	
文の総数	1	
単語の総数	1	
文字の総数	1	
音節の総数	1	
文の単語数のヒストグラム	12	1-3，4-5，6-7，8-9，10-11，12-13，14-15, 16-18，19-21，23-28，29-40，41-
単語の文字数のヒストグラム	11	1，2，..，10，11-
単語の音節数のヒストグラム	5	1，2，3，4，5-
各学年の単語	9	1 年生, 2 年生, .., 8 年生, SAT
合計	44	

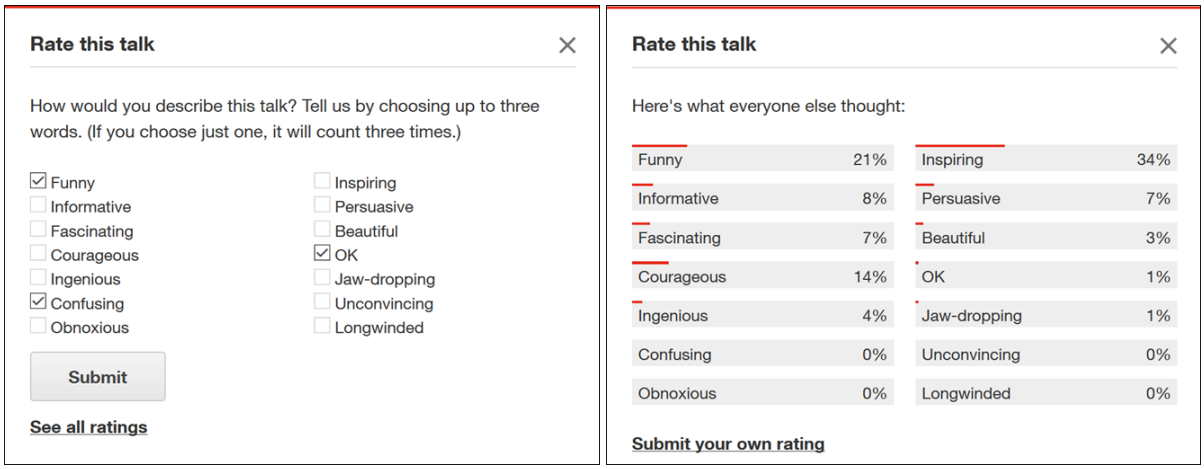
INTERSPEECH 2013 Computational Paralinguistics Challenge[*1]で使用されたものと同一のものを音声特徴量として用いた．これは 6,373 次元の特徴量であり，各微小区間におけるエネルギーの平均や標準偏差，MFCC などからなる．

2.4　データセット

2014 年 7 月時点で TED のウェブサイトで公開されていた 1,900 本以上のプレゼンテーション動画のうち，歌唱やダンスが中心のものを除いた 1,646 本をデータセットとして用いた．TED では Bill Clinton のような世界的な有名人やプロのアスリートをはじめとし，各分野で活躍する人々が講演者を務める．今回用いた動画の中で最も旧いものは 1984 年に行われたプレゼンテーションであるが，ごく一部を除き，大半は 2001 年以降に行われたプレゼンテーションである．

これらの各動画には人手により字幕が付けられており，英語だけでなく日本語，韓国

[*1] http://emotion-research.net/sigs/speech-sig/is13-compare/



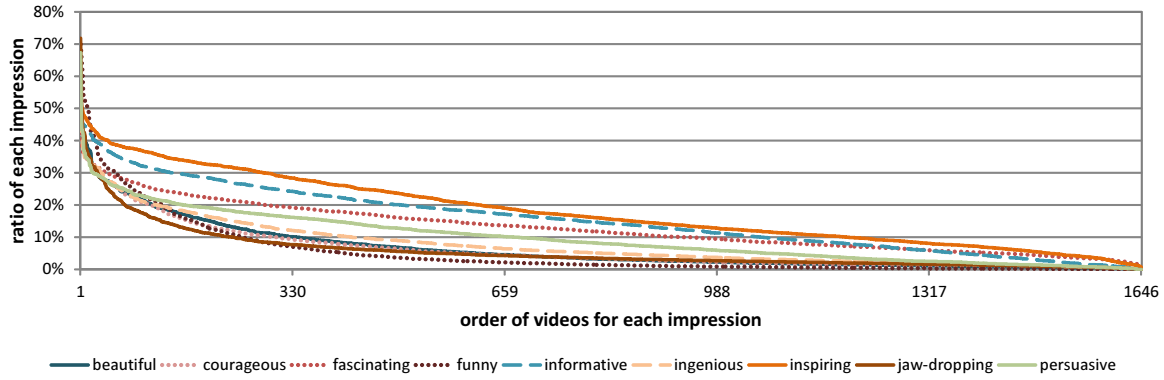
(a) 投稿フォーム　　(b) 他のユーザーの投稿の様子の表示

図 2.2: TED の各動画への印象の投票画面

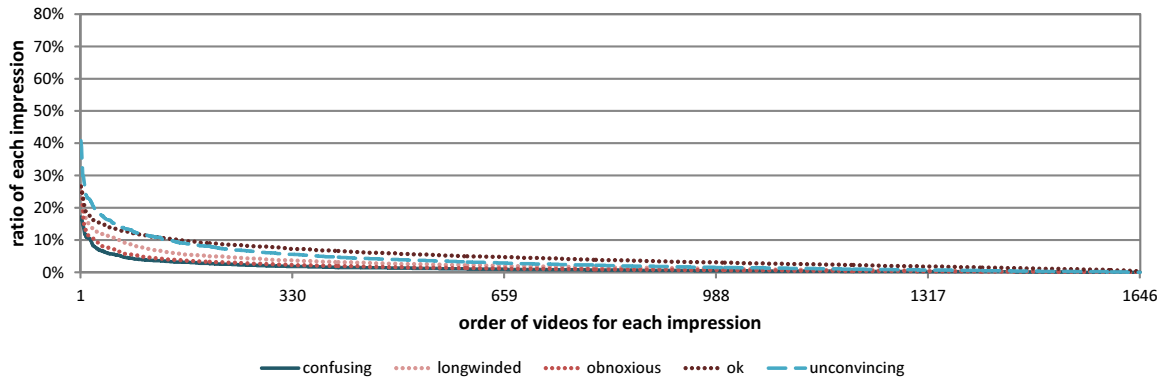
語，ドイツ語，スペイン語など様々な言語の字幕を付けての視聴が可能である．また，動画を視聴した各ユーザーは図 2.2 に示す画面からその動画を見て抱いた印象を 14 個（beautiful, confusing, courageous, fascinating, funny, informative, ingenious, inspiring, jaw-dropping, longwinded, obnoxious, ok, persuasive, unconvincing）の中から最大 3 つまで投票できる．なお，1 つの印象のみにチェックを付けて投稿した場合のみ，それは 3 票として集計される．我々は TED の公式 API[*1] を用いてこの印象の投票数を取得し，視聴者が各プレゼンテーションに抱く印象の正解データとして用いた．なお，この API は 2016 年 7 月をもって一般向けのサービスは停止されており，現在は TED にとって有益な Web サービスの製作者など，TED から認証を受けた者のみ利用が可能である．我々は本研究の継続のため，API のパートナーキーを発行して頂いた．

データセット内の 1646 本の動画について，各印象の投稿の傾向を図 2.3 に示す．横軸は各印象について，後述の図 2.2 の手法で投稿割合が高い順に動画を並べた時の順位，縦軸は各順位において，全印象に対して特定の印象が投稿された比率を示す．例えば，1646 本中 330 番目に inspiring の投稿比率の大きな動画では，視聴者による全投稿のうち 30% 弱が inspiring であることがわかる．

[*1] http://developer.ted.com/



(a) 良い印象群



(b) 悪い印象群

図 2.3: 印象別の投稿比率の順位と投稿比率

2.5　実験

2.5.1　概要

前項で述べた 1,646 本のプレゼンテーション動画について，視聴者がその動画について抱く印象を予測することが本研究の目的である．これを模した実験として，14 の印象に対して，特にその印象が抱かれるプレゼンテーション群と抱かれないプレゼンテーション群の 2 クラスの動画群について，これを Radial basis function (RBF) カーネルを用いた Support Vector Machine (SVM) [12] により識別する実験を行った．これが高い精度で行

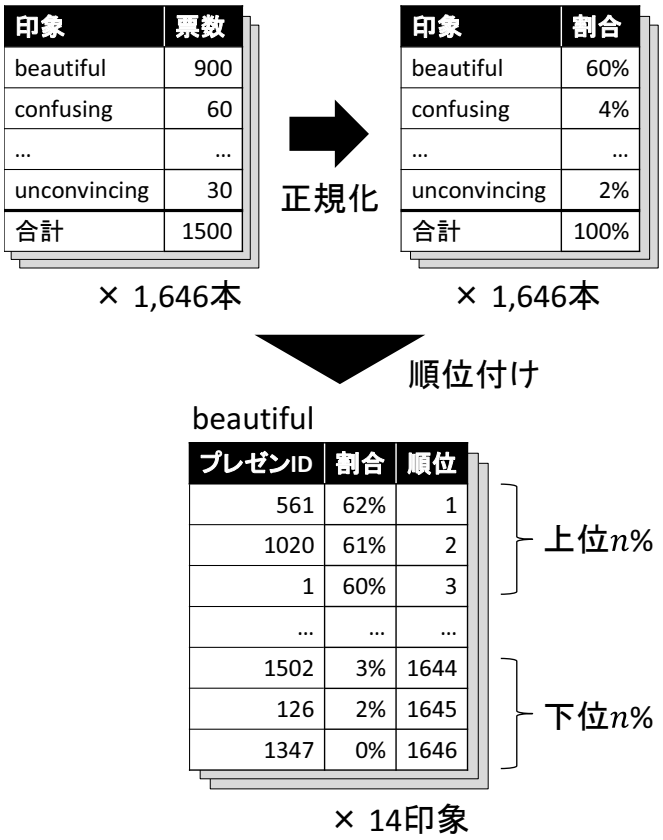
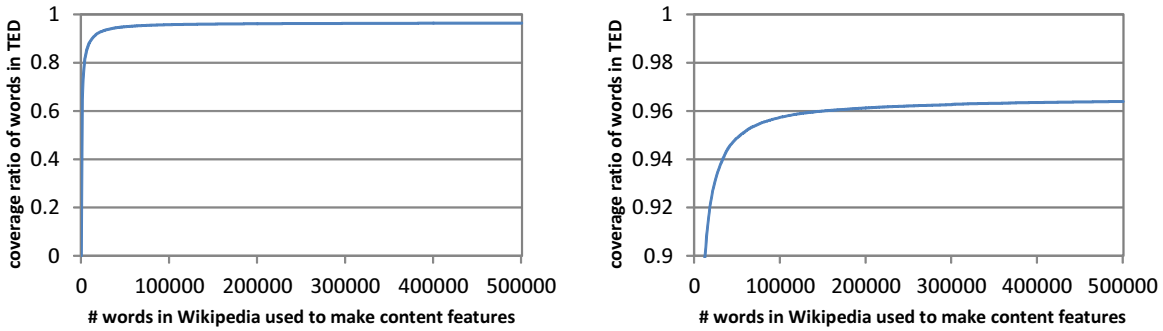


図 2.4: 各印象の上位・下位 $n\%$ のプレゼンテーション動画群の抽出課程

えるならば，ユーザーのプレゼンテーションに対して各印象を聴衆が抱くか抱かないかの識別も行えると考えられる．各印象について，2 クラスの動画群を抽出する流れを図 2.4 に示す．各プレゼンテーション動画について，各印象の投稿数を 14 印象の投稿数の和で割り，比率に直す．人気のあるプレゼンテーション動画は視聴者が多く，すべての印象の投稿数が多いため，この影響を取り除くことが目的である．データセット内の最も得票数の多い動画では計 66,636 票，少ない動画では計 154 票の投票があった．次に特定の印象のみに着目し，その印象の投稿割合をもとに全プレゼンテーション動画を降順に並べ，上位 $n\%$ をその印象が抱かれるプレゼンテーション群，下位 $n\%$ をその印象が抱かれないプレゼンテーション群として抽出する．$n = 10, 30, 50$ の 3 通りについて実験を行った．実験には Libsvm[8] を用い，評価は 10 分割交差検定により行った．上位・下位 $r\%$ は図 2.3 の左右端 $r\%$ に相当する．

特徴量の作成に関して，今回用いた 1,646 本のプレゼンテーション動画中には，一度しか出現しない単語を除いて 43,808 種類の単語が出現したため，これが BoW 特徴量の次



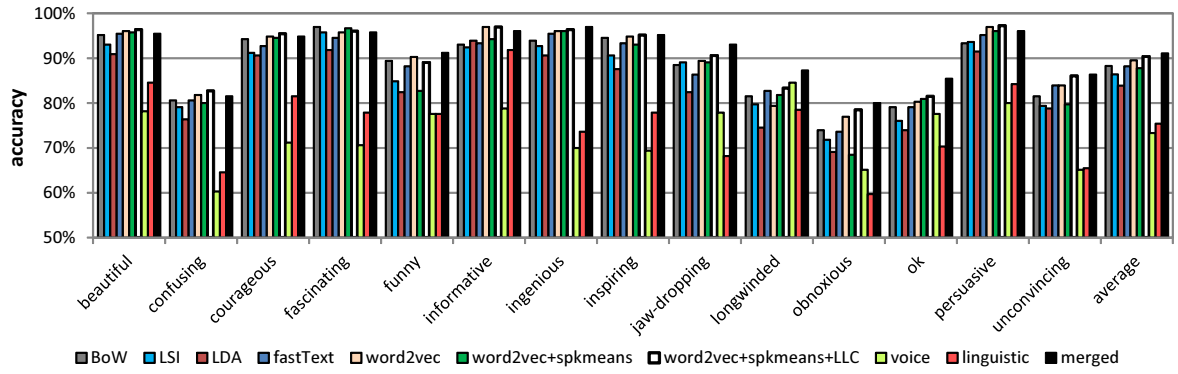
(a) 全範囲を描画したもの　　(b) $y > 0.9$ を拡大したもの

図 2.5: Wikipedia 中の出現頻度の高い n 単語（横軸）と TED 中の総単語の充足率（横軸）

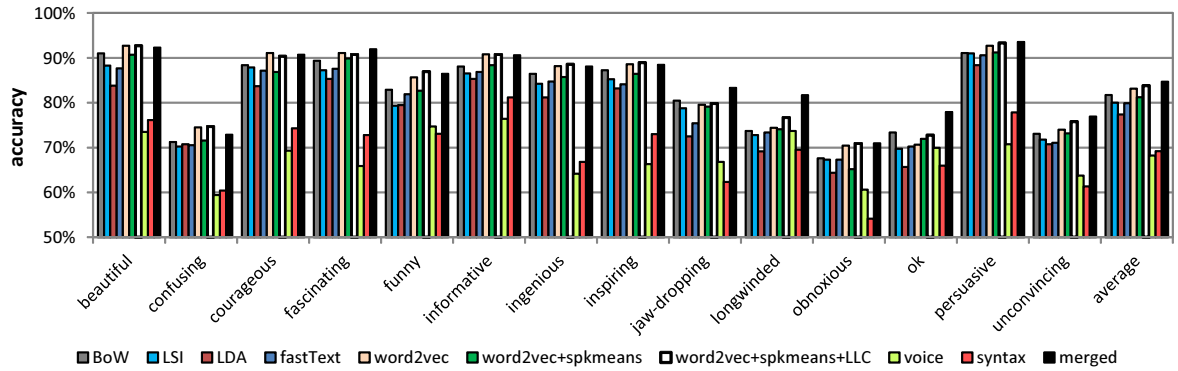
元となる．LSI と LDA では学習に 2014 年 11 月 8 日に取得した Wikipedia 英語版の全記事群を用いた．次元数については 100 から 3,000 次元までの 6 通りで実験を行い，LSI, LDA 共に最も精度の良かった 3,000 次元を用いた．Word2vec と fastText の学習には同様に Wikipedia の全記事群を用い，単語ベクトルの次元数は 500 とした．学習後，出現頻度の高い上位 100,000 単語のみを実験に用いた．図 2.5 は学習モデルから用いる単語数と，そこに TED のプレゼンテーション中の総単語数のうちどの程度が含まれているかを表すグラフであり，上位 100,000 単語には 95.7% が含まれている．Spherical K-means の k は 100 から 5,000 までの 6 通り，LLC の n は 1 から k までの最大 12 通りで実験を行い，最も精度の良かった $k = 5000, n = 2000$ のものを用いた．

2.5.2　実験結果

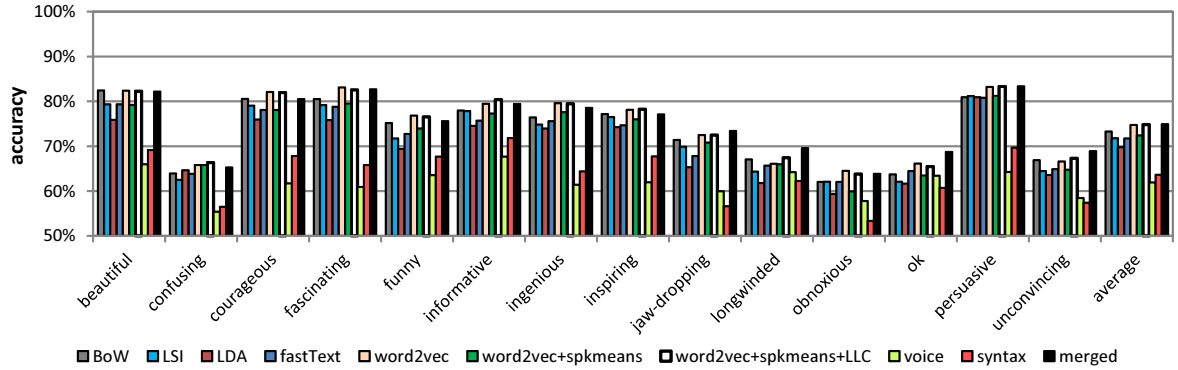
前節で述べた特徴量を用いて，各印象について 2 クラス間の識別を行った結果を図 2.6 に示す．各印象について，(a) 上位・下位 10% (計 330 本) のみを使用，(b) 上位・下位 30% (計 988 本) のみを使用，(c) 上位・下位 50% (1,646 本すべて) を使用したときの識別結果である．BoW, LSI, LDA, fastText, word2vec, word2vec+spkmeans+LLC はそれぞれ第 2.3.1 項で述べた内容特徴量であり，word2vec+spkmeans は LLC を適用する前の 1-of-K ベクトルで表現された単語ベクトルを足し合わせて正規化したものである．linguistic は第 2.3.3 項で述べた言語的特徴量，voice は第 3.4.8 項で述べた音声特徴量を表し，merged は内容特徴量の中で最も平均精度の良い word2vec+spkmeans+LLC と linguistic，voice の 3 つの特徴量を結合したものである．



(a) 上位・下位 10% の識別



(b) 上位・下位 30% の識別



(c) 上位・下位 50% の識別

図 2.6: 各特徴量での各印象の識別精度

単体の内容特徴量である 5 つを比べてみると，word2vec が平均で最も精度が良かった．fastText は word2vec の開発者らによる後継の特徴量であるが，前述したように [5] では英語のデータを対象とした実験において，言語的特性を問うような実験では word2vec に比べて fastText が，意味的特性を問うような実験では fastText に比べて word2vec が良い精度を達成している．本章で扱った，プレゼンテーション動画に対して視聴者が抱く印象の予測は，意味的な特性が重視されると考えられるため，word2vec がより適していたと考えられる．7 つの内容特徴量の精度を比べてみると，平均では提案手法である word2vec+spkmeans+LLC の精度が最も良く，$r = 10$ のとき 14 印象の平均で 90.4% の精度が出ている．

印象間での精度を比べてみると，confusing や obnoxious といった悪印象の識別精度が低いことがわかる．図 2.3 に示したように，悪い印象は良い印象に比べ，投稿数が大幅に少ない．例えば，$r = 10$ (上位，下位それぞれ 165 本) のときの unconvincing か否かの識別を考えてみると，これは unconvincing が 8.9% 以上投稿された動画と 0.4% 以下投稿された動画との 2 クラス分類となる．一方で良い印象である inspiring に関しては inspiring が 33.8% 以上投稿された動画と 5.5% 以下投稿された動画との分類であり，問題の難しさが異なることが理由として挙げられる．文書分類においては高い精度を誇る LSI や LDA は BoW にさえ劣る結果となった．LSI では一般に数百次元のものが用いられる [6] のに対し，我々の実験においては次元数が 3,000 と大きい場合に最もよい精度が出ており，この手法がうまく働いていないことが見受けられる．印象推定においては単語をトピックという括りでまとめることは不必要，すなわち話題での分類と視聴者の抱く印象での分類は異なる問題であると考えられる．Word2vec+spkmeans に関しても，LLC を適用する前は類似する単語が同一次元にまとめられただけの状態なため，精度は単なる word2vec より低いと考えられる．全体では merged が最も精度が良く，$r = 10$ のときに最も精度の良い ingenious で 97.0%，精度の悪い obnoxious で 80.0%，平均で 91.1% の精度となった．

$r = 10$ のときに従来特徴量に比べて提案手法の精度が特に高かった印象である persuasive について，BoW, LSI, LDA ではいずれも識別に失敗した一方で，word2vec+spkmeans+LLC では識別に成功した動画が 3 本[*1]存在した．この 3 本の共通点としていずれも「persuasive である」が正解である点，そして一般的に説得力がないとみなされている題材を扱っている点が挙げられる．一本は超常現象や疑似科学を科学的な観点から説明したもの，一本は手品師や霊媒師が人を騙すときに用いる手法を解説したもの，そしてもう一本は教科書を小説のようなフィクション仕立てに改編することの是非に関するものである．Word2vec による単語ベクトルは，LSI や LDA によるものに比べ，単

[*1] http://www.ted.com/talks/{22,835,1655}

表 2.2: 線形 SVM による識別精度と識別に強く影響を与える単語群

impression	beautiful	confusing	courageous	fascinating
accuracy	94.5%	77.0%	93.3%	95.8%
cluster1	sensed	everyone	equipped	agree
	aware	anybody	fitted	accept
	realize	things	customized	admit
cluster2	increase	assume	vivid	such
	excess	regard	beautiful	certain
	improve	believe	colorful	related
cluster3	hero	helpless	started	experiment
	narration	noticing	began	experimenter
	fictionalized	dazzled	moved	experimental

impression	funny	informative	ingenious	inspiring
accuracy	80.9%	93.3%	94.8%	89.4%
cluster1	funny	mainly	project	started
	amusing	widely	design	began
	playful	often	work	moved
cluster2	clinics	roughly	develop	life
	hospital	estimates	create	sense
	surgeon	tens	build	wildness
cluster3	aren't	sit	notion	such
	hasn't	leaning	belief	certain
	couldn't	stand	idea	related

impression	jaw-dropping	longwinded	obnoxious	ok
accuracy	86.1%	72.7%	65.5%	75.2%
cluster1	here	thank	centers	matter
	earliest	gratitude	sites	cases
	immemorial	farewell	places	basis
cluster2	twisted	careful	CBS	talked
	rope	pithy	WGN	spoke
	glued	copious	WBZ	chatted
cluster3	project	hardly	useful	hardly
	design	vaguely	helpful	vaguely
	work	partly	workable	partly

impression	persuasive	unconvincing
accuracy	93.9%	77.6%
cluster1	ask urge advising	rubbing sticking pinched
cluster2	problem tasks dilemma	arrest crimes charges
cluster3	country nation GPO	CBS WGN WBZ

表 2.3: “Before Avatar... a curious boy” (Speaker: James Cameron) の予測結果

imp.	bea.	con.	cou.	fas.	fun.	inf.	ing.	inp.	jaw.	lon.	obn.	ok	per.	unc.
true	0	-1	0	1	0	0	0	1	0	1	1	1	0	0
pred.	0	1	0	1	0	0	0	1	0	1	1	1	0	0

表 2.4: “Innovating to zero!” (Speaker: Bill Gates) の予測結果

imp.	bea.	con.	cou.	fas.	fun.	inf.	ing.	inp.	jaw.	lon.	obn.	ok	per.	unc.
true	-1	1	0	0	0	1	0	0	0	0	0	0	1	1
pred.	1	1	0	0	0	1	0	0	0	0	0	0	1	1

語の意味情報だけでなく文法的な情報も含む．加えて LSI や LDA の学習では tf-idf による重み付けにより，名詞に比べて，いかなる話題にも普遍的に使われやすい副詞や動詞は重みが小さくなる．提案手法では説得力のあるプレゼンテーションの属するトピックだけではなく，説得力のあるプレゼンテーションに用いられがちな文脈や副詞，動詞などの傾向を考慮することで識別に成功したと考えられる．

データセット内のプレゼンテーションの予測例について，映画『アバター』や『タイタニック』の監督である James Cameron の TED でのプレゼンテーションの予測例を表 2.3，マイクロソフトの創業者である Bill Gates の TED でのプレゼンテーションの予測例を表 2.4 に示す．上位・下位 10% ではほとんどの印象についてラベルが 0 であること，上位・下位 50% では境目付近にある印象の予測について信憑性に欠くことから，上位・下位 30% の識別結果を示す．いずれも結果のグラフ通り，80% 強の精度で識別が行えてい

ることが見て取れる．

2.5.3　各印象に影響を与える要素

音声特徴量が内容特徴量に唯一優っている印象として，$r = 10$ のときの longwinded が挙げられる．プレゼンテーションが退屈かどうかはその内容よりも話者の話し方による影響が大きいと言える．

TED の動画には聴衆の笑い声が入り込んでいることがあり，退屈か否かの識別ではこれが影響してしまった可能性がある．笑いが起きた部分については字幕に印がついているため，これを数えることで聴衆の笑い声が入った回数を数えることが可能である．前述の $r = 10$ のときの longwinded の上位・下位 10% の動画群について，上位 10% の笑いの回数の平均は 4.34 回，下位 10% の平均は 3.68 回であった．t 検定の結果，10% の有意水準でこの 2 つの間に有意差は見られなかった．ここから，退屈な講義群と退屈でない講義群の間に笑いの回数の有意差はないと考える．一方で笑いの回数と直接印象が結びつくと考えられる funny については，音声のみでの識別精度は高くないものの，funny の上位 10% では平均 12.4 回，下位 10% では平均 0.73 回の笑いがあり，有意水準 1% で有意差が見られた．また，longwinded に関しては内容特徴量と組み合わせることで大きく精度が向上しているため，内容が退屈なプレゼンテーションと話し方が退屈なプレゼンテーションは包含関係にはなく，多くの人が退屈だと感じたプレゼンテーションの一部は内容だけが退屈なものと考えられる．

線形 SVM を用いて同様の実験を行うことで，識別精度は低下する一方で特徴ベクトルの各次元の影響の強さ，すなわち重みを抽出することが可能である．表 2.2 に特徴量として word2vec+spkmeans を用いた際の，線形 SVM による 10 分割交差検定の識別精度を示す．実験には Liblinear[20] を用いた．文書の特徴ベクトルが $\boldsymbol{x}$ で表されるとき，線形 SVM における各印象に関する識別関数は以下のように表される．

$$f(\boldsymbol{x}) = sign(\boldsymbol{w}^t \boldsymbol{x} + b) \tag{2.6}$$

ただし b はバイアス項である．$\boldsymbol{w}$ の各次元は，特徴ベクトルの対応する各次元の影響の強さを表していると言える．

今回は word2vec の学習後，出現頻度の高い 100,000 単語のみを用いており，それを spherical K-means により 5,000 のクラスタに分類したため，word2vec+spkmeans による特徴ベクトルの次元は 5,000 であり，1 次元には平均で 20 の単語が属する．これと $\boldsymbol{w}$ の各次元を対応させることで，各印象の識別において重要である単語群を特定することが可能となる．表 2.2 に，各印象の識別において重みの大きかった上位 3 クラスタについ

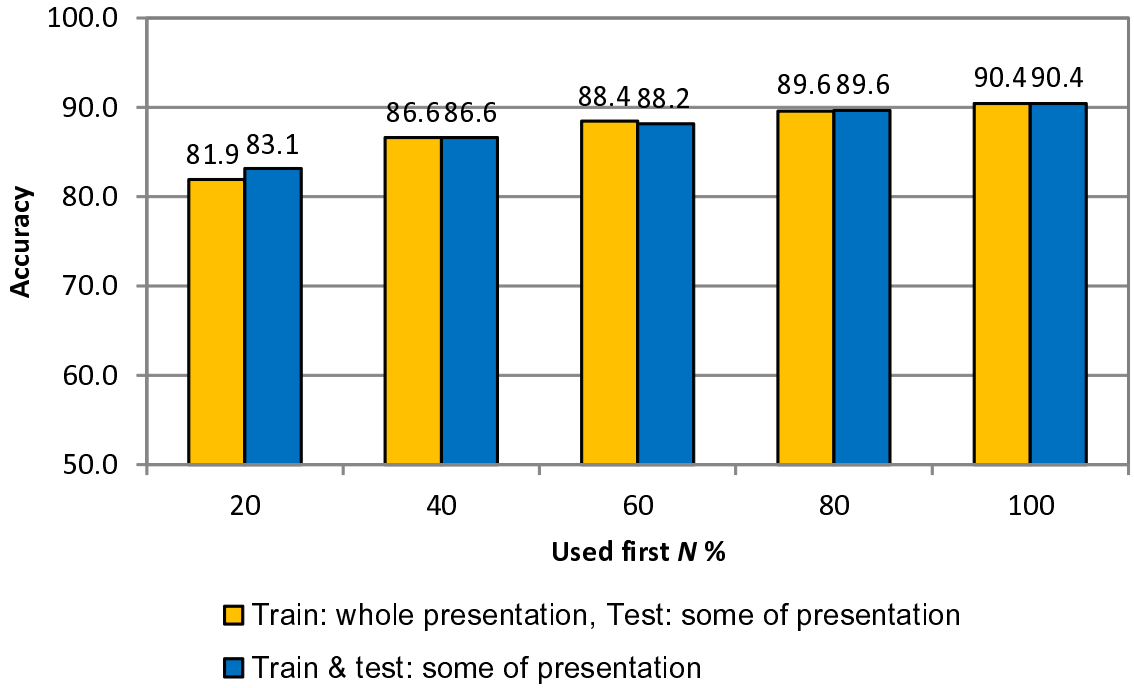


図 2.7: プレゼンテーションの冒頭の一部のみを用いた場合の識別精度

て，そのクラスタに属する単語を 3 つ例示した．courageous か否か，inspiring か否かでは start や begin といった何かを始める様子を表す動詞が上位に現れており，状況の変化と人を鼓舞するプレゼンテーション動画には結びつきがあることが見て取れる．一方で悪い印象に着目すると，例えば longwinded では thank のような謝意を表す単語群からなるクラスタが上位に存在している．これは出だしや結びに多く用いられる単語であり，それらの長さと退屈さとの間に関連があると考えられる．

2.5.4　リアルタイムなプレゼンテーション解析に向けた検討

プレゼンテーションの全体ではなく冒頭の一部分のみを用いた場合にも高い精度で印象の予測が可能であれば，ユーザーのプレゼンテーションの解析の際，入力中にリアルタイムで結果を出力することが可能である．図 2.7 に先の上位・下位 10% の識別実験を TED のプレゼンテーションの冒頭 N% のみを用いて行った場合の結果を示す．横軸は N，縦軸は 14 の印象の平均識別精度を表す．用いた特徴量は word2vec+spkmeans+LLC の $k = 5000, n = 2000$ のものである．既存のプレゼンテーションで訓練したモデルを用

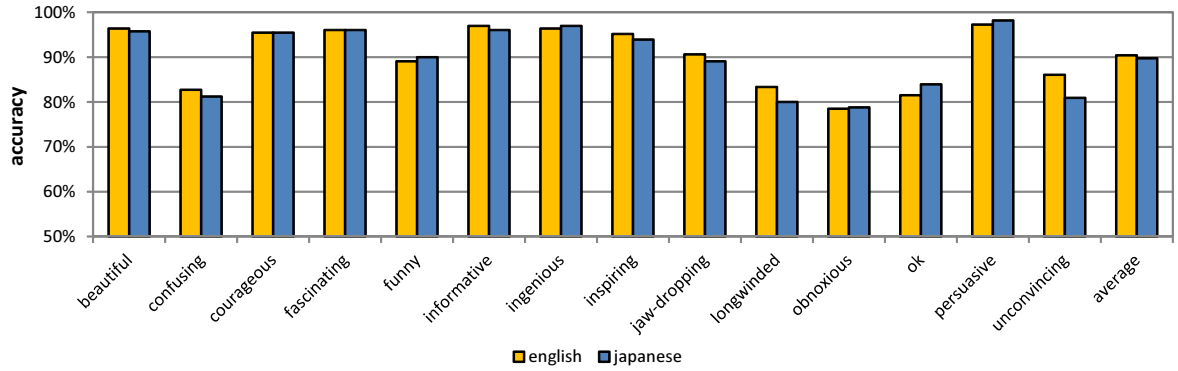


図 2.8: 日本語字幕から作成した特徴量を用いた場合の識別精度

いてユーザーのプレゼンテーションを解析する場合，訓練データについてはプレゼンテーション全体を用いることが可能なため，訓練時はプレゼンテーション全体を用いた場合と訓練時にも冒頭 $N\%$ のみを用いた場合の 2 通りについて実験を行った．いずれの場合についても，$N = 100$ の場合には訓練，テストともにプレゼンテーション全体を用いるため，先の実験と同条件となる．N が大きくなるにつれてやはり精度は向上するが，$N = 100$ のときの精度 90.4% に対し，$N = 40$ のときに精度 86.6% が出ており，プレゼンテーションの半分程度の入力があった時点で全体を用いた場合と遜色ない精度での識別が可能であると考えられる．

2.5.5　日本語のプレゼンテーションへの印象予測に向けた検討

2.4 節で述べたように，TED の多くのプレゼンテーションには多言語の字幕が付与されている．日本語字幕から作成した特徴量を用いて，上位・下位 10% の識別実験を行った結果を図 2.8 に示す．用いた特徴量は日本語，英語共に英語での実験において内容特徴量の中で最も精度の良かった word2vec+spkmeans+LLC の $k = 5000, n = 2000$ のものである．日本語の word2vec のモデルの学習には 2015 年 5 月 10 日時点の日本語 Wikipedia の全記事群を用い，次元数は 512 とした．longwinded と unconvincing の 2 つの印象については精度が大きく下がったものの，その他の印象では元の言語である英語のときと同等の精度が出ている．音節の扱いなどは日本語と英語とでは異なるため，言語的特徴量など一部の特徴量は日本語に適するものを作成する必要はあるものの，本章で述べた手法は日本語のプレゼンテーション動画に対しても有用であると考えられる．また，日本国内にはまだ TED のようにユーザーの抱いた印象のデータが豊富にあるウェブサービスは存在し

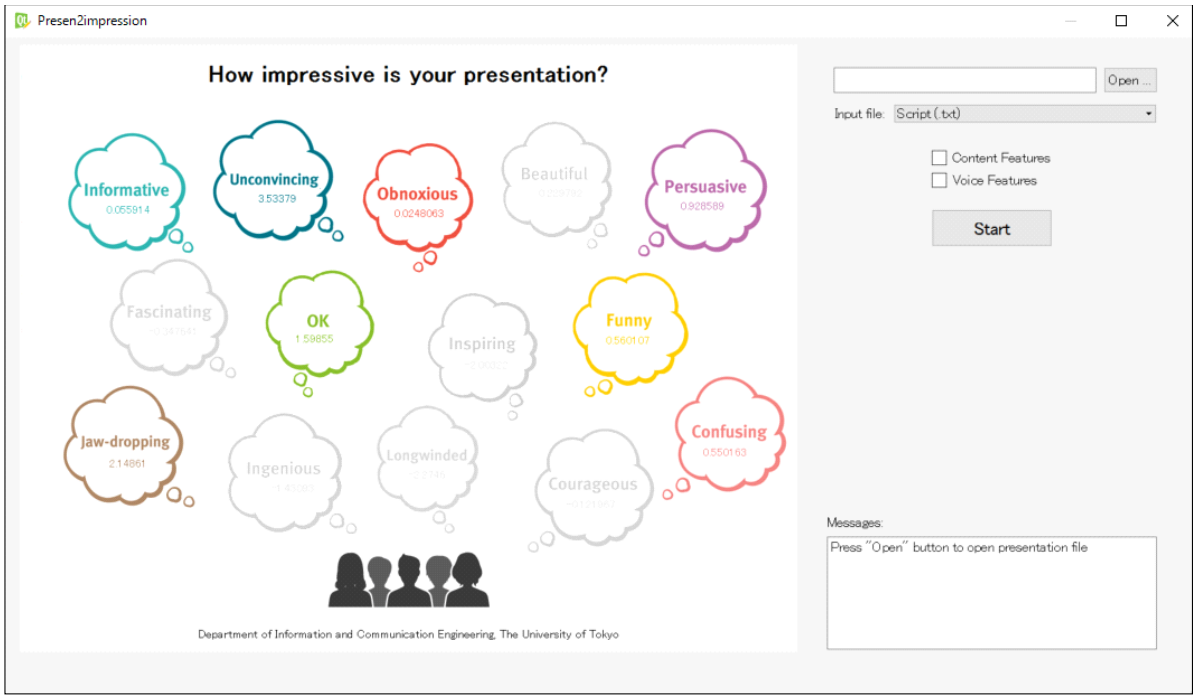


図 2.9: 印象推定ツールの画面

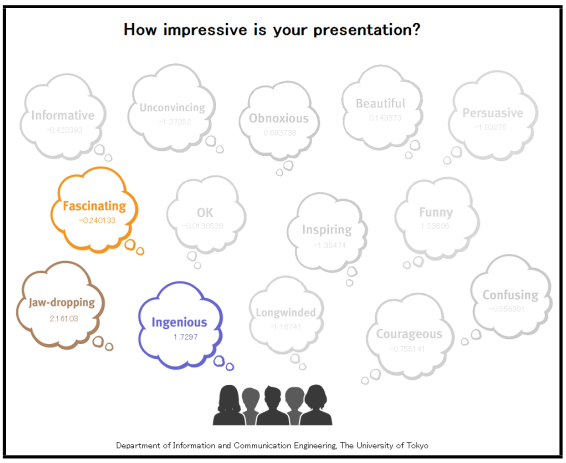
ないため，日本語のプレゼンテーションを本手法を用いて解析したい際，それが TED で行われる類のプレゼンテーションであれば TED の日本語字幕で学習したモデルを用いることが可能であるといえる．

2.6　印象推定ツールの作成

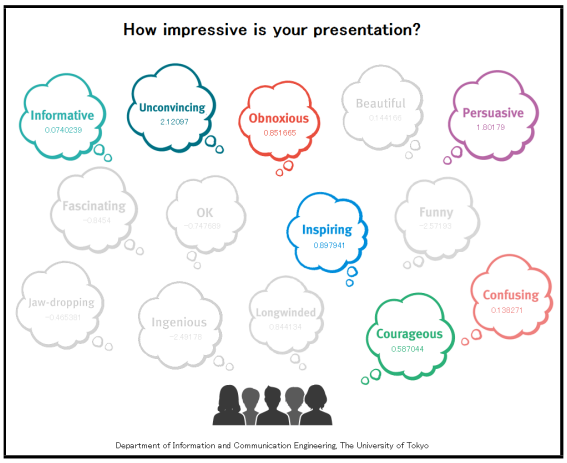
2.6.1　実装

実際にユーザーのプレゼンテーションに対して視聴者がどのような印象を抱くかの予測を行うため，図 2.9 のようなツールの作成を行った．本体は Python[*1]，ユーザーインタフェース部分は Qt[*2]を用いて作成した．ユーザーが解析を行いたいプレゼンテーションを入力すると，前項で述べた手法を用いて，TED の動画群で訓練されたモデルを用いて各印象について聴衆が各印象を抱くか否かを識別し，結果を出力する．なお，高階 MRF を用いて印象間のラベルの関係性，そして内容特徴量と音声特徴量など複数の特徴量を用い

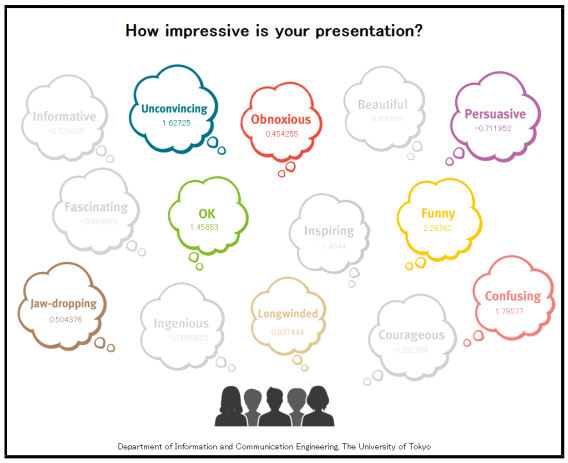
[*1] https://www.python.org/
[*2] https://www.qt.io/



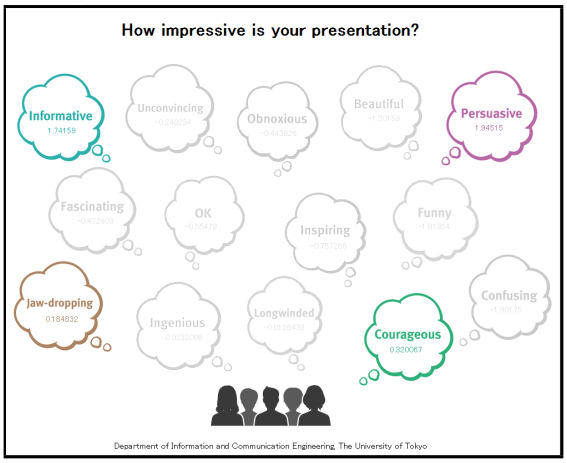
(a) Stave Jobs の iPhone 発表プレゼン



(b) Barack Obama の広島での演説



(c) Will Stephen による TED 風の中身のないプレゼン



(d) Bill Gates による疫病の大流行に関するプレゼン

図 2.10: データセット外のプレゼンテーションの本ツールによる解析結果

た場合には各特徴量間の関係性を考慮することで精度を向上させている．高階 MRF に関する詳細と，高階 MRF を用いたときの本章で行った実験に類似した実験の結果は発表文献 [11, 12, 16] に記されている．

プレゼンテーションの入力としては，音声ファイルとスクリプトファイルの 2 種類に対応している．音声ファイルが入力の場合，IBM Bluemix[*1]の API の一つである Speech to Text を用いた文書化がなされ，その文書から作成された内容特徴量を用いて各印象の推定を行う．

[*1] https://console.ng.bluemix.net/

2.6.2　適用例

本ツールにより，TED の聴衆によって視聴されたときにどのような印象が投稿されるか，という観点から任意のプレゼンテーションの評価が可能となる．図 2.10 は (a) Stave Jobs による iPhone 発表のプレゼンテーション，(b) Barack Obama の広島での演説，(c) Will Stephen による “How to sound smart in your TEDx Talk” というプレゼンテーション，(d) Bill Gates によるエボラ並のアウトブレイクが発生したときの対策に関するプレゼンテーションを上位・下位 10% の識別モデルを用いて本ツールにより解析した出力である．iPhone の発表プレゼンテーションに対しては，ingenious, jaw-dropping といった感嘆を表す複数の印象が予測されたのに対し，広島での原爆投下に関するプレゼンテーションに対しては，persuasive のような良い印象から obnoxious のような悪い印象まで様々な印象が予測された．TED 風プレゼンは，「TEDxNewYork で語られた数々の新しいアイデアを一挙に吹き飛ばすこの珍妙なプレゼンで、面白い人を生業とするウィル・スティーヴンが、何にも話すことがなくてもすごい話をしているように見せられる鉄板のプレゼンスキルを披露します」という動画説明が表すように，全く無意味なグラフや表を提示しながら中身のないプレゼンを 6 分間続けるものである．これに対し，我々のモデルは funny や confusing, 時には obnoxious など，中身は空っぽだが何故か面白い，という人間が持つのと似たような印象群を出力した．Bill Gates のプレゼンテーションについては，persuasive や informative など，聴衆が納得していることを示すような印象を出力した．

2.7　まとめ

本章ではプレゼンテーション動画に対して聴衆が抱く印象の推定手法を提案し，内容や音声から抽出した各種特徴量を用いた実験により精度の比較を行った．この印象推定という問題に対して，word2vec とソフトクラスタリングを組み合わせた手法が従来の文書分類に用いられてきた手法より有効であることを示した．また，プレゼンテーション動画の内容特徴のみ，言語的特徴のみ，そして音声特徴のみを用いて印象の推定を行った場合の結果を示し，各印象に対する各要素の影響を明らかにした．その結果，各印象の投稿率が上位・下位 10% のプレゼンテーションの識別を平均で 91.1% の精度で行えることを確認した．

第 3 章

Schoo の授業に対する受講者数と離脱率の推定

3.1 はじめに

前章では視聴者の抱く印象の解析を行ったが，実世界で行うプレゼンテーションとは異なり，オンラインで発信されるプレゼンテーション動画では視聴者の存在は担保されていない．ユーザーは自由に映像への出入りが可能なため，実世界では不可能なほどの大量の人々に発信することも可能である一方で，誰にも見られずに埋もれてしまう可能性もある．多くの人に映像を見てもらいたいならば，コンテンツ自体の面白さや興味深さとは別に，発信前に視聴者を集めるための工夫が必要となる．本章では国内最大級の MOOC である Schoo で行われる授業に対して，開講前に公開される情報を主に用いて，各授業を受講する人数と授業の途中で退席してしまうユーザーの割合の予測を行う．近年，オンラインコンテンツの作成や投稿，閲覧を容易に行えるサービスの普及とともに，それらのコンテンツの閲覧数やお気に入り数といった人気指標を対象とした研究が数を増している．その対象は，オンライン上のコンテンツがバラエティに富むのと同様に，動画 [36]，画像 [23, 35, 61]，ファッション [60] など多岐に渡る．それらの多くはコンテンツ自身の情報を用いるもの [23] や，投稿直後の一定期間のアクティビティを用いて最終的な人気度を予測する [52] ような投稿後の情報を用いるものである．これに対し，Schoo の授業は生放送であり，コンテンツの中身が揃ったときには授業は終了している．我々の目的はコンテンツのメタデータのみから受講者数と，授業にアクセスはしたものの退室してしまった生徒の数の予測を行うことである．また，放送時間帯や放送する授業のカテゴリ等が受講者数にどれほど影響するのかの要因解析も同時に行う．

edX[*1], Coursera[*2], Udacity[*3]などの代表的な他の MOOC と比べたとき，原則として生放送であるということが Schoo の最も特徴的な点である．受講中のユーザーのコメントを受けてリアルタイムに質問に答えたり，授業の内容を変化できるという利点がある一方で，どんなに授業の中身が素晴らしくても評判が広がる頃には授業は終了しているため，授業の開始前に多くのユーザーに興味を持ってもらう必要がある．次節で述べる MOOC を対象とした関連研究では，講義は録画された映像であることを前提としているものが多い一方で，本研究では授業の開始前にわかる情報のみを用いた解析に注力する．

3.2　関連研究

3.2.1　オンライン動画の閲覧者数の推定

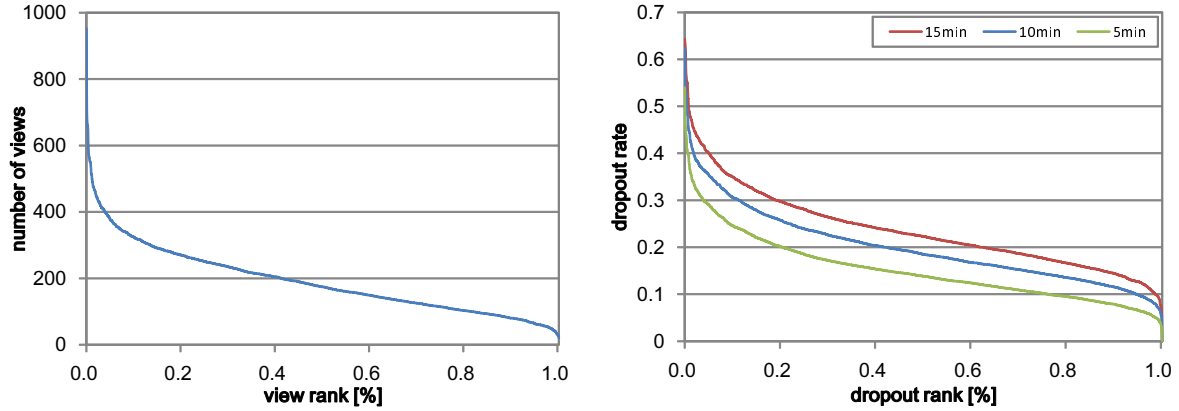
映像の中でも特にオンライン動画サービスの閲覧数や人気度を対象とした研究では，映像そのものの情報だけでなく，それがオンラインコンテンツであるという特徴を活かして Twitter[*4]での話題性など他のソーシャルメディアのデータを活用するもの [1, 58, 63] や，YouTube[*5]など特定のサービスに着目し，推薦システムなどそのサービスの特有の機能の情報を用いたもの [63, 65] が近年その数を増している．これはコンテンツ自体の魅力よりも，話題性や拡散度がオンラインコンテンツの閲覧数には重要なためである．これらの手法は特定のサービスについて高い精度で予測が行えるものの，機能やユーザインタフェースの異なる他のサービスに適用することは想定されておらず，容易ではない．

3.2.2　MOOC を対象とした研究

本稿と同様に生放送の授業データを利用した研究として，He[26] は大学が生配信する講義について，それをオンラインで受講した学生のチャットメッセージ（学生同士のやり取りと，先生と生徒のやり取りの両方を含む）に対してテキストマイニングを行い，質問回数と成績との相関などを調べた．これは実データに対する解析であり，モデルを構築して何かの予測を行うものではない．こうした解析を中心とするオンラインの教育動画を対象とした研究は，教育分野で数多く行われている [31, 46].

MOOC の離脱率について議論を行う際，二つの定義が存在する．一つはある授業の動

[*1] https://www.edx.org/
[*2] https://www.coursera.org/
[*3] https://www.udacity.com/
[*4] https://twitter.com/
[*5] https://www.youtube.com/



(a) 閲覧者数　　(b) 離脱者割合

図 3.1: Schoo の授業群の閲覧者と離脱率の傾向

画中に視聴をやめてしまうユーザーの割合を示す場合，もう一つは大学の講義のように複数回の授業からなるコースについて，コースの途中で履修をやめてしまうユーザーの割合を示す場合である．本章で行うのは前者の値の予測である．

前者の意味での離脱率をはじめて対象とした研究として，Kim ら [30] は MOOC の一つである edX の 862 の授業に対して，ビデオ長と離脱率との関係などの解析をはじめとして，自分の好きなペースで受講が可能であるという MOOC の特性に着目し，ユーザーが映像のスキップや巻き戻し，停止を行うタイミングの解析や，初受講生と再受講生の傾向の違いの解析を行った．ただし，これは離脱率の予測を行うものではない．

後者の離脱率の予測については，多くの手法が提案されてきた [33, 38, 45, 62]．例えば Li ら [38] は，授業を受ける，授業ページを閉じる，授業のフォーラムへのアクセスなど 6 種類の各ユーザーの行動を一週間ごとに区切ったものを特徴量とし，0.9 以上の高い F 値でコースから脱落するユーザーの識別が可能であることを示した．

3.3　データセット

本章では Schoo で放送された 3,000 以上の授業のうち，2012 年 12 月 11 日から 2016 年 5 月 2 日の間に行われた，授業時間が 30 分以上 120 分以下のもの 2,327 本を実験に用いた．図 3.1 はデータセット内の 2,327 本の授業群について，横軸に閲覧者数あるいは各時間以内の離脱率が高い割合に並べたときの順位，縦軸に閲覧者数あるいは離脱率を示したグラフである．例えば閲覧者数について，中央値は横軸が 0.5 のときの値である 180 人程

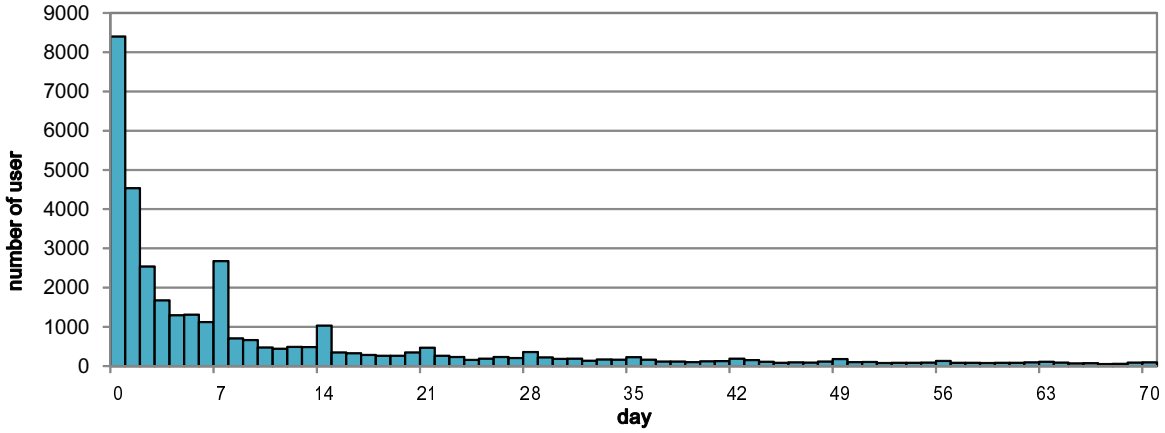


図 3.2: 各ユーザーの初受講から二度目の受講までの間隔（70 日まで）

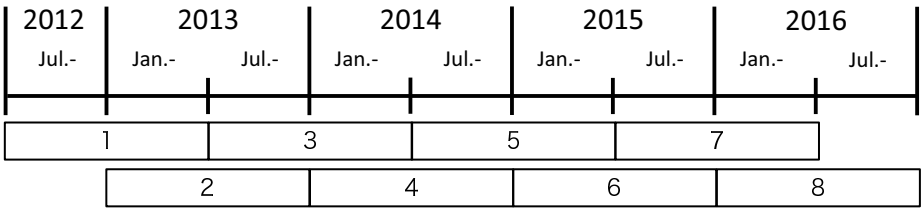


図 3.3: 授業時期を表す 8 次元の特徴量の作成

度であることがわかる．離脱率は，授業にアクセスした全ユーザーのうち，再生から 5 分以内，10 分以内，15 分以内に退室したユーザーの割合を示す．

2017 年 2 月現在，Schoo では会員登録をしたユーザーのみが授業を受講することができ，生放送については全ユーザーが，放送終了後の録画授業については無料ユーザーは月 1 本まで，有料ユーザーは何本でも視聴することができる．本稿では生放送の視聴データのみを扱う．データセット外の授業，すなわち 30 分以上 120 分以下のもの以外も含む授業から 2 本以上の受講をしたユーザーについて，1 本目から 2 本目までに空く期間のヒストグラムを図 3.2 に示す．同日に複数の授業にアクセスするユーザーが最も多く，その後は 7 日，14 日，21 日と 7 日ごとにピークがみられる．毎週日曜日など，決まった曜日に受講するユーザーが多いことがわかる．

3.4　特徴ベクトルの作成

Schoo の各授業の情報を表現するため，本稿では 8 種類の特徴量を用いた．うち 6 種類（番組表特徴量，カテゴリ特徴量，出演者特徴量，表題特徴量，サムネイル特徴量，アク

表 3.1: Schoo の番組表特徴量の詳細

特徴量	次元	説明
開始時刻	13	午前 10, 11, 12 時, 午後 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 時
時間帯	3	日中（-午後 3 時），夕方（午後 3 時-午後 7 時），夜（午後 7 時-）
放送時期	8	図 3.3 参照
授業の長さ	6	30-60，45-75，60-90，75-105，90-120，105-135 分
曜日	7	日，月，火，水，木，金，土
他授業	1	同日の他授業の有無
計	38	

ティブユーザー特徴量）は授業の放送前に入手が可能である特徴量，2 種類（内容特徴量，音声特徴量）は授業コンテンツから抽出される特徴量である．本節ではこれらの特徴量の詳細について述べる．

3.4.1　番組表特徴量

番組表特徴量の詳細を表 3.1 に示す．これは開始時刻や授業の長さ，曜日など放送枠に関する情報を表す特徴量である．それぞれの特徴量は表中に示されている値に量子化されており，各次元は値として 1 か 0 のいずれかを持つ．Schoo というサービスの成長に伴い閲覧者数は増大する傾向にあると考えられるため，放送時期に関しては図 3.3 に示す 8 次元の特徴量を用いた．例えば 2013 年 7 月に放送された授業であれば，対応する次元である 2, 3 次元目のみが 1 となる．

3.4.2　カテゴリ特徴量

Schoo の各授業にはカテゴリが設定されており，6 つの大カテゴリ（デザイン，WEB 開発，ビジネス，英語，スタートアップ，教養）とその下に 100 を超える小カテゴリ（経済，アート，PHP など）が存在する．大カテゴリと小カテゴリのそれぞれについて，各授業が属するカテゴリのみが 1 を持つ 1-of-K ベクトルとして表現したものがカテゴリ特徴量である．本章の実験に用いたデータセット内には 115 の小カテゴリが存在したため，大カテゴリとあわせて 121 次元の特徴量である．

(a) Schoo のトップページ　　(b) 各授業のページ

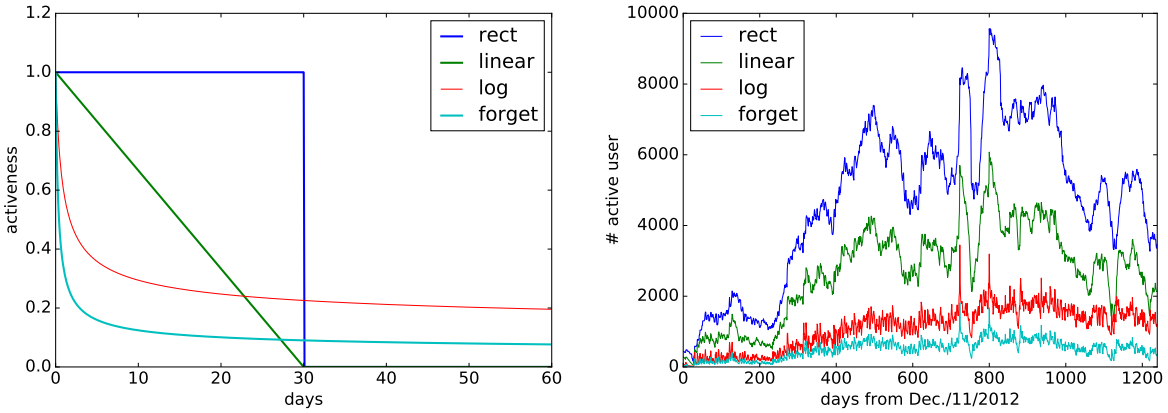
図 3.4: Schoo の授業群の閲覧者と離脱率の傾向

3.4.3　出演者特徴量

授業を行う先生はもちろん，Schoo では一部の授業についてそのアシスタントを行う学生代表が存在する．彼らを自身に対応する次元のみが 1 となる 1-of-K ベクトルで表し，授業ごとに出演者のベクトルの和をとったものが出演者特徴量である．データセット内の授業のうち 1 本のみしか担当していない先生と学生代表は除いた結果，439 人の先生と 16 人の学生代表からなる 455 次元の特徴量である．

3.4.4　表題特徴量

図 3.4 に Schoo のトップページと各授業ページを示す．トップページには本日の放送スケジュールが記載されており，授業タイトルとサムネイル画像が表示される．各授業ページにはさらに授業の概要を表す文書が記載されている．このタイトルと概要について，2 章で用いた内容特徴量の一つである word2vec と同様の手法で抽出した特徴量が表題特徴量である．Word2vec のモデルの学習には 2015 年 5 月 10 日時点の日本語 Wikipedia の全記事群を用い，次元数は 512 とした．タイトルと概要の特徴量を結合した 1,024 次元の特徴量である．



(a) アクティブ度合いを表す関数　　(b) 各関数での日毎のアクティブユーザー数

図 3.5: アクティブユーザー数の算出

3.4.5　サムネイル特徴量

図 3.4 に示す各授業ページに掲載される，授業のイメージ画像を用いて作成したものがサムネイル特徴量である．Imagenet[17] で学習された GoogLeNet[55] を用いて，この画像を入力としたときに最後の全結合層への入力となる 1,024 次元の特徴量をサムネイル特徴量として用いた．なお，画像は左右に空白領域がある場合は除去した後，224×224 に縮小を行ったものを入力とした．

3.4.6　アクティブユーザー特徴量

Schoo は会員制のサービスのため，閲覧者数の上限は登録されたユーザー数である．そのうち何人がアクティブユーザーかは定かではないが，授業の放送自転でどの程度のユーザーが Schoo を定期的に利用しているかの概数は，授業の閲覧者数を予測するための要素として活用できると考えられる．各ユーザーに対して，最後に Schoo の授業を受講してから一定期間はアクティブユーザーとみなせるという仮定をおき，日毎のアクティブユーザー数の算出を行い，特徴量として用いた．各ユーザーのアクティブ度合いの算出として，最後に受講してからの日数を x として以下の 4 つを用いた．$x \leq 60$ の概形と，各ユーザーのアクティブ度合いの和をとることで算出したアクティブユーザー数を図 3.5 に示す．データセット外の授業も含めて視聴データの存在する最も古い授業の放送日が 2012 年 12

月 11 日のため，アクティブユーザー数はこれ以降のものが算出される．

1. $1\ (x \leq 30)$（受講から 30 日間はアクティブユーザー）（rect）
2. $1 - \dfrac{x}{30}\ (x \leq 30)$（線形にアクティブ度合いが減少）（linear）
3. $\dfrac{1}{\log(x+1)+1}\ (\leq 1000)$（非線形にアクティブ度合いが減少）（log）
4. $\dfrac{1.023}{\log_{2.166}(x+1)^{2.324}+1.023}\ (x \leq 1000)$（図 3.2 を曲線回帰したもの）（forget）

これらを用いて各授業のアクティブユーザー特徴量の算出を行う．各授業の放送日の直前の n 週間について，図 3.5b のアクティブユーザー数を一週間ごとに区切り，各週の和を取ることで n 次元の特徴量を得ることができる．これは絶対値であり，放送直前に Schoo を利用するユーザーがどの程度いるかの指標となる．また，放送 n 週前の一週間のアクティブユーザー数の和に対してのその後の $n-1$ 週間分の各週の和の比率を取ることで，$n-1$ 次元の特徴量を得ることができる．これは相対値であり，放送前の n 週間の Schoo を利用するユーザー数の推移の指標となる．これらの特徴量を，3.4.2 項で述べた大カテゴリ 6 つと，その総和について個別に算出することで $7n + 7(n-1)$ 次元の特徴量が得られる．また，自身が属するカテゴリの特徴量を授業間で同じ次元に対応させるため，複製して末尾につけることで，$8n + 8(n-1)$ 次元となる．

さらに，例えば WEB 開発に人気が集中しているなど，各カテゴリの人気度を考慮するため，大カテゴリ 6 つの一週間のアクティブユーザーの総和を，全アクティブユーザー数の一週間の総和で割ることで，6 次元の特徴量が得られる．自身が属するカテゴリのものを末尾につけることで 7 次元となる．これを n 週間に渡って算出することで，$7n$ 次元の特徴量が得られる．すなわち，各授業について計 $23n - 8$ 次元の特徴量が得られる．

3.4.7　内容特徴量

内容特徴量の算出のため，IBM Bluemix の API の一つである Speech to Text を用いて授業動画の全スクリプトの作成を行った．これを 3.4.4 項で述べたのと同様の手法により，512 次元の特徴量化したものが各授業の内容特徴量である．

3.4.8　音声特徴量

項で述べたのと同様の手法で特徴量を抽出したものが各授業の音声特徴量である．次元数も同様であり，6,373 次元である．

3.5　実験

3.5.1　手法

閲覧者数と離脱率の予測モデルとして，線形サポートベクター回帰（SVR）を用いた．精度の検証について，交差検定が用いられることが多いが，本章で扱う授業データは時系列の情報を持つデータであり，単なる交差検定は適さない．例えば 2014 年に放送された授業の予測は，それ以前に放送された授業のみを用いて訓練されたモデルにより行われるべきである．同様に，ある授業の後編の閲覧者数を前編を含むモデルで予測するのは適切であるが，その逆は不適切である．本章では精度の検証において，訓練時のパラメータの決定の際に以下に示す時系列を考慮した交差検定 [25] を用いた．

1. n 本の授業群 $D = d_1, d_2, ..., d_n$ を時系列順に並べる．
2. 回帰モデルを $d_1, ..., d_t$ から作成し，d_{t+1} の閲覧者数（離脱率）の予測を行う．ただし t は時系列を表す．
3. 手順 2 を $t = k, ..., n-1$ について繰り返す．ただし k は回帰モデルを作成するのに最低限必要なドラマ数とする．
4. $d_{k+1}, ..., d_n$ について，実際の閲覧者数（離脱率）と予測閲覧者数（離脱率）との間の Mean Squared Error (MSE) を計算する．

本章では $k = 215$ とした．これはデータセット内の最初の 1 年間（2012 年 12 月 11 日から 2013 年 12 月 10 日）の授業数である．データセット内の 2,327 本の授業のうち，前述の 215 本を除いた残りの 2,112 本が予測の対象となる．

各授業の閲覧者数について，本章では以下の 2 通りの方法で算出を行った．総ユーザー数は授業の長さの影響を強く受けると考えられるため，ある時点での瞬間の閲覧者数として，中間時点の閲覧者数を用いた．

- 授業にアクセスした総ユーザー数（全閲覧者数）
- 授業の中間時点の閲覧者数（30 分授業なら 15 分時点，60 分授業なら 30 分時点）（中間閲覧者数）

各ユーザーの閲覧の算入について，Schoo のサービスの存在を知り登録したユーザーは，その当日に行われる授業を興味の有無に関わらず視聴する，すなわち各ユーザーの初回の視聴はノイズであるという仮定を置き，以下の 2 通りの算入を行った．

表 3.2: 番組表特徴量のみを用いた場合の結果

	RMSE	相関係数
全閲覧者数&全履歴	93.5	0.45
中間閲覧者数&全履歴	71.1	0.41
全閲覧者数&初回除外履歴	79.0	0.47
中間閲覧者数&初回除外履歴	61.7	0.44

表 3.3: 番組表特徴量にアクティブユーザー特徴量を加えたときの RMSE

n	rect	linear	log	forget
1	76.9	77.0	76.9	76.8
2	77.0	77.0	76.9	76.6
3	76.9	77.1	76.9	76.7
4	76.9	77.0	77.0	76.8
5	76.7	76.9	76.9	76.8
6	76.5	76.8	76.7	76.8
7	76.3	76.7	76.6	76.8
8	**76.2**	76.6	76.4	76.8

- 各ユーザーの全視聴履歴を用いる（全履歴）
- 各ユーザーの二本目以降の視聴履歴を用いる（初回除外履歴）

前述の 2 点を考慮し，閲覧者数の予測では予測する値として計 4 種類が存在する．

離脱率の予測においては，授業にアクセスした総ユーザー数のうち，授業にアクセスしてから一定時間以内に離脱したユーザーの割合を予測する値とする．本稿ではアクセスから 5 分，10 分，15 分以内に授業から切断したユーザーを離脱ユーザーとした．各ユーザーの初回の視聴を用いるか否かで予測する値として計 2 種類が存在する．

3.5.2　閲覧者数の予測

3.4 節で述べた特徴量のうち，閲覧者数の予測は授業の放送前に入手が可能である番組表特徴量，カテゴリ特徴量，出演者特徴量，表題特徴量，サムネイル特徴量，アクティブユーザー特徴量の 6 つを用いて行った．番組表特徴量をベースとして，他の特徴量を加え

表 3.4: 閲覧者数予測の RMSE（全閲覧者数&全履歴）

RMSE	カテゴリ無, 出演者無	カテゴリ無, 出演者有	カテゴリ有, 出演者無	カテゴリ有, 出演者有
表題無, サムネイル無	90.6	81.8	82.2	78.2
表題無, サムネイル有	86.2	84.1	82.0	80.5
表題有, サムネイル無	79.7	78.2	77.9	**76.5**
表題有, サムネイル有	80.0	79.1	78.5	77.6

表 3.5: 閲覧者数予測の RMSE（中間閲覧者数&全履歴）

RMSE	カテゴリ無, 出演者無	カテゴリ無, 出演者有	カテゴリ有, 出演者無	カテゴリ有, 出演者有
表題無, サムネイル無	69.3	61.6	62.5	58.7
表題無, サムネイル有	65.4	63.3	61.8	60.4
表題有, サムネイル無	60.0	58.5	58.5	**57.1**
表題有, サムネイル有	60.1	59.2	58.7	57.9

表 3.6: 閲覧者数予測の RMSE（全閲覧者数&初回除外履歴）

RMSE	カテゴリ無, 出演者無	カテゴリ無, 出演者有	カテゴリ有, 出演者無	カテゴリ有, 出演者有
表題無, サムネイル無	76.2	66.6	66.7	62.8
表題無, サムネイル有	70.7	68.1	66.0	64.5
表題有, サムネイル無	64.6	62.8	62.6	**61.2**
表題有, サムネイル有	64.5	63.5	62.8	61.9

表 3.7: 閲覧者数予測の RMSE（中間閲覧者数&初回除外履歴）

RMSE	カテゴリ無, 出演者無	カテゴリ無, 出演者有	カテゴリ有, 出演者無	カテゴリ有, 出演者有
表題無, サムネイル無	59.8	51.2	52.2	48.2
表題無, サムネイル有	55.0	52.4	51.1	49.5
表題有, サムネイル無	49.7	48.1	48.2	**46.8**
表題有, サムネイル有	49.7	48.7	48.2	47.4

ていく形式で実験を進めた．番組表特徴量単体での実験結果を表 3.2 に示す．予測する値が異なるため，RMSE の比較は不可能だが，相関係数を比較すると各ユーザーの初回の視聴を用いない方が精度が良いことがわかる．すなわち，初回の視聴の多くはノイズであると考えられる．

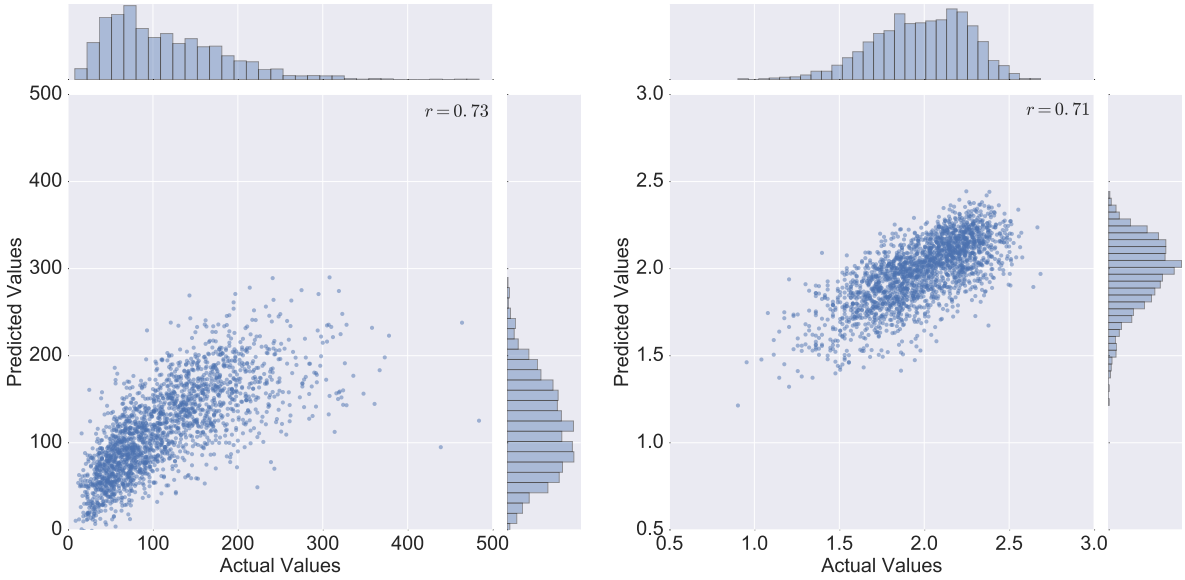
アクティブユーザー特徴量にはパラメータが存在するため，この選定を行うため，はじめに番組表特徴量にアクティブユーザー特徴量を加えた実験を行った．番組表特徴量単体での相関係数が最も良かった全閲覧者数&初回除外履歴の予測について，アクティブユーザー特徴量を加えたときの結果を表 3.3 に示す．データセット内の時間的に最初の授業について，最長でその 8 週前までのアクティブユーザー数が計算可能であったため，n を

1 から 8 まで変化させて実験を行った．最も RMSE の小さかったのは各ユーザーのアクティブ度合いを 3.4.6 項で述べた rect 関数で算出した際の $n = 8$ であった．これを閲覧者数の予測に用いるアクティブユーザー特徴量とする．

番組表特徴量と前述のアクティブユーザー特徴量に，カテゴリ特徴量，出演者特徴量，表題特徴量，サムネイル特徴量の 4 つを加えたときの結果をそれぞれ表 3.4, 3.5, 3.6, 3.7 に示す． 4 つのうちいずれかを単体で加えた場合，表題特徴量が最も精度の向上に貢献していることがわかる．他の特徴量と比べ，授業のタイトルや概要はその授業の内容を直接表現する文章であり，これを密に表現した特徴量が表題特徴量であるため，予測への影響が大きかったと考えられる．各単体で特徴量を加えた際，どの特徴量も精度を上げる働きをしているものの，すべての特徴量を考慮した場合，いずれの場合もサムネイル特徴量以外を連結した場合が最も精度が良いことがわかる．

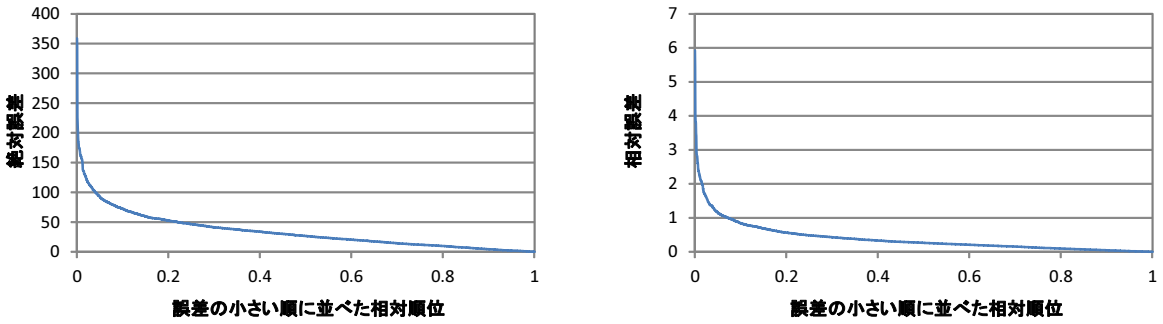
実際の閲覧数と予測閲覧数の予測例として，中間閲覧者数&初回除外履歴の場合の，サムネイル特徴量以外の特徴量を用いた場合の散布図を図 3.6a に示す．以下ではこの予測の結果について議論を行う．このときの相関係数は $r = 0.73$ であった．どの程度の誤差に収まった授業がどの程度あるのかの傾向を図 3.7 に示す．横軸が各授業を誤差順に並べたときの順位，縦軸が誤差を示す．絶対誤差について，約 8 割の授業は誤差が 50 人以内に収まっている一方で，図 3.6a の散布図からもわかるように一部の授業に対しては大きな誤差が生じている．実際の閲覧者数については 300 を超える授業がいくつかあるのに対し，予測値はすべて 300 以下であり，これらの人気授業の一部について予測が正しく行えていない．相対誤差について，約半数の授業で誤差は実際の閲覧者数の 4 分の 1 以下，約 7 割の授業で誤差は実際の閲覧者数の半分以下である．一部の授業について，誤差が 400% を超えているようなものがあるが，これは実際の閲覧者数が 10 人程度のため，割合を計算したときに誤差が大きくなってしまった授業群である．

実際の閲覧者数よりもモデルが閲覧者数を少なく予測してしまう授業群として，我々が特徴量として用いた情報以外，すなわち Schoo の外部の情報が閲覧者数に影響したものが見られる．特に先生が有名人であるものとして，タレントの池澤あやかが担当した 2 本の授業では実際の閲覧者数が 276 人と 290 人に対し，予測閲覧者数は 168.0 人と 180.6 人であった．お笑い芸人の西野亮廣が担当した授業では実際の閲覧者数 90 人に対して予測値が 47.3 人，教育評論家の尾木直樹が担当した授業では実際の閲覧者数 141 人に対して予測値が 61.9 人と，有名人が突発的に行う授業については実際よりも低く予測を行ってしまう傾向が見られた．彼らが Twitter など外部ツールで自分の授業の告知を行うことの影響や，有名人の単発授業においては出演者特徴量がうまく働かないことが原因と考えられる．



(a) 対数を取らない場合の閲覧者数の予測　　(b) 対数を取ったときの閲覧者数の予測

図 3.6: 実際の閲覧者数（横軸）と予測閲覧者数（縦軸）の散布図



(a) 図 3.6a の結果を絶対誤差順に並べた順位（横軸）と絶対誤差（縦軸）

(b) 図 3.6a の結果を相対誤差順に並べた順位（横軸）と絶対誤差を実際の閲覧者数で割った相対誤差（縦軸）

図 3.7: 図 3.6a の誤差の傾向

表 3.8: 閲覧者数に影響を与えた上位 30 要因

特徴量	重み
横田幸信（先生）	+51.4
たにぐちまこと（先生）	+50.1
堀口誠人（先生）	+28.1
日比野ななえ（先生）	+26.6
HTML/CSS（小カテゴリ）	+25.3
大倉芙美子（先生）	+24.0
ライフハック/仕事術（小カテゴリ）	+22.8
ウェブ解析（小カテゴリ）	+21.5
デザイン（大カテゴリ）	+21.4
まきのゆみ（先生）	+21.4
宇都出雅巳（先生）	+21.3
谷本有香（先生）	+21.0
鈴木満里乃（学生代表）	+20.6
小川卓（先生）	+19.8
日曜日に放送	+19.0
上野美香（先生）	+18.7
間島ゆかり（先生）	+18.7
朽木誠一郎（先生）	+18.6
塩原桜（学生代表）	+18.5
淵上真一（先生）	+18.4
シバタアキラ（先生）	+18.0
大串肇（メガネ）（先生）	+17.5
有山圭二（先生）	+ 17.3
放送時間 75-105 分	+17.1
英語学習法（小カテゴリ）	+16.6
池澤あやか（先生）	+16.5
大木しのぶ（学生代表）	+16.3
浅野桜（先生）	+16.0
社会（小カテゴリ）	+15.4
ハードウェア（小カテゴリ）	15.3

表 3.9: 閲覧者数に影響を与えた下位 30 要因

特徴量	重み
Java（小カテゴリ）	-26.8
I（先生）	-26.4
F（先生）	-25.7
M（先生）	-25.7
Android Studio（小カテゴリ）	-24.5
プロジェクトマネジメント（小カテゴリ）	-22.2
教養（大カテゴリ）	-21.4
K（先生）	-17.1
N（先生）	-17.0
K（先生）	-16.6
T（先生）	-16.5
I（先生）	-16.5
Word（小カテゴリ）	-16.1
iOS（小カテゴリ）	-15.8
放送時間 30-60 分	-15.8
数学/データ分析（小カテゴリ）	-15.5
N（先生）	-15.4
T（先生）	-15.4
T（先生）	-15.4
T（先生）	-15.4
O（先生）	-15.2
ビジネスシーン別（小カテゴリ）	-14.9
同日に他授業あり	-14.5
Y（先生）	-14.5
N（先生）	-14.5
N（先生）	-14.4
K（先生）	-14.3
S（先生）	-14.0
N（先生）	-13.9
K（先生）	-13.7

表 3.10: 5 分以内の離脱率予測の RMSE（初回除外履歴）

RMSE	カテゴリ無, 出演者無	カテゴリ無, 出演者有	カテゴリ有, 出演者無	カテゴリ有, 出演者有
表題無, サムネイル無	0.0708	0.0692	0.0681	0.0674
表題無, サムネイル有	0.0709	0.0705	0.0693	0.0690
表題有, サムネイル無	0.0665	0.0670	0.0664	**0.0662**
表題有, サムネイル有	0.0673	0.0674	0.0668	0.0669

表 3.11: 10 分以内の離脱率予測の RMSE（初回除外履歴）

RMSE	カテゴリ無, 出演者無	カテゴリ無, 出演者有	カテゴリ有, 出演者無	カテゴリ有, 出演者有
表題無, サムネイル無	0.0794	0.0771	0.0758	0.0752
表題無, サムネイル有	0.0788	0.0781	0.0773	0.0770
表題有, サムネイル無	0.0730	0.0726	0.0727	**0.0725**
表題有, サムネイル有	0.0742	0.0741	0.0740	0.0739

表 3.12: 15 分以内の離脱率予測の RMSE（初回除外履歴）

RMSE	カテゴリ無, 出演者無	カテゴリ無, 出演者有	カテゴリ有, 出演者無	カテゴリ有, 出演者有
表題無, サムネイル無	0.0838	0.0807	0.0799	0.0788
表題無, サムネイル有	0.0829	0.0821	0.0810	0.0806
表題有, サムネイル無	0.0768	0.0765	0.0762	**0.0760**
表題有, サムネイル有	0.0775	0.0772	0.0769	0.0768

3.5.3　閲覧者数への各要素の影響

回帰モデルとして線形 SVR を用いたため，各次元に対応する重みを得ることができ，それが予測閲覧者数への影響の度合いとなる．表 3.8 に中間閲覧者数&初回除外履歴の予測モデルの重み上位 30 件，表 3.9 に下位 30 件を示す．ただし，表題特徴量のような密な特徴量は除いた，1-of-K ベクトルとして表現された特徴量のみ抽出したものである．負の影響を特定の人物の出演が与えている場合は匿名化を行う．ただし閲覧数へ大きな影響を与えることは，その先生の担当する講義のターゲット層が広い，あるいは狭いなど様々な要因が考えられ，その要因が原因で閲覧者が増減することを直接示すものではない．上位も下位もその多くを先生が占めていることがわかる．

3.5.4　放送前情報のみを用いた離脱率の予測

閲覧者数予測の結果から各ユーザーの初回の視聴はノイズと考え，これを除いた視聴履歴について実験を行う．また，閲覧数とは異なり離脱率はその授業を閲覧した人数が母数

表 3.13: 授業中の特徴量を用いた 5 分以内の離脱率予測の RMSE（初回除外履歴）

特徴量	内容無, 音声無	内容無, 音声有	内容有, 音声無	内容有, 音声有
RMSE	0.0662	0.0659	**0.0656**	0.0661
相関係数	0.499	0.500	**0.517**	0.506

表 3.14: 授業中の特徴量を用いた 10 分以内の離脱率予測の RMSE（初回除外履歴）

特徴量	内容無, 音声無	内容無, 音声有	内容有, 音声無	内容有, 音声有
RMSE	0.0725	0.0727	**0.0721**	0.0724
相関係数	0.519	0.505	**0.530**	0.514

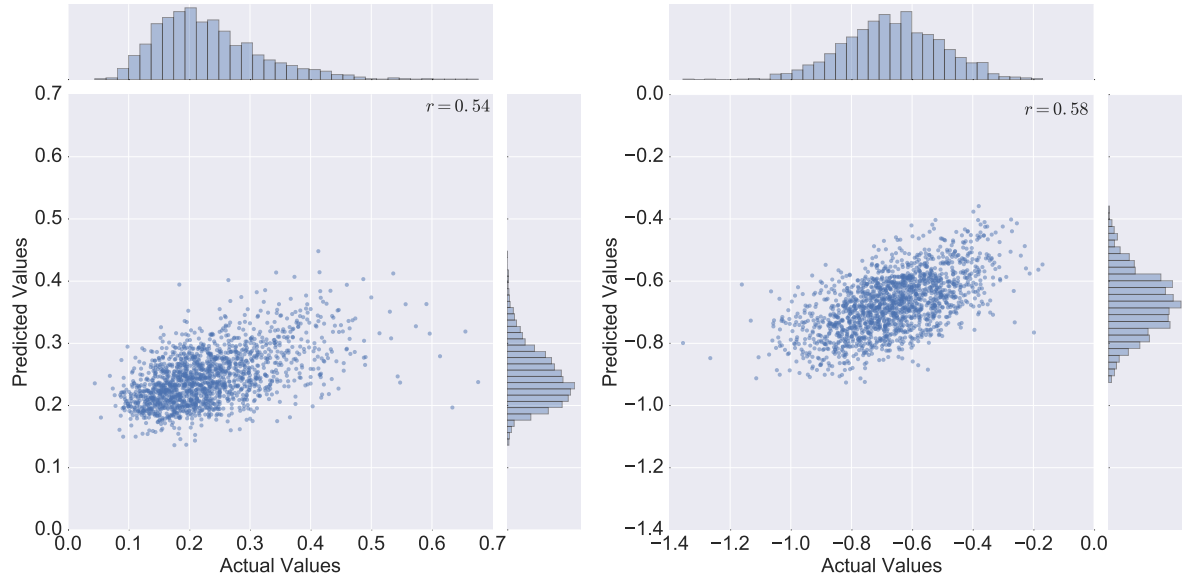
表 3.15: 授業中の特徴量を用いた 15 分以内の離脱率予測の RMSE（初回除外履歴）

特徴量	内容無, 音声無	内容無, 音声有	内容有, 音声無	内容有, 音声有
RMSE	0.0760	0.0756	0.0751	**0.0750**
相関係数	0.527	0.529	0.542	**0.543**

であり，アクティブユーザー数の影響はないと考えられるため，アクティブユーザー数は含めず番組表特徴量に他の 4 種類（カテゴリ，出演者，表題，サムネイル）の特徴量を加えた実験を行う．なお，実際には 15 分以内の離脱率の予測において，$n = 1$ の linear 関数を用いて計算したアクティブユーザー特徴量を用いた場合にわずかな精度の向上が見られたが，$n = 1$ は本来のアクティブユーザー特徴量の目的を反映しているとはいえないこと，向上率が RMSE の有効数字 3 桁目に影響する程度であることから誤差の範囲と考える．結果を表 3.10, 3.11, 3.12 に示す．閲覧者数の予測と同様に，離脱率の予測でもサムネイル特徴量以外のすべてを用いたときが最も良い精度を得られた．

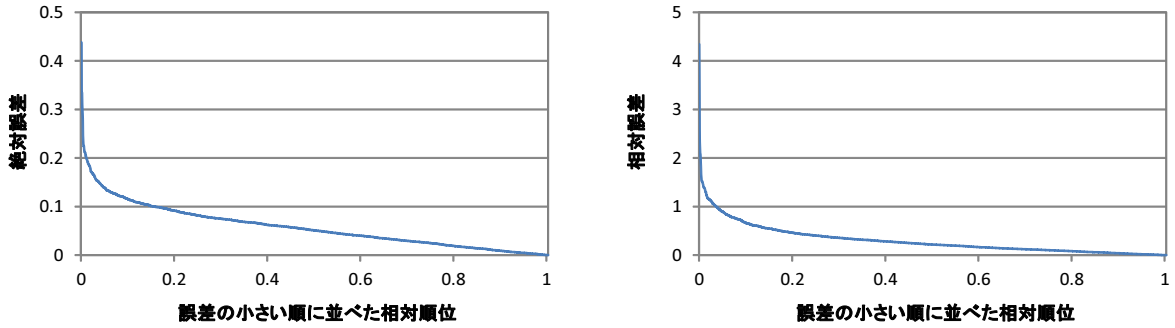
3.5.5　授業映像の情報も用いた離脱率の予測

閲覧と異なり，授業からの離脱は授業に一旦アクセスして視聴してからのアクションであるため，授業の内容に関する情報が影響すると考えられる．3.4 節で述べた内容特徴量と音声特徴量も用いた場合の結果を表 3.13, 3.14, 3.15 に示す．わずかではあるが，内容特徴量と音声特徴量のいずれか，あるいは両方を加えることで精度の向上が見られる．15 分以内の離脱率予測において，内容と音声も含む全特徴量を用いたときの予測の散布図を図 3.8a に示す．また，どの程度の誤差にどの程度の授業が収まっているかの傾向を図 3.9 に示す．



(a) 対数を取らない場合の離脱率の予測　　(b) 対数を取ったときの離脱率の予測

図 3.8: 実際の離脱率（横軸）と予測離脱率（縦軸）の散布図



(a) 図 3.8a の結果を絶対誤差順に並べた順位（横軸）と絶対誤差（縦軸）

(b) 図 3.8a の結果を相対誤差順に並べた順位（横軸）と絶対誤差を実際の離脱率で割った相対誤差（縦軸）

図 3.9: 図 3.8a の誤差の傾向

表 3.16: 離脱率に影響を与えた上位 10 要因

特徴量	重み
同日に他授業あり	+0.0124
2014 年 7 月-2015 年 6 月に放送	+0.0120
午後 7 時に開始	+0.0119
水曜日に放送	+0.0114
2015 年 1 月-2015 年 12 月に放送	+0.0101
教養（大カテゴリ）	+0.00959
夜（午後 7 時から午後 11 時）に放送	+0.00845
午後 8 時に開始	+0.00772
O（先生）	+0.00631
O（学生代表）	+0.00591

表 3.17: 離脱率に影響を与えた下位 10 要因

特徴量	重み
午後 10 時に開始	-0.00930
デザイン（大カテゴリ）	-0.00801
2013 年 7 月-2014 年 6 月に放送	-0.00744
日曜日に放送	-0.00503
2012 年 7 月-2013 年 6 月に放送	-0.00495
土曜日に放送	-0.00470
昼（午前 10 時から午後 2 時に放送）	-0.00453
スタートアップ（大カテゴリ）	-0.00453
SEO（小カテゴリ）	-0.00434
小林あつし（先生）	-0.00395

3.5.6　離脱率への各要素の影響

閲覧者数について行ったのと同様に，離脱率に影響を与えた上位・下位 10 件を表 3.16, 3.17 に示す．ただし 1-of-K ベクトルで表される特徴量のみを抽出したものである．離脱率に最も影響を与えた要因は同日の他授業の有無であり，ザッピングしたまま他の授業へ移ってしまうユーザーの存在が考えられる．また，午後 7 時や 8 時の授業に比べ，夜深い午後 10 時や日の高い昼の授業では離脱率が低い傾向も見られた．授業を放送する年の要因が上位・下位ともにランキング入りしていることもわかる．これらは，時間帯や期間によってスクーのユーザー層が異なることが原因であると考えられる．特に 2014 年夏から 2015 年夏までは，Schoo の公式ではなく公認団体が行った授業が乱立した時期であり，通常の MOOC では見られないような様々なジャンルの授業が放送されたため，これが上位に 2014 年や 2015 年の要因がランクインしている一因であると考えられる．

3.6　まとめ

本章では時間帯や出演者といった情報を用いて，オンライン授業における閲覧者数と離脱率の予測を行う手法を提案した．我々の手法を Schoo で行われた 2,327 本からなるデータセットに適用し，閲覧者数は $r = 0.73$，離脱率は $r = 0.54$ の相関係数で予測できることを確認した．また，これらの予測に大きな影響を与える要因を明らかにした．

MOOC で行われる授業の一部は一般ユーザーを対象としたものではなく，起業やプログラミングに関するものなど，特定の層をターゲットとした授業である．今後の課題として，これらに着目し，ユーザーの年齢やスキル別の予測が挙げられる．

第 4 章

放送前の情報のみを用いたテレビドラマの視聴率予測

4.1 はじめに

前章ではオンラインコンテンツの一種である MOOC に対して閲覧者数の予測を行ったが，用いたモデルや特徴量はオンライン特有でないものも多く，他のコンテンツの閲覧者数の予測にも適用可能であると考えられる．その応用の一つとして前章と類似した手法を用いて，本章では実世界で放送されるテレビ番組の視聴率の予測を行う．テレビメディアは長年にわたり最大の広告媒体であり，2018 年には全世界の総広告費のうち 34.1% を占めると考えられている[*1]．番組の視聴率は人気度を表す主たる指標であり，コマーシャルの広告料金を決める最大の要因は視聴率であるため，その予測は放送局と広告主の両者にとって重要な問題である．本稿では，各局が高視聴率が見込まれる時間帯であるプライムタイム（一般に午後 7 時から午後 11 時）前後に放送するテレビドラマに注目し，放送局や時間帯，出演者や脚本家，そして役者の話題性など様々な情報を用いて，ドラマの視聴率の予測を行う．すなわち，映像や音声などの情報は用いず，配役やスタッフといった企画段階に得られる情報のみを用いて視聴率を高精度で予測することが目的である．また，近年急速に普及する Twitter などのソーシャルメディアにおいて，テレビ番組の話題は日常生活でのそれと同じように人気のトピックである．本章では放送直前のソーシャルメディア上での反響を利用することで，更なる予測の改善を行う．

ドラマの視聴率に対して，それを構成する要素，例えば各役者の出演がどれほど影響す

[*1] http://www.zenithmedia.com/wp-content/uploads/2016/04/Adspend-forecasts-March-2016-executive-summary.pdf

るかは定かではない．著名な役者や脚本家を多用すれば視聴率は上がるだろうが，制作費も膨れ上がる．各局が追求するものは利益であり，それを最大化するためには各要素の視聴率への影響を明らかにする必要がある．本稿では実際に視聴率の向上に大きく寄与する要因，そして阻害する要因を特定し，それらがどの程度影響するのかの解析を行う．

4.2　関連研究

4.2.1　テレビ番組の視聴率予測

視聴率の予測を目的とした研究は多数行われてきたが，その多くは数個のテレビ局の数週間や数ヶ月といった短期間を対象としたものである [14]．我々と同様に長期間のデータを用いた研究として，Danaher ら [15] は述べ 36,000 時間のテレビ番組を対象として，Nested Logit Model を適用し，番組の放送時間やジャンル，クリスマスや年末年始といった特別な出来事を考慮した予測を行った．これは潜在視聴者を放送中の各番組に振り分けるようなモデルであり，裏番組を正確に考慮できる一方で，番組観の相対的な情報から視聴率の予測を行うため，同時間帯の主要な全番組の情報が必要となる．一方で我々のモデルは，ドラマの放送データのみから生成が可能である．また，従来のテレビ番組を対象とした研究 [47, 51] では番組の放送回は考慮されず，何年も同じ番組が続いている場合には明らかに視聴率の予測が容易であることに言及しないのに対し，我々は各ドラマの初回の視聴率，すなわちその番組の放送歴がない状態での視聴率を予測する．すなわち，局と時間帯によって定まる放送枠に対する視聴率の予測ではなく，番組に対する視聴率の予測である点が差異である．放送枠とドラマの情報は独立にモデルに組み込まれているため，例え同じ内容のドラマが異なる放送枠で放送されたとしても柔軟に対応が可能である．

4.2.2　テレビドラマや映画の視聴率予測

特にテレビドラマを対象とした研究として，本章と同様に Facebook[*1]など外部ソーシャルメディアの情報を活用したものがある [11, 28]．ただしいずれも役者など番組の制作に携わる人々に注目したものではなく，本章とは異なり前回へのソーシャルメディア上での反響を利用して次回以降の視聴率の予測を行うものである．本章と同様にソーシャルメディアとテレビ番組の関係に着目し，日本国内の番組を対象としたもの [56] もあるが，番組への反響を利用して視聴率の計測を行うものであり，放送前の予測を行うものではない．

役者が出演し，監督や脚本家といったスタッフがいるという点で，ドラマと多くの共通

[*1] https://www.facebook.com/

点を持つものが映画である．映画の興行収入を予測する研究はこれまでにいくつか行われてきたが，その中でも俳優や監督に焦点を当てたものとして，彼らの受賞歴を用いたもの [4]，過去の収益を用いたもの [50]，Twitter[*1]でのフォロワー数を用いたもの [2] などがある．我々の研究と同様に Wikipedia[2][*2]の情報を用いたものとして，各映画の Wikipedia 上の記事の編集回数と閲覧数を用いることで，興行収入の予測を行ったもの [42] がある．

4.3　データセット

4.3.1　ドラマのデータ

我々は 2008 年 4 月から 2015 年 6 月の間に放送されたドラマ 678 本からなるデータセットを作成し，実験に用いた．再放送や単発のスペシャルドラマは含まれておらず，データセット内のすべてのドラマは午後 6 時から午前 2 時の間に 3 回以上放送された連続ドラマである．すべてのドラマは公共放送である NHK，民放である日本テレビ，テレビ朝日，TBS，フジテレビ，テレビ東京の計 6 局のいずれかが関東地方で放送したものであり，予測のターゲットである視聴率は関東地方での視聴率を用いた．期間内のこの条件を満たすドラマ群は「テレビドラマデータベース」[*3]から機械的に選定したのち，目視での選定を行った．各ドラマの出演者やスタッフ，放送時間といったデータは前述のウェブサイトから機械的に収集を行ったのち，役者や放送時間などドラマに必ず存在するメタデータに漏れがある場合やその疑いがある場合，Wikipedia や各ドラマの公式サイトの情報と照合を行い，誤りのある部分については人手で修正を行った．その結果，すべてのドラマには少なくとも 3 人の役者が存在し，脚本家と監督が 1 人以上存在する．678 本中 545 本には音楽担当，507 本には主題歌担当アーティスト，326 本には原作者のデータが付与されている．

4.3.2　ソーシャルメディアのデータ

各役者の放送時の人気や話題性を考慮するため，本稿では Wikipedia と Twitter の二つのソーシャルメディアのデータを用いた．Wikipedia に関しては Page view statistics[*4]から 2007 年 12 月から 2015 年 6 月までの各記事の閲覧数を取得し，本章で扱うドラマに出

[*1] https://twitter.com/
[*2] https://ja.wikipedia.org/
[*3] http://www.tvdrama-db.com/
[*4] https://dumps.wikimedia.org/other/pagecounts-raw/

表 4.1: テレビドラマの番組表特徴量の詳細

特徴量	次元	説明
開始時刻	10	6, 7, 8, 9, 10, 11, 12 p.m. 1, 2, 3 a.m.
終了時刻	9	7, 8, 9, 10, 11, 12 p.m. 1, 2, 3 a.m.
放送時間	12	0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.2, 1.5, 1.7, 2.0 時間
放送時期	26	図 3.3 参照
曜日	7	
放送回数	15	3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15-19, 20-24, 25-
放送局	6	
放送局，曜日，時刻の組	420	6 (放送局)×7 (曜日)×10 (開始時刻)
裏番組	1	同時刻の他局でのドラマの放送の有無
Total	506	

演している役者のページのもののみ抽出して用いた．Twitter に関しては公式の API[*1]から取得した 2011 年 6 月 5 日から 2015 年 6 月 30 日までの約 1 億 2,500 万の国内ツイートのうち，役者名を含むもののみ抽出して用いた．Wikipedia の閲覧数について，女優の杏のように名前が一般名詞の場合，果物としての杏と女優としての杏のページは分かれているため，後者の閲覧数のみ用いた．

4.4　特徴ベクトルの作成

ドラマに関するあらゆる特徴を考慮するため，本章では番組表特徴量，役者特徴量，スタッフ特徴量，役者人気特徴量の 4 種の特徴量を定めた．特徴量の一部は前章で述べた Schoo の授業を表現するための特徴量に類似するものである．本節ではこれらの特徴量の詳細について述べる．

[*1] https://dev.twitter.com/streaming/

4.4.1　番組表特徴量

番組表特徴量の詳細を表 4.1 に示す．これは Schoo の番組表特徴量と同様に，放送時間や放送局などドラマの内容ではなく放送枠に関する情報のみからなる特徴量である．ただし放送時期を表す 26 次元の特徴量について，テレビドラマは一般に 1 クールと呼ばれる 1-3 月，4-6 月，7-9 月，10-12 月のいずれか 3 ヶ月間に放送されるため，図 3.3 に示す半年を一区切りとする特徴量化ではなく，3 ヶ月を一区切りとする特徴量化を行った．また，月曜夜 9 時からフジテレビで放送されるドラマをその時間帯の名を取って「月 9」と呼ぶなど，ドラマにおいては放送枠が重要な役割を果たす．これを考慮するため，放送局（6 種類）と曜日（7 種類）と放送時間（10 種類）の組み合わせである 420 次元の特徴量も番組表特徴量に含む．

4.4.2　役者特徴量

各役者を自身に対応する次元のみが 1 となる 1-of-K ベクトルで表し，各ドラマの役者特徴量は出演する役者のベクトルの和とする．役者を単語，ドラマを文書として見ると，これは配役の bag-of-words (BoW) 表現となる．我々のデータセット内には 14,768 人の役者が出現するため，これは最大で 14,768 次元の特徴量となる．Schoo の出演者特徴量と類似する特徴量であるが，Schoo の授業における先生とは異なりドラマにおいては最大 100 人を超える役者が一本のドラマに出現する．そのすべてが予測において重要であるとは限らず，実際に主要な役者のみを用いることで精度の向上が見られたため，次に述べる 2 つの閾値を設ける．一つはデータセット内で $m_a(minimum\ appearance)$ 本以上のドラマに出演した役者のみ用いる，という閾値であり，もう一つは各ドラマについて重要である $n_a(number\ of\ actors)$ 人のみ用いる，という閾値である．「テレビドラマデータベース」では出演者を含む各ドラマの情報を視聴者が掲示板で提供し，ウェブサイトの管理人がこれに精査を行い掲載する，という流れになっている．キャスト欄における掲載順を各ドラマの配役の重要度として用いた．これは原則として，ドラマのオープニングやエンディングに流れるクレジットの順である．これにより，いくつかのドラマに出演しており，かつ主要な人物を演じる役者のみが用いられることとなる．m_a は 1 から 10 までの 10 通り，n_a は 1, 2, 3, 5, 10, 20, 50, 100 の 8 通りについて，すなわち計 80 通りについて実験を行い，最も精度の良いものを採用した．

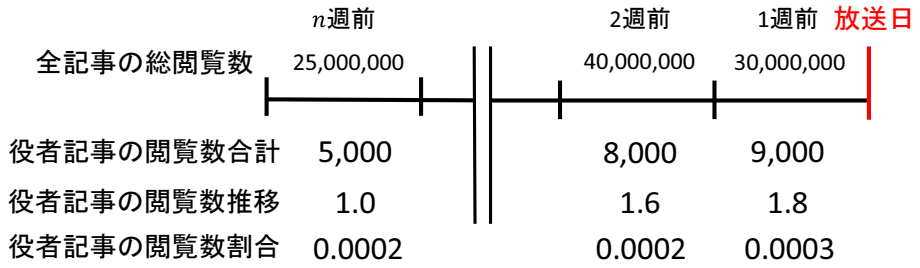


図 4.1: 各役者の Wikipedia 閲覧数からの特徴量計算

4.4.3　スタッフ特徴量

ドラマの作成に携わる役者以外のすべての人々を考慮するため，我々は役者特徴量と同じものを監督（454），脚本家（412），音楽（183），原作者（272），そして主題歌を担当するアーティスト（272）の 5 種類の人々について算出し，スタッフ特徴量として用いた．括弧内の数値は我々の学習データセット内に登場する各スタッフの人数である．各スタッフに関しても，主要なスタッフのみを考慮するため，前項で述べた閾値を用いて考慮するスタッフの絞り込みを行った．ただし，役者とは異なり一本のドラマの制作に関わる特定の職種のスタッフは高々数人のため，二つの閾値のうち m_a のみを用いた．5 種のスタッフについて，m_a を 1, 2, 3, 無限の 4 通りの計 4^5 通りについて実験を行い，最も精度の良いものを採用する．ただし m_a が無限とは，文字通り捉えれば無限本のドラマに携わったスタッフのみを用いる，すなわちそのスタッフの情報を用いないことを表す．

4.4.4　役者人気特徴量

図 4.1 に各役者の Wikipedia の記事の閲覧数から特徴量を作成する過程を示す．ドラマの初回放送日の直前の n 週間について，それに出演するある役者について，記事の各週の閲覧数，それを初回放送日の n 週前の一週間の総閲覧数で割ったもの，そして各週の閲覧数を Wikipedia の全記事の総閲覧数で割ったものの三つを特徴量として用いる．役者の人気や知名度がどれほどのものであるか，そしてそれがドラマ放送直前にどのように変化しているかを数値化することが本特徴量の目的である．各ドラマについて，重要である m 人の役者からこの特徴量を抽出し，それぞれの値そのものと対数をとったものを用いると，計 $6mn$（(3 features) × (raw, log) × (m actors) × (n weeks)）の特徴量が各ドラマについて得られる．役者人気特徴量を用いる実験では，n は 1 から 14 の 14 通り，m は 1

から 3 の 3 通りの計 42 通りについて実験を行い，最も精度の良いものを採用する．

Twitter のデータからも同様の特徴量の抽出を行う．ただし，Twitter では Wikipedia の各役者の記事の閲覧数の代わりに，各役者の名前が含まれるつぶやき（tweet）の数を用いる．

4.5　実験

4.5.1　手法

我々は視聴率予測のモデルとして，線形サポートベクター回帰（SVR）[20] とランダムフォレスト（RF）[7] の二つを用いた．いずれも説明変数の寄与度が求められる手法であり，視聴率に影響を与える要因の解析に適している．例えばある主役とヒロインのペアといった役者の組み合わせが非線形カーネルにより考慮され，その影響が大きいと判明したとしても，ドラマの制作においては続編でない限り，既存のものと主役級の役者が被らないよう考慮されるのが一般的である．精度の検証には前章で用いたのと同様の時系列を考慮した交差検定を用いた．本章では，モデルを学習するための最小データ数である k について，$k = 67$ とした．これはデータセット内の 2008 年の間に放送されたドラマ数である．通常の 1 クール 3 ヶ月間に渡り放送されるドラマと，NHK で放送される大河ドラマのような通年のドラマの両者を揃えるため，年の変わり目を境目とした．本稿では各ドラマの初回視聴率を予測の対象とする．第 2 話以降の視聴率は第 1 話の面白さや評判に大きく依存すると考えられること，そして第 1 話の視聴率は与えられているという仮定のもとで実験を行っている先行研究 [11] において第 2 話以降の視聴率は各回に対するソーシャルメディア上の反響と，前回の視聴率から高い精度で予測を行えているためである．

4.5.2　各特徴量単体での予測結果

前節で述べた 4 種の特徴量を単体で用いた視聴率の予測結果を図 4.2 に示す．特に役者特徴量のみ，スタッフ特徴量のみを用いた場合において，RF による予測が良い結果を示した．放送枠がまだ決まっていない段階で，配役や監督を選定する必要がある場合，RF を用いた予測が適していることがわかる．なお，役者人気特徴量について，ドラマのデータセットが 2008 年からなのに対して Twitter のデータは 2011 年以降のもののため，すべてのドラマを用いた実験においては Wikipedia から抽出された特徴量のみを用いた．

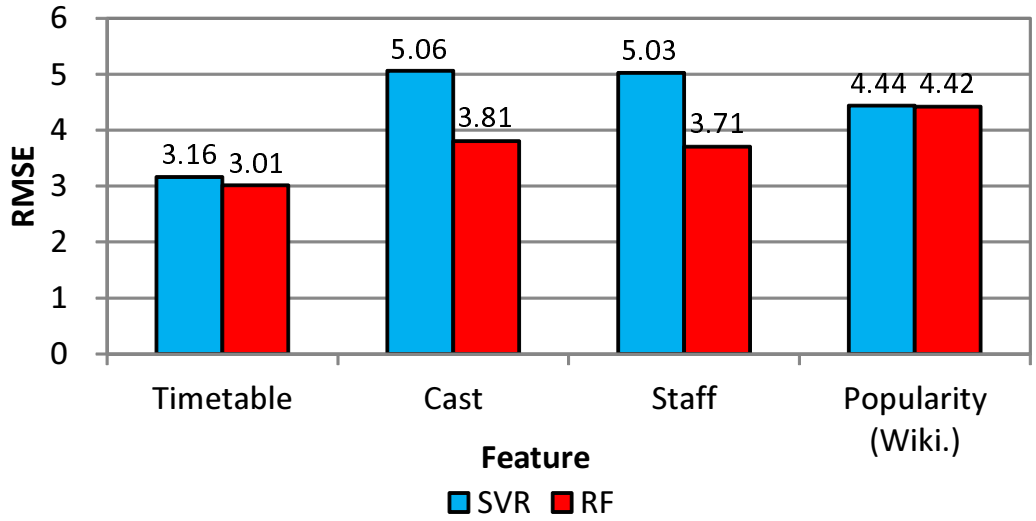


図 4.2: 実際の視聴率と予測視聴率との間の RMSE

表 4.2: 番組表特徴量と役者特徴量を用いた時の実際の視聴率と予測視聴率の間の RMSE

$ma \backslash na$	1	2	3	5	10	20	50	100
1	3.10	3.10	3.09	3.15	3.17	3.19	3.21	3.23
2	3.05	3.06	3.08	3.13	3.17	3.21	3.20	3.21
3	3.03	3.07	3.09	3.07	3.13	3.17	3.19	3.19
4	3.05	3.04	3.07	3.06	3.08	3.16	3.18	3.20
5	3.04	3.01	3.01	3.04	3.07	3.11	3.17	3.18
6	3.01	3.03	**3.00**	3.04	3.09	3.07	3.20	3.19
7	3.01	3.05	3.04	3.01	3.10	3.07	3.17	3.18
8	3.02	3.04	3.07	3.04	3.10	3.09	3.15	3.19
9	3.14	3.16	3.18	3.14	3.16	3.15	3.16	3.17
10	3.14	3.15	3.17	3.15	3.15	3.15	3.17	3.21
10	3.14	3.15	3.17	3.15	3.15	3.15	3.17	3.21

4.5.3　特徴量を組み合わせた際の予測結果

役者特徴量のパラメータである m_a と n_a を変化させたときの結果を表 4.2 に示す．$m_a = 6, n_a = 3$ のとき，すなわちデータセット内の 6 本以上のドラマに出演しており，か

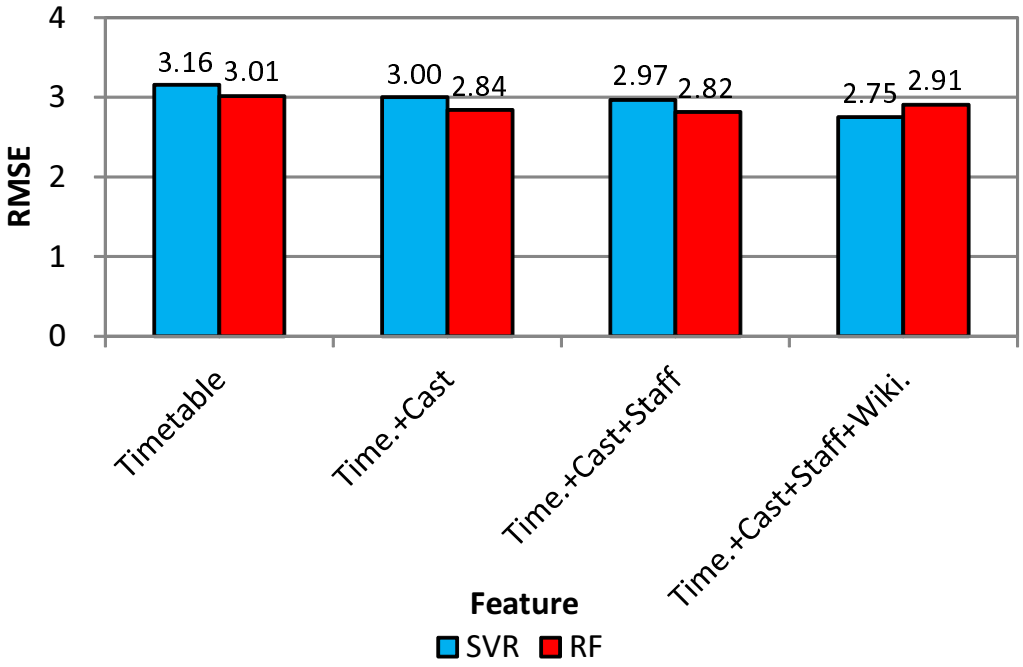


図 4.3: 番組表特徴量に他の特徴量を付した際の結果

つ各ドラマの重要な 3 人のみを考慮した時が最も良い精度を示した．このとき，データセット内の 111 人の役者が残った．以後，これを役者特徴量とする．同様にスタッフ特徴量についても，番組表特徴量と役者特徴量と組み合わせたときに最も精度の良い ma を調べたところ，脚本家とアーティストについては $ma = 3$，音楽については $ma = 2$，監督と原作者については ma は無限となった．同様に役者人気特徴量のパラメータについて，すべての特徴量を結合したときの実験から，m と n のどちらも 3 のときに最も良い精度を示した．番組表特徴量に役者，スタッフ，役者人気特徴量を付け加えた際の結果を図 4.3 に示す．Twitter のデータも用いた場合の結果として，同様の実験を 2012 年以降の 346 本のドラマのみを用いて行った場合の結果を図 4.4 に示す．このとき，時系列を考慮した交差検定のパラメータである $k = 73$ とした．これは 2012 年 1 月から 2012 年 9 月までに放送されたドラマの数である．いずれの場合も，番組表，役者，スタッフ特徴量までを用いた場合には RF が良い精度を，役者人気特徴量も含めた場合には SVR が良い精度を示した．

4.5.4　視聴率への各要素の影響

データセット内のすべてのドラマを使った実験について，すべての特徴量を用いた場合の SVR モデルにおいて予測への影響の大きかった要因の上位・下位 30 個を表 4.3，4.4 に

表 4.3: 視聴率に影響を与えた上位 30 要因

特徴量	重み
放送回数 25 回以上	+2.50
木村拓哉（俳優）	+2.45
福田靖（脚本家）	+2.45
午後 8 時開始	+1.85
午後 10 時終了	+1.73
米倉涼子（女優）	+1.68
NHK，日曜，午後 8 時開始	+1.67
フジテレビ，月曜，午後 9 時開始	+1.66
綾瀬はるか（女優）	+1.63
大島ミチル（音楽）	+1.62
阿部寛（俳優）	+1.56
北大路欣也（俳優）	+1.50
Superfly（アーティスト）	+1.44
DREAMS COME TRUE（アーティスト）	+1.39
放送時間 0.9 時間	+1.36
黒木メイサ（女優）	+1.30
森下佳子（脚本家）	+1.29
TBS，金曜，午後 10 時開始	+1.29
2010 年 6 月から 2011 年 5 月の間に放送	+1.27
午後 9 時開始	+1.27
徳永富彦（脚本家）	+1.24
吉俣良（音楽）	+1.23
櫻井武晴（脚本家）	+1.17
杏（女優）	+1.15
日本テレビ，土曜，午後 9 時開始	+1.14
橋田壽賀子（脚本家）	+1.13
テレビ朝日で放送	+1.11
テレビ朝日，金曜，午後 11 次開始	+1.11
服部隆之（音楽）	+1.09
TBS，日曜，午後 9 時開始	+1.07

表 4.4: 視聴率に影響を与えた下位 30 要因

特徴量	重み
フジテレビ，日曜，午後 9 時開始	-1.55
NHK，土曜，午後 9 時開始	-1.45
午前 0 時開始	-1.20
午前 1 時終了	-1.15
放送回数 4 回	-1.09
A（脚本家）	-1.09
午前 2 時終了	-1.07
テレビ朝日，日曜，午後 11 時開始	-1.01
午前 1 時開始	-0.99
放送時間 0.5 時間	-0.96
H（女優）	-0.95
A（脚本家）	-0.94
フジテレビ，金曜，午後 8 時開始	-0.89
テレビ東京で放送	-0.85
B（アーティスト）	-0.84
M（アーティスト）	-0.80
放送時間 0.3 時間	-0.78
日本テレビ，水曜，午前 1 時開始	-0.78
午前 3 時終了	-0.78
K（アーティスト）	-0.75
T（俳優）	-0.73
NHK，金曜，午後 10 時開始	-0.70
K（女優）	-0.68
C（音楽）	-0.68
N（音楽）	-0.67
M（女優）	-0.67
N（脚本家）	-0.66
午後 7 時開始	-0.64
H（音楽）	-0.63
K（脚本家）	-0.63

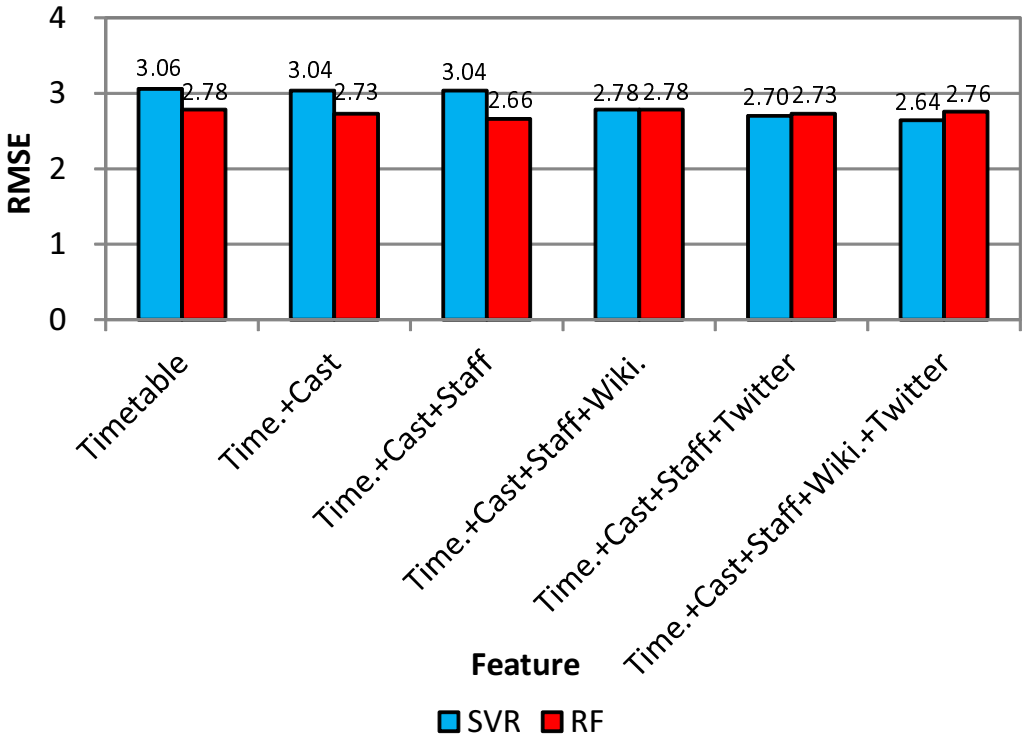


図 4.4: 2012 年以降のドラマを用いた場合の結果

示す．本章では線形 SVR を用いたため，この数値はそのまま予測視聴率への寄与となる．例えば，表 4.3 の 6 行目は，米倉涼子の出演したドラマの予測視聴率は一律で 1.68% 加算されることを示す．表 4.3 から，一部の著名な俳優は視聴率に 1% を超えるほどの影響を与えていることがわかる．また，フジテレビの月 9 枠や，NHK の大河ドラマ枠など，特定のドラマ枠を表す要素はどちらの表にも頻出しており，ドラマの内容に関わらず，多くの人々は特定の曜日の特定の時間に，ドラマの内容によらず定期的にある局を視聴する傾向が見られる．

一方で，特定のドラマに対して過学習が行われた形跡も見られる．表 4.3 では視聴率に最も影響を与える要因として放送回数 25 回以上とあるが，これは国内ドラマにおいてはほぼ大河ドラマに限られている．これは大河ドラマである，という事実が与える影響が 2.5% であることを示唆しており，他のドラマが単に放送回数を増すだけで視聴率が増すとは言えない．また，今回の手法では役者人気特徴量により役者の人気の時系列の変化を考慮している一方で，役者特徴量は時系列の変化を含まないため，どの時点での潜在視聴率かが不明瞭であるという問題がある．

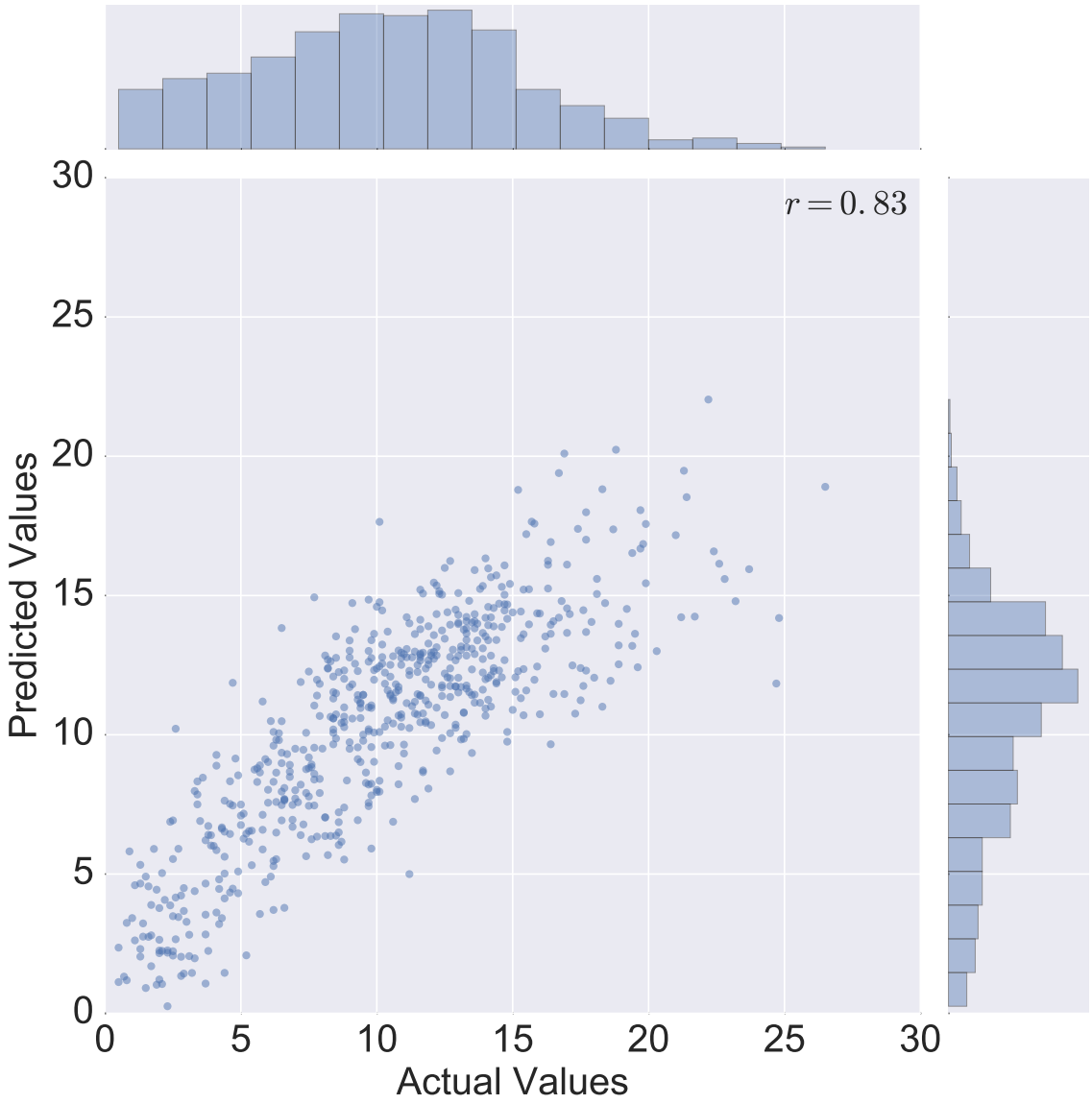
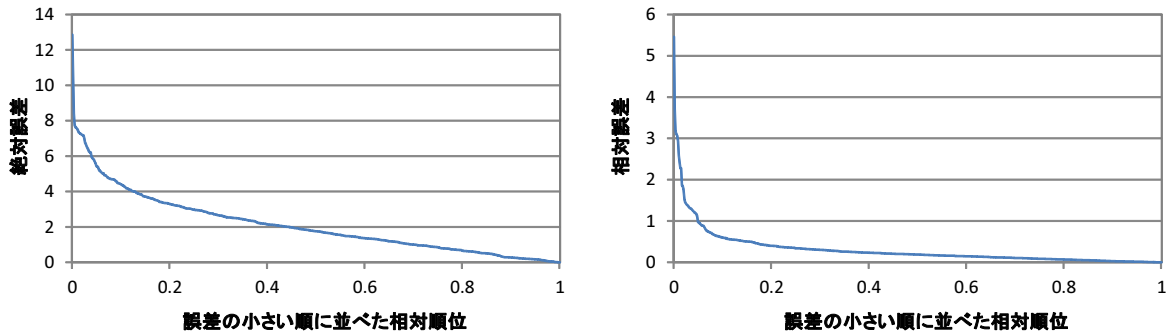


図 4.5: 実際の初回視聴率（横軸）と予測初回視聴率（縦軸）の散布図



(a) 絶対誤差順に並べた順位（横軸）と絶対誤差（縦軸）

(b) 相対誤差順に並べた順位（横軸）と絶対誤差を実際の視聴率で割った相対誤差（縦軸）

図 4.6: 図 4.5 の誤差の傾向

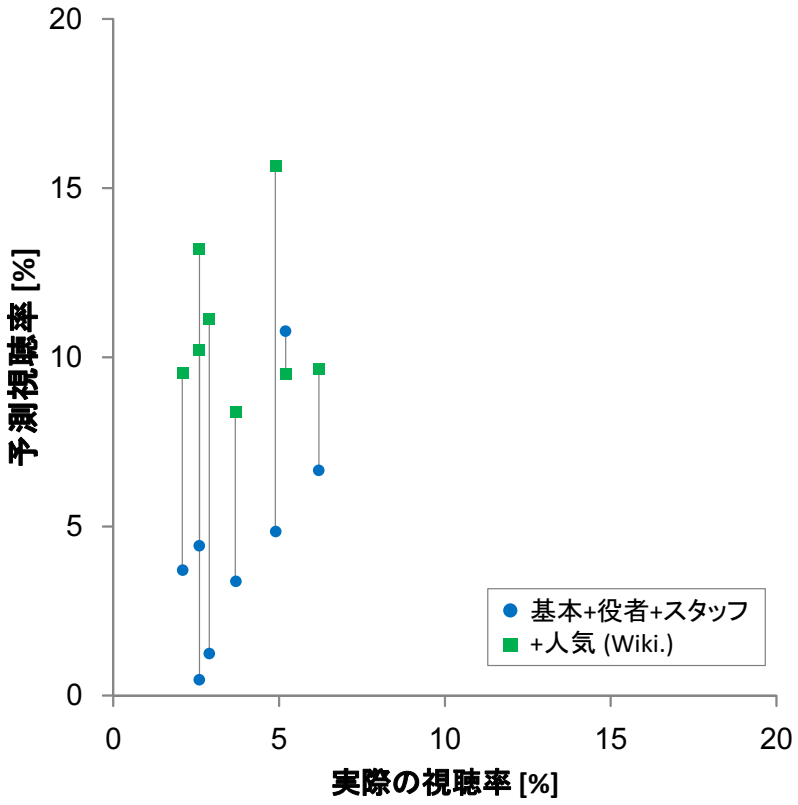


図 4.7: AKB48 グループが主役か準主役のドラマの特徴量による予測の差

4.5.5　結果の考察

表 4.3，4.4 で参照したモデルにおける，実際の視聴率と予測視聴率との散布図を図 4.5 に示す．このときの両者の相関係数は 0.832，平均誤差は 2.13% であった．予測の誤差の傾向を図 4.6 に示す．全体の半数以上のドラマについて，2% 以内の誤差で予測が行えていることがわかる．概ね予測がうまく行えている一方で，外れ値としていくつかのドラマが点在していることもわかる．その顕著な例として，データセット中で初出の大河ドラマである 2009 年「天地人」は 12.8%，翌年の 2010 年「龍馬伝」は 8.4% の誤差で予測に失敗した．いずれも実際の視聴率よりも低く視聴率を予測してしまうことが原因であり，日本人にとって大河ドラマは，出演する役者やスタッフ，あるいは内容に加えて，大河ドラマであるということ自体が視聴者を惹きつける力を持っていると考えられる．一方で，2013 年「八重の桜」では 2.8%，2014 年「軍師官兵衛」では 6.3%，2015 年「花燃ゆ」では 2.4% の誤差であり，時代が下るにつれて予測誤差は減少する傾向にある．大河ドラマは毎年固定

の放送枠を持つため，この放送枠に結びつけたモデルを作成できたと考えられる．

他に実際の視聴率に対して低く視聴率を予測したものとして，過去のドラマの続編のものが目立った．実際の視聴率に比べ，「HERO2」では 7.6%，「リーガル・ハイ 2」では 7.0%，「ショムニ 2013」では 7.3%，「DOCTORS2」では 7.2% も視聴率が低く予測された．続編では大半の役者やスタッフは同じであるため，我々のモデルでは前作の視聴率に準じた値が予測される．一方で，続編が制作されるドラマは前作で大きな成功を収めたものが多く，その期待から第 2 シーズンの初回視聴率は前情報の一切ないドラマに比べて高い傾向がある．続編の視聴率をよく予測するためには前作の人気や反響を考慮する必要があると考えられる．

また，渡辺謙主演の「負けて，勝つ」は実際の視聴率 11.2% に対して，予測視聴率が 5.0% と大きく外れていた．人気特徴量を用いず，基本 + 役者 + スタッフ特徴量のみを用いた場合は，さらに低い 3.3% という予測結果であった．渡辺謙はハリウッド映画でも活躍する指折りの俳優であるが，国内の連続ドラマに出演することはほとんどなく，今回データとして用いた 2008 年以降の連続ドラマでは「負けて，勝つ」が唯一である．我々のモデルでは過去のドラマからの予測が主のため，渡辺謙は他に出演作のない無名俳優と同様に扱われてしまい，予測を大きく外したと考えられる．人気特徴量を用いることで本作の予測視聴率は高まったものの，より良い予測のためには映画や舞台など，他の分野での活躍も考慮する必要がある．

一方，予測された視聴率よりも実際の視聴率が低いものとしては，裏番組が強いものが挙げられる．22.8% の視聴率を獲得した「ドクター X2」の裏である「夫のカノジョ」は 11.9% と予測したものの実際は 4.7%，16.0% の視聴率であった「獣医ドリトル」の裏である「パーフェクト・リポート」は 14.9% と予測したものの実際は 7.7% の視聴率に留まった．

他に視聴率の予測が外れたドラマの多くは，AKB48 やジャニーズを筆頭としたアイドルが主演を務めたものが占めていた．前田敦子主演の「花ざかりの君たちへ 2011」は 17.6% と予測したものの実際は 10.1%，錦戸亮主演の「ごめんね青春！」は 14.8% と予測したものの実際は 10.1% であった．アイドルは役者業の他にも，ライブやバラエティ番組への出演など，役者を生業とする人々に比べメディアへの露出の機会は多い．アイドルの数ある仕事のうち，いずれかに興味を持った人は Wikipedia の記事を閲覧する可能性があるため，専業の俳優とは閲覧の傾向が異なることが理由として考えられる．実際に，例えば指原莉乃主演の「ミューズの鏡」では，実際の視聴率 2.6% に対し，基本 + 役者 + スタッフ特徴量で予測した場合は 5.6%，そこに人気特徴量を加えた場合は 10.2% という予測結果であった．図 4.7 にアイドルの一例として，AKB48 グループのメンバーが主役か

準主役を演じたデータセット内の 8 本のドラマに対して，スタッフ特徴量までを用いた場合と役者人気特徴量を加えた場合との予測結果を示す．線で結ばれたドラマは同一のドラマであることを表す．1 本を除き，他のすべては役者人気特徴を考慮することで予測誤差が大きくなっていることがわかる．この 8 本については，基本 + 役者 + スタッフ特徴量を用いた場合の MSE が 2.36 に対し，役者人気特徴量を併せて用いた場合の MSE は 7.60 であり，Wikipedia の閲覧数の考慮が負の影響を及ぼしていることがわかる．

4.6　まとめ

本章では役者やスタッフなどテレビドラマに関連する様々な情報を用いて，ドラマの初回視聴率を予測する手法を提案した．我々の作成した 678 本の国内ドラマからなるデータセットに手法を適用し，0.832 の相関係数で視聴率の予測を行えることを示した．

さらに予測モデルからの視聴率に影響を与える要因の解明も行い，特定の役者は視聴率に数パーセントもの影響を与えうることや，番組の内容に関わらず特定の時間帯に特定の局を視聴する傾向を明らかにした．

今後の課題として，全体の視聴率ではなく，性別や年令ごとの視聴率の予測が挙げられる．化粧品をはじめとする多くの商品は特定の視聴者層をターゲットとしたものであり，自社が広告を出している番組が，どの年代の，どちらの性別の人々に見られているかは，特定の層をターゲットとしたい広告主にとって重要な問題となる．

第 5 章

まとめと展望

5.1　本研究の成果

本研究の目的は視聴者に向けて話を行うプレゼンテーション動画の解析の一環として，以下の 2 点を行うことであった．

- 視聴者の抱いた印象の予測
- 閲覧者数と離脱率の予測

1 つ目については TED の 1,646 本のプレゼンテーション動画群を用いて，各印象の投稿率が上位・下位 10% の 2 クラス分類を 14 印象の平均で 91.1% の精度で行えることを示した．また，本手法を実装した，任意のプレゼンテーションについて視聴者が抱くであろう印象の予測が行えるツールを作成した．2 つ目については Schoo の授業群を用いて，閲覧者数については $r = 0.73$，離脱率については $r = 0.54$ の相関係数で予測できることを示した．後者の手法についてオンライン動画ではない映像コンテンツであるテレビドラマに対して適用し，視聴率の予測を $r = 0.83$ の相関係数で行えることを示した．また，すべての予測においてターゲットとなる値の予測に影響の大きな要因の解析を行った．

5.2　今後の展望

今後の展望として，プレゼンテーションへの印象の予測について，予測に留まらずユーザーのプレゼンテーションの改善に向けた支援を行うシステムの開発が挙げられる．また，本稿では TED のプレゼンテーション群を用いたが，TED では科学や社会，アートなど内容に関しては多種多様なプレゼンテーションが存在する一方で，形式に関しては学会発表やビジネスの場でのプレゼンテーションとは異なる．様々な形式のプレゼンテーショ

ンに対する解析を行うことも今後の課題であるといえる．

閲覧者数の予測について，本稿では 2 つのデータセットについていずれも高い精度で予測が行えることを示した．一方，現在オンラインコンテンツを配信するサービスは激動の時代であり，本稿で用いたテレビドラマをはじめとするテレビ番組についても，各キー局が一定期間ウェブ上で見逃した番組の視聴を行えるサービスを開始している．4 章で舞台や映画など，テレビドラマ以外の情報を用いることが精度を向上させる案の一つであると述べたが，時代とともに変わりゆく配信形式の変化をモデル内に組み込むことは今後の課題の一つである．MOOC の授業については，本稿では各授業に対する予測を行ったが，関連研究で述べたようにコース途中での脱落者の予測など，授業間をまたがる情報を対象する研究も存在する．それらの手法と組み合わせることで更なる精度の向上や，より広い視野での MOOC に関する問題を見つけることも今後の課題である．

参考文献

[1] A. Abisheva, V. R. K. Garimella, D. Garcia, and I. Weber. Who watches (and shares) what on youtube? and when?: using twitter to understand youtube viewership. In *WSDM*, pp. 593–602, 2014.

[2] K. R. Apala, M. Jose, S. Motnam, and C. C. Chan. Prediction of movies box office performance using social media. In *ASONAM*, pp. 1209–1214, 2013.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, pp. 993–1022, 2003.

[4] P. Boccardelli, F. Brunetta, and F. Vicentini. What is critical to success in the movie industry? a study on key success factors in the italian motion picture industry. *Dynamics of Institutions and Markets in Europe*, No. 46.

[5] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov. Enriching word vectors with subword information. In *arXiv*, 2016.

[6] R. B. Bradford. An empirical study of required dimensionality for large-scale latent semantic indexing applications. In *CIKM '08*, pp. 153–162, 2008.

[7] L. Breiman. Random forests. *Machine learning*, Vol. 45, No. 1, pp. 5–32, 2001.

[8] C. C. Chang and C. J. Lin. Libsvm: A library for support vector machines, 2011.

[9] H. Chen, M. Sun, C. Tu, Y. Lin, and Z. Liu. Neural sentiment classification with user and product attention. In *EMNLP*, pp. 1650–1659, 2016.

[10] L. Chen, C. W. Leong, G. Feng, and C. M. Lee. Using multimodal cues to analyze mla'14 oral presetations quality corpus. In *MLA*, pp. 45–52, 2014.

[11] Y. H. Cheng, C. M. Wu, T. Ku, and G. D. Chen. A predicting model of tv audience rating based on the facebook. In *SocialCom*, pp. 1034–1037, 2013.

[12] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, Vol. 20, No. 3, pp. 273–297, 1995.

[13] A. M. Dai and Q. V. Le. Semi-supervised sequence learning. In *NIPS*, pp. 3079–

3087, 2015.

[14] P. J. Danaher, T. S. Dagger, and M. S. Smith. Forecasting television ratings. *International Journal of Forecasting*, Vol. 27, No. 4, pp. 1215–1240, 2011.

[15] P. Danaher and T. Dagger. Using a nested logit model to forecast television ratings. *International Journal of Forecasting*, Vol. 28, No. 3, pp. 607–622, 2012.

[16] S. C. Deerwester, S. T. Dumals, G. W. Furnas, T. K. Landauer, and R. A. Harshman. Indexing by latent semantic analysis. *JASIS*, Vol. 41, pp. 391–407, 1990.

[17] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, pp. 248–255, 2009.

[18] I. S. Dhillon and D. S. Modha. Concept decompositions for large sparse text data using clustering. *Journal of Machine Learning*, Vol. 42, pp. 143–175, 2001.

[19] V. Echeverría, A. Avenda no, K. Chiluiza, A. Vásquez, and X. Ochoa. Presentation skills estimation based on video and kinect data analysis. In *ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, pp. 53–60, 2014.

[20] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin. Liblinear: A library for large linear classification. *Journal of Machine Learning Research*, pp. 1871–1874, 2008.

[21] T. Gan, Y. Wong, B. Mandal, V. Chandarasekhar, and M. S. Kankanhalli. Multi-sensor self-quantification of presentations. In *ACMMM*, pp. 601–610, 2015.

[22] T. Gan, Y. Wong, B. Mandal, V. Chandrasekhar, and M. S. Kankanhalli. Multi-sensor self-quantification of presentations. In *ACMMM*, pp. 601–610, 2015.

[23] F. Gelli, T. Uricchio, M. Bertini, A. Del Bimbo, and S. F. Chang. Image popularity prediction in social media using sentiment and context features. In *ACMMM*, pp. 907–910, 2015.

[24] Z. S. Harris. Distributional structure. *Word*, Vol. 10, pp. 146–162, 1954.

[25] J. D. Hart. Automated kernel smoothing of dependent data by using time series cross-validation. *Journal of the Royal Statistical Society*, pp. 529–542, 1994.

[26] W. He. Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, Vol. 29, No. 1, pp. 90–102, 2013.

[27] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingbury. Deep neural networks

for acoustic modeling in speech recognition. *Signal Processing Magazine*, Vol. 29, No. 6, pp. 82–97, 2012.

[28] Y. Y. Huang, Y. A. Yen, T. W. Ku, S. D. Lin, W. T. Hsieh, and T. Ku. A weight-sharing gaussian process model using web-based information for audience rating prediction. *Technologies and Applications of Artificial Intelligence*, pp. 198–208, 2014.

[29] R. Johnson and T. Zhang. Supervised and semi-supervised text categorization using lstm for region embeddings. In *ICML*, pp. 526–534, 2016.

[30] J. Kim, P. J. Guo, D. T. Seaton, P. Mitros, K. Z. Gajos, and R. C. Miller. Understanding in-video dropouts and interaction peaks inonline lecture videos. In *L@S*, pp. 31–40, 2014.

[31] Y. Kim and J. Thayne. Effects of learner-instructor relationship-building strategies in online video instruction. *Distance Education*, Vol. 36, No. 1, pp. 100–114, 2015.

[32] J. P. Kincaid, R. P. Fishburn, R. L. Roers, and B. S. Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, CNTECHTRA Research Branch, 1975.

[33] M. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart. Predicting mooc dropout over weeks using machine learning methods. In *EMNLP Workshop on Analysis of Large Scale Social Interaction in MOOCs*, pp. 60–65, 2014.

[34] K. Kurihara, M. Goto, J. Ogata, Y. Matsusaka, and T. Igarashi. Presentation sensei: A presentation training system using speech and image processing. In *ICMI '07*, pp. 358–365, 2007.

[35] J. G. Lee, S. Moon, and K. Salamatian. An approach to model and predict the popularity of online contents with explanatory factors. In *WI-IAT*, pp. 623–630, 2010.

[36] H. Li, X. Ma, F. Wang, J. Liu, and K. Xu. On popularity prediction of videos shared in online social networks. In *CIKM*, pp. 169–178, 2013.

[37] T. Li and M. Ogihara. Toward intelligent music information retrieval. *IEEE TMM*, Vol. 8, No. 3, pp. 564–574, 2006.

[38] W. Li, M. Gao, H. Li, Q. Xiong, J. Wen, and Z. Wu. Dropout prediction in moocs using behavior features and multi-view semi-supervised learning. In *IJCNN*, pp.

3130–3137, 2016.

[39] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang. Rapid: Rating pictorial aesthetics using deep learning. In *ACMMM*, pp. 457–466, 2014.

[40] G. Luzardo, B. Guamán, K. Chiluiza, J. Castells, and X. Ochoa. Estimation of presentations skills based on slides and audio features. In *MLA*, pp. 37–44, 2014.

[41] G. H. McLaughlin. Smog grading: A new readability formula. *Journal of reading*, Vol. 12, pp. 639–646, 1969.

[42] M. Mestyán, T. Yasseri, and J. Kertész. Early prediction of movie box office success based on wikipedia activity big data. *PloS one*, Vol. 8, No. 8, 2013.

[43] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Adv. NIPS*, pp. 3111–3119, 2013.

[44] A. T. Nguyen, W. Chen, and M. Rauterberg. Online feedback system for public speakers. In *IEEE Symposium on IS3e*, pp. 1–5, 2012.

[45] D. F. Onah, J. Sinclair, and R. Boyatt. Dropout rates of massive open online courses: behavioural patterns. In *EDULEARN14 Proceedings*, pp. 5825–5834, 2014.

[46] O. Ozan and Y. Ozarslan. Video lecture watching behaviors of learners in online courses. *Educational Medial International*, Vol. 53, No. 1, pp. 27–41, 2016.

[47] R. Pagano, M. Quadrana, P. Cremonesi, S. Bittanti, S. Formwentin, and A. Mosconi. Prediction of tv ratings with dynamic models. In *ACM Workshop on Reccomendation Systems for Television and Online Videos*, 2015.

[48] B. Pang and L. Lee. pinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, Vol. 2, pp. 1–135, 2008.

[49] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: Sentiment classification using machine learning techniques. *ACL-02*, Vol. 10, pp. 79–86, 2002.

[50] R. Parimi and D. Caragea. Pre-release box-office success prediction for motion pictures. In *MLDM*, pp. 571–585, 2013.

[51] A. Patelis, K. Metaxiotis, K. Nikolopoulos, and V. Assimakopoulos. Fortv: decision support system for forecasting television viewership. *Journal of Computer Information Systems*, Vol. 43, No. 4, 2003.

[52] H. Pinto, J. M. Almeida, and M. A. Gonçalves. Using early view patterns to predict the popularity of youtube videos. In *WSDM*, pp. 365–374, 2013.

[53] R. J. Senter and E. A. Smith. Automated readability index. Technical report, AMRL-TR-6620, 1967.

[54] M. Soleymani, M. Pantic, and T. Pun. Multimodal emotion recognition in response to videos. *IEEE TAC*, Vol. 3, pp. 211–223, 2012.

[55] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, and D. Anguelov. Going deeper with convolutions. In *CVPR*, pp. 1–9, 2014.

[56] S. Wakamiya, L. E. E. Ryong, and K. Sumiya. Twitter-based tv audience behavior estimation for better tv ratings. In *DEIM Forum*, 2011.

[57] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. CVPR, 2010.

[58] Z. Wang, W. Zhu, X. Chen, L. Sun, J. Liu, and M. Chen. Propagation-based social-aware multimedia content distribution. *TOMM*, Vol. 9, No. 1, 2013.

[59] F. Weininger, P. Staudt, and B. Schuller. Words that fascinate the listener: Predicting affective ratings of on-line lectures. *IJDET*, Vol. 11, No. 2, pp. 110–123, 2013.

[60] K. Yamaguchi, T. L. Berg, and L. E. Ortiz. Chic or social: Visual popularity analysis in online fashion networks. In *ACMMM*, pp. 773–776, 2014.

[61] T. Yamasaki, J. Hu, K. Aizawa, and T. Mei. The power of tags: Predicting the popularity of your content in social media in geo-spatial and temporal content. In *PCM*, pp. 149–158, 2015.

[62] D. Yang, T. Shiha, D. adamson, and C. P. Rosé. "turn on, tune in, drop out: Anticipating studen dropouts in massive open online courses. In *NIPS Data-driven education workshop*, 2013.

[63] H. Yu, L. Xie, and S. Sanner. Twitter-driven youtube views: Beyond individual influencers. In *ACMMM*, pp. 869–872, 2014.

[64] D. Zhang, H. Xu, Z. Su, and Y. Xu. Chinese comments sentiment classification based on word2vec and svm perf. *Expert Systems with Applications*, Vol. 42, No. 4, pp. 1857–1863, 2015.

[65] R. Zhou, S. Khemmarat, L. Gao, J. Wan, and J. Zhang. How youtube videos are discovered and its impact on video views. *Multimedia Tools and Applications*, Vol. 75, No. 10, pp. 6035–6058, 2016.

発表文献

国内論文誌

[1] 福島悠介，山崎俊彦，相澤清晴. 文書と音声解析に基づくプレゼンテーション動画の印象予測. 電子情報通信学会論文誌，Vol. J99-D，pp. 699–708，2016.

[2] 福島悠介，山崎俊彦，相澤清晴. 放送前の情報のみを用いたテレビドラマの視聴率予測. 映像情報メディア学会誌，Vol. 70，No. 11，pp. 255–261，2016.

国際会議

[3] T. Yamasaki, Y. Fukushima, R. Furuta, L. Sun, K. Aizawa, and D. Bollegala. Prediction of user ratings of oral presentations using label relations. In *ACMMM-ASM*, pp. 33–38, 2015.

[4] Y. Fukushima, T. Yamasaki, and K. Aizawa. Audience ratings prediction of TV dramas based on the cast and their popularity. In *IEEE BigMM*, pp. 279–286, 2016.

[5] T. Yamasaki, Y. Fukushima, R. Furuta, and K. Aizawa. Towards online impressions prediction of oral presentations using soft coding. In *1st workshop on Attractiveness Computing in Multimedia in conjuction with BigMM*, pp. 462–465, 2016.

[6] R. Furuta, Y. Fukushima, T. Yamasaki, and K. Aizawa. Multi-label classification using class relations based on higher-order mrf optimization. In *BigVision Workshop in conjuction with CVPR*,2016.

国内会議・研究会

[7] 福島悠介, 山崎俊彦, 相澤清晴. 大量の事例に基づくプレゼンテーションの分類・解析. 映像情報メディア学会ヒューマンインフォメーション研究会 (HI), Vol. 38, No. 46, 2014.

[8] 福島悠介, 山崎俊彦, 相澤清晴. 文書解析に基づくプレゼンテーション動画の分類と印象判定. 映像情報メディア学会冬季大会, 11-1, 2014.

[9] 福島悠介, 山崎俊彦, 相澤清晴. ソフトクラスタリングを用いたプレゼンテーションの印象推定精度の改善. 画像の認識・理解シンポジウム (MIRU), 2015.

[10] 山崎俊彦, 福島悠介, 徐建鋒, 酒澤茂之. 文書特徴と音声特徴を用いたプレゼンテーションの印象推定. 電子情報通信学会マルチメディア・仮想環境基礎研究会 (MVE), pp. 119–122, 2015.

[11] R. Furuta, Y. Fukushima, T. Yamasaki, and K. Aizawa. MRF-based multi-label classification using label relations. *IEICE technical report*, 115(224):83-90, 2015.

[12] 古田諒佑, 福島悠介, 山崎俊彦, 相澤清晴. マルコフ確率場に基づくラベルの関係性を考慮したマルチラベル分類. 情報処理学会コンピュータビジョンとイメージメディア研究会, 2015.

[13] 古田諒佑, 福島悠介, 山崎俊彦, 相澤清晴. 高階エネルギーの MRF 最適化によるラベル共起を考慮したマルチラベル分類. 映像情報メディア処理シンポジウム (IMPS2015), I-4-13, 2015.

[14] 山崎俊彦, 福島悠介, 古田諒佑, 相澤清晴. 文書・音声特徴によるリアルタイム・プレゼンテーション解析に向けた検討. 画像符号化・映像情報メディア処理シンポジウム (PCSJ・IMPS2015), pp. 82–83, 2015.

[15] 福島悠介, 山崎俊彦, 相澤清晴. 番組除法と出演者の話題性を考慮したテレビドラマの視聴率予測. 映像情報メディア学会冬季大会, 14C-5, 2015.

[16] 古田諒佑, 福島悠介, 山崎俊彦, 相澤清晴. 高階 MRF によるクラス間の共起情報を用いたマルチラベル分類の精度改善. 電子情報通信学会画像工学研究会, 2016.

[17] Y. Fukushima, T. Yamasaki, and K. Aizawa. Predicting Ratings of TV Dramas Using Social Media Activities. 画像符号化･映像メディア処理シンポジウム (MIRU), OS2-01, 2016.

[18] 高田裕樹, 福島悠介, 山崎俊彦, 相澤清晴, 森健志郎, 鈴木顕照. 協調フィルタリングとアイテム間類似度に基づくオンライン授業の推薦技術の検討. 映像情報メディア学会冬季大会, 13C-2, 2016.

[19] 福島悠介, 山崎俊彦, 相澤清晴, 森健志郎, 鈴木顕照. オンライン動画学習サービスにおける閲覧数・離脱率の推定. ヒューマンインフォメーション研究会 (HI), 2017. (発表予定)

受賞

[20] 山崎俊彦, 福島悠介, 古田諒佑, 相澤清晴. IMPS2015 優秀論文フロンティア賞, 2015.

[21] 山崎俊彦, 福島悠介, 徐建鋒, 酒澤茂之. MVE 賞, 2015.

[22] 古田諒佑, 福島悠介, 山崎俊彦, 相澤清晴. 画像工学研究会 IE 賞, 2016.

[23] 福島悠介. MIRU 学生奨励賞, 2016.

[24] 福島悠介. 第 32 回電気通信普及財団賞 テレコムシステム技術学生賞 佳作, 2017. (受賞予定)

解説記事

[25] 古田諒佑, 福島悠介. 株式会社スクー訪問レポート. 映像情報メディア学会誌, Vol. 70, No. 3, pp. 498–499, 2016.

謝辞

本研究を進める上で，そして相澤山崎研究室で 3 年間を過ごす上で，たくさんの方にお世話になりました．

指導教員である山崎俊彦准教授には，研究の “け” の字も知らなかった学部 4 年の頃からこれまで，テーマ決めから外部発表に至るまでのすべてのプロセスにおいて手厚く面倒を見ていただきました．本稿の 3 つのテーマをアイデアで終わらせず形に出来たのは他ならぬ山崎先生のおかげです．就職後も，私が研究に関わる部門に配属されたならば，どこかでまた携わる機会があるかもしれません．その時にはよろしくお願いします．

相澤清晴教授には，ミーティングで毎回貴重な意見を頂きました．また，発表予行においては鋭い質問や指摘も多く，外部で発表を行う本番には自信を持って挑むことができました．ありがとうございました．

学術支援職員の松林真幸さんには，研究室での生活面において大変お世話になりました．出張の準備をはじめとする煩雑な手続きの遂行や，研究室のコーヒーの用意により，研究に集中できる環境を作っていただきました．

共同研究を行った株式会社スクーの皆様に感謝申し上げます．共同研究ミーティングは自分の研究について外部の方から意見を頂く貴重な時間でした．また，本研究で用いた Twitter のデータを提供して下さった電気通信大学の柳井啓司教授に感謝申し上げます．

同期の堀口くん，胡さん，藤本くん，山本さん，グェンくんは今日まで切磋琢磨してきた仲間です．パソコンとネット環境さえあればどこでも研究できるような分野でしたが，研究室に来れば同期の誰かしらがいたからこそ，毎日研究室に足を運ぶことができました．小野くん，糸川くん，小宮山くんとは共に修了することはできませんでしたが，以前のように，思い立ったときにふらりと遊びに行ければと思います．

その他の研究室の皆さんとも，先輩後輩を問わず研究室ミーティングでは貴重な意見を頂いたり，研究が一段落したときには雑談や遊戯に興じたりしました．今ではどれも懐かしい思い出です．またお会いする機会はあると思うので，その時を楽しみにしています．

最後に，6 年間大学に通わせてくれた両親に感謝します．ありがとうございました．

2017 年 2 月 3 日