

指導教員（主査）：山本祐輔 講師

副査：森田純哉 准教授

2018 年度 静岡大学情報学部 卒業論文

脚本の内容と構成要素に基づく 映画印象推定

静岡大学 情報学部 IS プログラム 所属

学籍番号 70511059

中野 裕介

2019 年 2 月 14 日

概要

本稿では、映画の脚本データに着目し、脚本の内容と構成要素を分析することで映画の印象を推定する手法を提案する。ここで分析に用いる脚本の構成要素とは、脚本の基本的な記述ルールによって定められた文章のまとまりを指す。提案システムでは、映像化されたとき視聴者に伝わる情報である脚本内容に Doc2Vec アルゴリズムを適用し、得られた分散表現を映画の特徴量とする。2 つ目のアプローチでは、「ト書き」や「シーンの柱」など脚本の構成要素の出現頻度や長さに着目し、映画の特徴量とする。以上の 2 つの要素について考慮した映画特徴ベクトルを入力として受け取り、SVM を用いた映画の印象推定器を構築した。本稿では印象推定器の構築および評価のために、IMSDb に収録された映画脚本データを用いた。評価実験として、特徴量として脚本内容のみを考慮した場合、脚本の構成要素のみを考慮した場合、両方を考慮した場合の映画の印象推定精度の評価を行った。その結果、脚本の内容と構成要素の両方を考慮した場合、より高い精度で映画の印象を分類することが可能であることを明らかになった。

目次

第 1 章	はじめに	5
第 2 章	関連研究	7
2.1	ジャンル分類	7
2.2	ユーザー嗜好に合わせた映画推薦	8
2.3	本研究との比較	8
第 3 章	利用技術	9
3.1	Doc2vec	9
3.2	サポートベクタマシン	10
第 4 章	提案手法	12
4.1	脚本の内容と構造	12
4.2	学習に用いたデータについて	13
4.3	特徴ベクトルの作成	17
4.4	印象推定器の構築	17
第 5 章	評価実験	19
第 6 章	結果	21
第 7 章	考察	24
7.1	構成要素について	24
7.2	データ数について	24
7.3	分類におけるラベル付けについて	25
第 8 章	おわりに	26

目次

3.1	線形分離可能なデータモデル	10
4.1	IMSDb の実際の脚本データ 引用元 https://www.imsdb.com/scripts/La-La-Land.html に加筆	14

表目次

4.1	構成要素の意味と本稿での定義	15
4.2	映画ジャンル毎のデータ件数と統計量	16
4.3	構成要素の抽出法についての本稿での定義	18
6.1	手法毎のジャンル分類精度一覧.	22
6.2	分類手法とジャンルのデータ件数の比較.	23

第 1 章

はじめに

情報通信，および検索技術の進歩により，人々が取得可能な映画データは増加の一途をたどっている．その一方で，自らの視聴したい映画を主体的に選択するにあたり，一般に広まっている手法は限定的であると考えられる．映画の内容について直接触れる判断材料は公式に提供されたものとして宣伝用の画像やトレーラー動画，あらすじなどが挙げられ，その他には映画に設定されたジャンルや監督，演者などの情報を調べることができる．またユーザの投稿したレビューも内容に触れるものとして挙げられるが，各ユーザの主観に大きく左右されるほか，サイトによってはデータ数が膨大で適切なものを効率的に探すことが難しいという問題も挙げられる．結果として増加する映画候補からの選択はますます難しくなっているといえる．

映画の内容を表す情報を調べるユーザの心理として，映画が自分の好む要素を含んでいるかを判断しているということが考えられる．これはすなわち，映画から受ける印象を映画の内容を表す雑多な判断材料から推測する試みにほかならない．映画から受ける印象をユーザが細かく設定し，該当する映画を検索することができれば，膨大な映画候補の中から好みの映画を効率よく検索できるようになると考えられる．

一般に，映画を分類する切り口としてはジャンルが用いられることが多い．しかし，映画を分類するジャンルには明確な定義が存在しない．そのため映画のジャンル分けはジャンル分けを行う者の主観によるところが大きい．また，映画によっては複数のジャンルに対してそのジャンルであるといえる要素を含んでいるものも存在する．例えば，登場人物の誇張された動作や陽気な会話によって視聴者にユーモラスな印象を与える映画はコメディというジャンルに属するといえる．しかしその物語の展開上で倒すべき悪役が設定され，登場人物同士が武器などを用いて戦う様が描かれれば，アクションというジャンルにも属するとみなすことができる．そのため，映画のどの部分に着目するかによって，ジャ

ジャンル分けの判断が大きく異なってくる可能性が考えられる。さらに映画によってはジャンル分けが難しく、一概にどのジャンルであると判断することが難しいものも存在する。それゆえに映画のもつ情報に基づいて機械的に複数のジャンル要素を持つ映画のジャンルを自動的に推定することは、ユーザの映画選択をより効果的に支援できるものであると考えられる。

本稿では、映画の脚本データを用いて映画のジャンルを推定する手法を提案する。提案手法は脚本内容と脚本構成要素という 2 要素から映画を特徴ベクトル化し、SVM を用いて映画のジャンル（印象）を分類する。映画の特徴ベクトル化および提案手法の評価に当たっては、IMSDb^{*1}の脚本データを利用する。評価実験の結果、提案手法は脚本内容、構成要素それぞれを単独で特徴量として用いた場合と比べ、脚本データ数上位 10 ジャンルにおいて最多の 5 ジャンルで最も高い F 値を示した。この手法の性能をさらに向上させることができれば、それぞれのジャンルを構成する要素の特徴を抜き出すことでジャンルがラベル付けされていない映画に対してもその分類が可能となるほか、人間の恣意的なジャンル設定では取得できない、その映画の内包する未知の要素を明らかにすることも期待される。

^{*1} <https://www.imsdb.com/>

第 2 章

関連研究

2.1 ジャンル分類

速水は映画ジャンルの分類における定義の曖昧さを指摘した上で、インタビューを通して一般に認知されている映画ジャンルを体系化した [11]。また、それぞれの映画ジャンルが含む属性をカテゴリー要素と定義し、各ジャンルの内包するカテゴリー要素やジャンル名称に着目し、ジャンル間にみられる重複関係などについて様々な考察を行った。

Zhou らは映画の予告などに用いられるトレーラー動画を利用したジャンル推定を試みた [7]。まず動画をショット単位に分割し、シーン特徴検出器を用いることによってシーン単位での分類を行った。次にそれらを Bag-of-Visual Words によって特徴ベクトル化し、k 近傍法によって IMDb^{*1}で設定された各トレーラーのジャンル进行分类するアルゴリズムを提案した。

Chu らは映画の宣伝に用いられるポスター画像を利用してジャンルの推定を試みた [2]。さまざまなメタデータの付随したポスターデータセットを収集し、ポスター内の物体进行分类できるようあらかじめ調整された畳み込みニューラルネットワークを用いて学習を行った。IMDb において付与されたジャンルのラベルに基づいてその傾向を得てジャンルの分類を試みた。

Ertugrul らは映画のあらすじ情報からジャンルを推定することを試みた [3]。あらすじに含まれる複数の文章を用いて Bi-LSTM と呼ばれる手法のネットワークをトレーニングし、文章毎に推定されたジャンルの多数決によってその映画のジャンルを決定した。

^{*1} <https://www.imdb.com/>

2.2 ユーザー嗜好に合わせた映画推薦

林らはインターネット上に投稿された映画のレビュー記事を利用し、ユーザーの入力したレビューから別の映画のレビューを提示することで映画を推薦するシステムを考案した [12]. その手法としてレビューの肯定、否定の表現に着目して興味語と定義し、TF-IDF に基づく従来の文書間の類似度の計算手法を改良した. それによって興味語の抽出を可能とし、それに基づいた類似レビューの提示を可能とした.

櫛見らもまたインターネット上の映画レビュー記事に着目し、レビューから映画の特徴ベクトルを生成し、ユーザが選択した映画と他の映画の特徴ベクトルの類似度から映画の推薦を行う手法を考案した [10]. 具体的な方法として、あらかじめ分かち書きされたレビューを Doc2vec の学習データとして用い、コサイン類似度によってレビュー間の類似度を求めた.

2.3 本研究との比較

上述の通り、従来の映画を対象とした研究においては、その目的の達成のためにトレーラー動画やレビューといった、映画のストーリーを部分的に写像したものを用いる研究が多い. その点で本研究で着目する脚本データはストーリー全体の内容を把握可能な形態であるといえ、その内容に対するより正確な分析を行うことが期待できる. また、脚本データの分析方法として、他の研究では扱われていない、その記述法に着目した構成要素という概念を取り入れることによって、ストーリーに付随する演出などといった特徴を間接的に取得することが期待できる. 映画ジャンル分類問題としても、これらの脚本内容と構成要素という概念によって従来手法より高い精度での分類が可能であると考えられる.

第 3 章

利用技術

本稿では，脚本データの分散表現を得るために Doc2vec を用い，また対象となる各ジャンルの分類器生成のためにサポートベクタマシン (以下，SVM) を利用した．

3.1 Doc2vec

Doc2vec[5] は，文書の分散表現を得ることのできる手法である．この手法は，Mikolov らの考案した単語の分散表現を得るための手法である Word2vec[6] を拡張したものである．Word2vec は同じ文脈で使われる単語であればその意味は類似する傾向にあるという Harris の分布仮説 [4] に基づき，対象とする単語とその周辺の語を関連性があるとみなすことで単語の類似度を推定する．この仮説は，周辺語を入力として対象とする単語を得る ContinuousBOW と，対象とする単語から周辺語を予測する Skip-gram と呼ばれる 2 つのアルゴリズムを用いることで Word2vec 上で実現されている．これらのニューラルネットワークは入力層，中間層，出力層に分かれ，ベクトルとして表現された単語が入力され，それが 2 度の重み付けを経てその他の単語との類似度を示すベクトルに変換され出力される．出力されたベクトルは Softmax 関数によって各単語間の出現率を示す確率を表すベクトルへと変換される．これによって出力されたベクトルは訓練データから算出された正解値と比較可能な形に正規化され，その誤差を修正するために重み付けに用いられた行列の値を調整することを繰り返して精度を高めていく．そして最終的にこの試行で調整された重み付け用の行列の各列が対象とする単語の特徴を示すベクトルとして様々な応用可能な分散表現となる．

Doc2vec は上述の Word2vec の考え方を拡張し，重み付け用の行列として文書そのものをベクトル化したものを並べたものを用意する．この際それらのベクトルの次元数は

単語の分散表現と同一である．前述の Word2vec で用いた 2 つのアルゴリズムである，ContinuousBOW と Skip-gram をそれぞれ拡張した PV-DM，PV-DBOW というアルゴリズムがこの手法では用いられる．PV-DM では対象とする単語を得るために周辺語に加えて文書 ID を与え，PV-DBOW では入力として文書 ID のみを用いる．そしてこれらもまた Softmax 関数によって変換されたのち，正解値と比較されて精度を高めていくことで文書ベクトルが分散表現として適切なものとなる．

本稿では脚本内容の特徴ベクトル化には PV-DM を採用した．

3.2 サポートベクタマシン

SVM[1] はパターン認識手法の一つであり，分類や回帰の教師あり学習に用いられる機械学習法である．一般的な利用方法として，本稿のような 2 値分類問題が挙げられる．これは図 3.1 のような A と B の 2 つにラベル付けされたデータが存在する線形分離可能な多次元空間において，それらを適切に分類することのできる超平面を求める問題である．

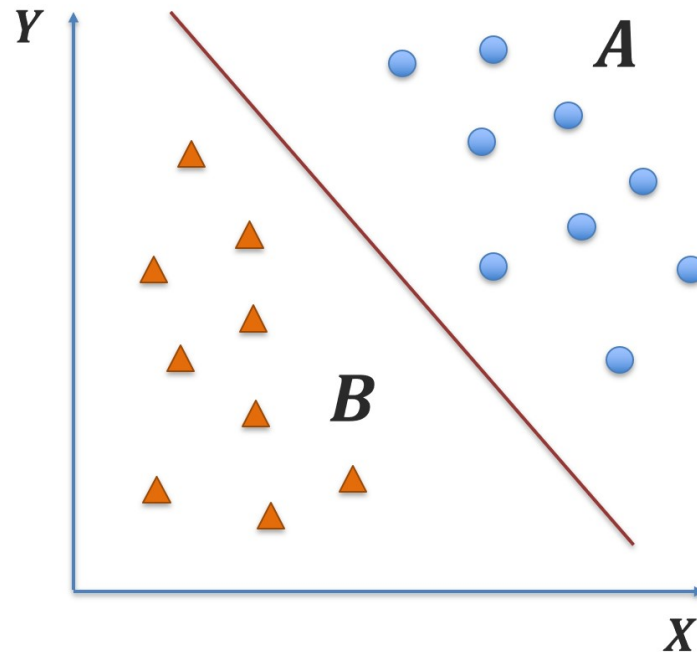


図 3.1 線形分離可能なデータモデル

他のパターン認識モデルと比較した場合の SVM の特徴として，マージン最大化が挙げられる．マージンとは，データのうちでも超平面に隣接するものであるサポートベクトル

と超平面のユークリッド距離を示すものである。この距離を最大とすることによって、生成された分類器の汎化性能の向上が期待できる。汎化性能とは分類器が学習に用いられたデータだけでなく、同じ分類問題を対象とした未知のデータに対しても適切な分類を行うことのできる能力を示すものである。マージンが小さく、サポートベクトルと超平面の距離が近い場合、微細なデータの値の揺れによって誤った分類をしてしまうリスクが高くなるといえ、マージン最大化を行うことは分類性能の向上につながるといえる。

また扱う問題によっては、明確に線形分離可能でないデータが存在することもある。SVM においては、カーネル関数を用いて対象となるデータをさらに高次元化することによって分離のできないベクトルを線形分離可能に変形することができ、これを非線形 SVM と呼ぶ。さらにカーネル関数をさらに用いることで、高次元化によって増大した計算量を軽減することが可能であり、これをカーネルトリックと呼ぶ。

本稿では非線形 SVM を用い、カーネルには RBF カーネルを使用した。

第 4 章

提案手法

我々は、脚本データを利用し、脚本の内容と構成要素という二つの点に着目してジャンル分類を行う手法を提案する。提案手法の主なステップは下記の通りである：

1. IMSDb データベースから脚本データを取得する
2. 脚本の内容と構成要素に着目し IMSDb に格納された各映画から特徴ベクトルを生成する
3. IMSDb で映画それぞれに設定されたジャンルタグをラベルとして SVM による機械学習を行い、分類器を構築する
4. 構築した分類器を用いてジャンルが未知の映画のジャンル推定を行う

以下では、脚本データに関する概念を整理した上で、各ステップの詳細について述べる。

4.1 脚本の内容と構造

本節では、本稿で用いる脚本内容と構成要素という用語について定義し、また用いる理由を示す。

構成要素とは、脚本を記述するにあたり定められた典型的なルールによって分類できる文章のまとまりを表す。Field[8]によると、脚本は大きく分けて「シーンの柱」、「ト書き」、「登場人物のセリフ」、「トランジション」によって構成されている。各構成要素の定義について、表 4.1 に記す。本稿ではこれら 4 つを脚本の構成要素とする。吉高 [9] らによれば、映画はカメラワークやショット長の遷移などの演出技法によって緊張感や安心感など、視聴者が映画から受け取ることのできる印象を操作できるという。映画はその脚本のページ数が映画上映時間に対応するといわれ、その点から記述の長さを実際の映画での

描写に相関があるといえる．構成要素は映画内での各種の描写の長さや遷移を表しているため，各ジャンルを推定するための一要因になると考えられる．本稿ではこれらを記法に基づいて取得するために表 4.1 のルールに基づいて，脚本データからシーン，ト書き，会話，トランジションの 4 つの部分を取り出した．

脚本内容については，脚本データのうち上述の脚本記述のためのルールに則り記述され，映像化した場合に視聴者に伝えられる場面の状況や登場人物の会話，動作について主に示された部分を指す．直接的な情景や出来事の描写は，ジャンル毎に受け取る典型的印象を推定する要因になると考えられる．

4.2 学習に用いたデータについて

本稿では，映画脚本データベースである IMSDb に保管された脚本データを，ジャンル分類器の構築及び評価に用いた．これらの脚本データには図 4.1 に示すように映像や音声として映画で描写されるすべての事物に関する情報が記載されており，共通の記述ルールによって構造化されている．これらは表 4.1 で定義された各構成要素の対象とする部分と対応している．映像として視聴者に見える部分の情報を記述したト書きや，登場人物のセリフなどといった要素についてその記述形式によって判別することが可能である．

本稿では IMSDb のウェブサイトをクロールし，計 1007 件の映画脚本データおよび各映画に割り振られた映画ジャンルを収集した．収集した脚本データ数及びその統計量をジャンル別で表 4.2 に記す．ここでの統計量については，表 4.1 によって定義された各構成要素のシーン別行数に対して，平均値，標準偏差，中央値の三つを当てはめたものである．なお，一つの脚本データに対して複数のジャンルのタグが付与される場合がある．

シーンの柱	—	[130	EXT. COFFEE SHOP - MOMENTS LATER
ト書き	—	[Mia exits the coffee shop...and is met by a CREW MEMBER on a GOLF CART. She gets on the cart -- and is driven away...
			トランジション — [CUT TO:
ト書き	—	[CLOSE ON hands on piano keys, fluttering across the ivories. We PULL BACK: it's SEBASTIAN. We're in...
シーンの柱	—	[131	INT. JAZZ CLUB - DAY
ト書き	—	[...a small jazz club. Simple, tasteful, cool. Stone arches in 1940's style. The seats close to the band, the piano in the center. The club has the same old-school character as the Lighthouse -- but it's not run-down. It's polished, inviting.
			The place is empty save for Sebastian and an EMPLOYEE. It's before-hours. Sebastian finishes playing. Feels out the lowest keys once more, then the highest. Then turns and --
会話	—	[SEBASTIAN Alright, I'm done. (gets up) Harris did a nice job with it.
			EMPLOYEE Took him long enough.

図 4.1 IMSDb の実際の脚本データ 引用元 <https://www.imsdb.com/scripts/La-La-Land.html> に加筆

表 4.1 構成要素の意味と本稿での定義

要素名	本稿における定義
シーン	<p>映画においてある場面での一連の動作の区切りを示すもの。下位概念としてカメラが切り替わるまでの部分を指し、これらのまとまりをシーンと見なせる「カット」、上位概念として物語的なシーンのまとまりを指す「シークエンス」が存在し、階層構造を成す。</p> <p>本稿においてはシーンの開始を示すシーンの柱から次のシーンの柱までの行。ここにはト書き、会話、トランジションもまた含まれる。なおシーンの柱とはその場所が建物の外であるか、中であるか、具体的な場所の名前、時間帯、の3つの順が左詰めに記載された行を指す。</p>
ト書き	<p>映像として実際に表現する場面の情景や登場人物の動作を記述した部分。</p> <p>本稿においては文章の開始位置がシーンの柱と同様に左詰めに記述された行。</p>
会話	<p>映画の登場人物が発するセリフについて記述した部分。</p> <p>本稿においては一定以上の空白を開けて行中央に揃えて記述された文章のうち、話者の名前を指す行でない部分。</p>
トランジション	<p>シーンの切り替わりの演出方法について説明した部分。</p> <p>本稿においては右詰めで記述され、各シーンの最後に記載される文章。</p>

表 4.2 映画ジャンル毎のデータ件数と統計量

ジャンル名	件数	シーン			ト書き			会話			トランジション		
		平均値	標準偏差	中央値	平均値	標準偏差	中央値	平均値	標準偏差	中央値	平均値	標準偏差	中央値
action	279	38.1	79.1	17.0	9.85	21.2	5.0	16.8	41.6	4.0	0.215	0.410	0.0
adventure	160	39.9	109.8	16.0	10.1	29.4	5.0	18.2	53.3	3.0	0.206	0.404	0.0
comedy	327	46.6	82.0	22.0	9.1	17.7	5.0	24.6	49.3	8.0	0.238	0.426	0.0
crime	198	39.8	72.8	18.0	9.4	18.7	5.0	18.7	40.8	4.0	0.242	0.428	0.0
drama	576	42.3	80.5	20.0	9.5	19.0	5.0	20.9	46.8	5.0	0.237	0.426	0.0
horror	136	39.6	74.1	18.0	11.3	20.3	5.0	15.2	36.6	2.0	0.253	0.435	0.0
mystery	100	39.2	63.9	18.0	9.5	15.2	5.0	18.4	37.2	4.0	0.229	0.420	0.0
romance	183	44.6	80.6	21.0	9.2	17.4	5.0	22.7	47.8	6.0	0.209	0.407	0.0
sci-fi	146	36.4	58.2	16.0	10.1	18.4	5.0	15.6	32.4	3.0	0.195	0.397	0.0
thriller	356	39.1	74.5	17.0	10.2	18.7	5.0	16.9	42.1	3.0	0.232	0.422	0.0

4.3 特徴ベクトルの作成

SVM によるジャンル推定器の構築にあたり、各脚本データを特徴ベクトル化した。特徴ベクトルは特徴量は下記の通りである：

- 300 次元：Doc2Vec によって得られた脚本内容の分散表現
- 28 次元：脚本の各構成要素に関する統計量

構成要素に関する特徴量については、表 4.3 の定義に基づき脚本データのシーンごとに分類して行数を取得し、各映画データについて「シーン数」「ト書き数」「会話数」「トランジション数」の最大値、最小値、総数、平均、分散、中央値、歪度を計算し、計 28 個の特徴量を抽出した。

脚本内容の特徴量化については、Doc2Vec[5] を用いた。Doc2vec は文中である単語の周辺に出現する単語の出現傾向を学習することによって、文書の分散表現を得る技術である。これは同じ文脈で用いられる単語は意味が類似している傾向があるという仮説に基づき、ある単語から周辺単語を予測するという問題をニューラルネットワークに学習させて単語の分散表現を得る Word2Vec[6] という技術を発展させたものである。Doc2vec では上記の考え方を文書に適用させる。

本稿では Doc2Vec を用い、IMSDb から取得したすべての脚本データの分散表現を得た。この際、IMSDb の脚本データからメタタグを取り除いたもの脚本内容を表すテキストデータと見なし、IMSDb から取得できる全脚本データを Doc2Vec の学習コーパスとした。

以上の構成要素および脚本内容に基づいたそれぞれの特徴量を組み合わせ 328 次元のベクトルとした。

4.4 印象推定器の構築

上述の 328 次元のベクトルを用いてジャンル毎に分類器を構築する。分類器の生成には教師あり機械学習法である SVM[1] を用いる。カーネルには RBF カーネルを指定した。分類器毎に対象とするジャンルのタグが付与されていた脚本データを正解データとして扱い、対象以外のジャンルの脚本データを不正解データとした。学習はジャンル毎に行い、ラベル付けを対象とするジャンルであるかそうでないかに限定し 2 値分類問題とした。

表 4.3 構成要素の抽出法についての本稿での定義

要素名	本稿における定義
最大値	シーン毎に対象の構成要素に該当する行を数えたとき，そのデータの中で最大となるものの値．
最小値	シーン毎に対象の構成要素に該当する行を数えたとき，そのデータの中で最小となるものの値．
総数	脚本データにおける対象の構成要素に該当する行の総数．
平均	シーン毎に対象の構成要素に該当する行を数え，そのデータの中での平均をとった値．
分散	シーン毎に対象の構成要素に該当する行を数えた集合の中で，そのデータの分散をとった値．
中央値	シーン毎に対象の構成要素に該当する行を数えた集合の中で，そのデータの中央値となるものの値．
歪度	シーン毎に対象の構成要素に該当する行を数えた集合の中で，そのデータの歪度をとった値．

第 5 章

評価実験

作成した印象推定器について評価実験を行った。本実験では、以下の 3 手法について性能の比較を行った。

- **d2v+esf** : 提案手法である脚本内容と構成要素を特徴量とした印象推定器
- **d2v** : 脚本内容のみを特徴量とした印象推定器
- **esf** : 脚本の構成要素のみを特徴量とした印象推定器

d2v+esf, **d2v**, **esf** 手法ともに、学習には SVM (RBF カーネル) を用いた。

性能評価は、4.2 節で述べた脚本データのうち、総数が 100 件以上であった 10 ジャンルの映画に対して行った。これは各映画ジャンルでそのジャンルタグが付与された脚本データの数にばらつきがあるためである。具体的には、以下の手順で各ジャンルの推定性能評価を行った。

1. ターゲットとなるジャンルを決める。
2. ターゲットジャンルの映画を正解データ、非ターゲットジャンルの映画を不正解データとする
3. 5 分割交差検証によって推定性能を評価するために、データを 4:1 の割合で訓練データと評価データに分割する
4. 訓練データに対してグリッドサーチと 5 分割交差検証^{*1}を適用し、最適な印象推定器を構築する
5. 構築した最適な印象推定器を評価データに適用し、評価指標スコアを算出する

^{*1} グリッドサーチの評価関数には F 値を用いた。

評価指標には，以下の適合率 (*precision*)，再現率 (*recall*)，F 値 (*F-value*)，精度 (*accuracy*) を用いた．本稿における定義を以下に示す．

$$precision = \frac{\text{正解と分類できた正解データ数}}{\text{正解と分類したデータ数}} \quad (5.1)$$

$$recall = \frac{\text{正解と分類できた正解データ数}}{\text{実際に正解であるデータ数}} \quad (5.2)$$

$$F\text{-value} = \frac{2recall \cdot precision}{recall + precision} \quad (5.3)$$

$$accuracy = \frac{\text{適切に分類できたデータ数}}{\text{すべての分類対象データ数}} \quad (5.4)$$

第 6 章

結果

5 章で述べた実験の結果について、表 6.1 に示す。d2v+esf は、対象としたすべてのジャンルで 6 割以上の F 値を示した。また esf に対しては F 値、精度ともにすべてのジャンルで値が上回ったほか、d2v に対しても F 値に関して drama, crime, adventure, horror, mystery の 5 つのジャンルでその値を上回り、また thriller, romance の 2 つのジャンルで同等の性能を示すことが明らかになった。

d2v+esf と d2v の F 値及びその脚本データの件数とデータ総数に対する割合を表 6.2 に示す。ジャンルはその脚本データの件数が多い順に上から並べた。また F 値について d2v+esf と d2v を比較した場合に値の大きかった方を太字で示したほか、それぞれのジャンルのデータの件数及びそれが全体に占める割合を示した。それぞれのジャンルのデータ件数が 200 個を超えるか否かでジャンルを区分した場合に、200 件を上回る 4 件のジャンルのうち 1 件が d2v の性能を上回り、2 件がその性能を下回った。それに対して、データ件数が 200 以下の 6 件のジャンルに着目した時、4 件のジャンルで d2v の性能を上回り、1 件がその性能を下回った。

表 6.1 手法毎のジャンル分類精度一覧.

ジャンル名	precision			recall			F-value			accuracy		
	d2v+esf	d2v	esf	d2v+esf	d2v	esf	d2v+esf	d2v	esf	d2v+esf	d2v	esf
action	0.518	0.535	0.369	0.770	0.767	0.537	0.711	0.722	0.580	0.741	0.753	0.629
adventure	0.500	0.530	0.185	0.562	0.493	0.262	0.701	0.689	0.511	0.835	0.841	0.701
comedy	0.607	0.601	0.406	0.718	0.746	0.627	0.734	0.739	0.574	0.759	0.761	0.590
crime	0.426	0.424	0.240	0.717	0.681	0.475	0.686	0.681	0.522	0.759	0.760	0.610
drama	0.858	0.822	0.616	0.616	0.652	0.602	0.724	0.720	0.560	0.725	0.722	0.566
horror	0.298	0.340	0.233	0.542	0.559	0.580	0.622	0.607	0.561	0.773	0.767	0.683
mystery	0.328	0.327	0.136	0.559	0.620	0.300	0.655	0.642	0.519	0.842	0.827	0.748
romance	1.000	1.000	0.225	0.781	0.781	0.448	0.926	0.926	0.521	0.961	0.961	0.626
sci-fi	0.740	0.758	0.194	0.718	0.732	0.376	0.839	0.849	0.531	0.923	0.927	0.694
thriller	1.000	1.000	0.457	0.974	0.974	0.601	0.990	0.990	0.594	0.991	0.991	0.610

表 6.2 分類手法とジャンルのデータ件数の比較.

ジャンル名	F-value		データ件数	ジャンル割合 (%)
	d2v+esf	d2v		
drama	0.724	0.720	576	57.2
thriller	0.990	0.990	356	35.3
comedy	0.734	0.739	327	32.4
action	0.711	0.722	279	27.7
crime	0.686	0.681	198	19.7
romance	0.926	0.926	183	18.2
adventure	0.701	0.689	160	15.8
sci-fi	0.839	0.849	146	14.5
horror	0.622	0.607	136	13.5
mystery	0.655	0.642	100	9.9

第 7 章

考察

本章では、6 章で記した結果について考察を行う。表 6.1 に示す通り、F 値に着目したとき、d2v+esf は d2v より多くのジャンルで最も高い数値を示した。しかし、ジャンルによってその性能の変化に差異が生じたほか、データ件数が 200 件以上のグループではあまり性能の変化が見られなかった。これを踏まえて 3 点について考察を行う。

7.1 構成要素について

本稿では脚本データの利用法として、脚本の内容に加えて構成要素という観点を用いた。これは映画の映像上の演出を間接的に表現するとして着目したものであるが、本研究では脚本における「シーン」「ト書き」「会話」「トランジション」の数を最大値、最小値など 7 つの統計量で表現することで、構成要素の特徴量を 28 次元に圧縮している。すなわち、提案手法では映画を構成する各シーンの情報およびシーン間の順序関係を無視していることになる。映画における演出はカメラワークやショット時間の変化量によってとらえることが可能 [9] であり、ト書きの 1 文ごとにカットが切り替わるという考え方や脚本の長さと映像の長さは比例するという考え方から、ショット時間の遷移を予測できることが考えられる。それゆえ、映画の特徴量を生成する際にシーン間の順序関係や時系列情報を踏まえることで、d2v+esf の性能が向上する可能性が考えられる。

7.2 データ数について

本稿では IMSDb から 1007 件の脚本データを取得し、それをジャンル別に扱って分類器を生成した。4.1 節で述べた通り、ジャンルタグは一つの映画に対して複数付けられて

いる場合があり，多いものでは6つ以上のジャンルが1つの映画に対して付与されている．ジャンルを複数もつ映画については複数のジャンルの印象を有することとなり，またそれらのジャンルのうちどれが最も映画に反映されているかの優劣は判断できず，そのジャンルタグを付与するに至る基準も存在しない．仮説として，複数のジャンルが与えられた脚本データというものは，そのシーン毎に異なるジャンルが割り当てられるということが考えられる．例えば全編を通してコメディ的な要素を持つ映画のあるシーンにおいてアクション的な要素が含まれていれば，その映画はコメディ及びアクションというジャンルを有する映画と見なされる．それゆえにこの映画をアクションジャンルに該当する物と見なすとき，アクション映画の分類器がコメディ的な要素を含んでしまう可能性が考えられる．この点から表 6.2 で示された脚本データが 200 件より多いジャンルとそれ以下のジャンルで d2v+esf が d2v の性能を上回った数に差が見られたのは，件数が多いために上述の理由によって他ジャンルとの差異が小さくなり，適切な構成要素の分類ができなかったためと考えることができる．

7.3 分類におけるラベル付けについて

本稿では印象推定器による分類の対象としてジャンルを用いた．ジャンルは映画の印象の分類において，一般に広く認知されている方法であり，映画を用いた商業的な活動の様々な局面で利用されている．しかしジャンルというものの区分について明確な定義は存在せず，人によっては同一の作品であっても異なるジャンルを想定する場合も考えられる．

速水 [11] によれば映画のジャンル分けとそのジャンルにおける印象をさらに細分化した「カテゴリー要素」を照らし合わせた場合，異なるジャンル間で同一のカテゴリー要素を共有している場合も存在する．ここから映画はジャンル分けによってそれぞれが完全な独立関係にはならず，共通している部分も多いといえる．

上述の観点から，映画の分類にあたってジャンルよりもさらに粒度の小さい印象に着目して分類を行うほうが適切な分類を行えると考えられる．具体的には，映画内の登場人物の立場が逆転した場合などに用いられる「どんでん返し」といった考え方が挙げられる．

第 8 章

おわりに

本稿では，映画の印象推定を行うための手法として脚本データに着目し，脚本データの捉え方として脚本内容と構成要素という 2 つの指標を提案した．またそれらを特徴量として SVM による映画ジャンルの分類器を構築した．IMSDb の脚本データを用いた評価実験を通じて，提案手法は他の手法と比較して最も有用であるということを示した．

今後の課題として，ジャンルに代わる新たな分類指標の模索が挙げられる．今回行ったジャンルに基づく印象推定の方法では，そのデータ数が多いジャンルについては提案手法があまり良い性能を示さなかった．これについては以下の理由が考えられる．第 1 の理由として，複数のジャンルに属している映画が存在し，その種の映画に付与されたジャンルは今回の実験においてはジャンル分類の正解データとして機能しない可能性が考えられる．提案手法が性能を発揮しなかった第 2 の理由としては，映画を構成するシーン単位のデータを扱わず，映画全体を 1 つのデータと見なして分析を行ったことが挙げられる．このためにシーン間の変化量などを扱えず，映画ジャンル毎の差異を適切に写像できていなかったことが考えられる．このことは，ジャンル以外の指標で映画の印象を捉えるという課題にもつながる．例として「どんでん返し」などの映画全体を通して感じられる印象が挙げられるほか，映画全体でなくシーン単位に表現される，視聴者の感情に影響が存在することも考えられる．ジャンルに対して一般的ではないことから，それらを分類指標として用いるためにはこちらから一定の基準を設ける必要も考えられるため，その基準を得るための方法や理論についても探っていく必要があるといえる．

参考文献

- [1] Boser, B. E., Guyon, I. M. and Vapnik, V. N.: A Training Algorithm for Optimal Margin Classifiers, *Proceedings of the Fifth Annual Workshop on Computational Learning Theory (COLT 1992)*, ACM, pp. 144–152 (1992).
- [2] Chu, W.-T. and Guo, H.-J.: Movie Genre Classification Based on Poster Images with Deep Neural Networks, *Proceedings of the Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes (MUSA2 2017)*, ACM, pp. 39–45 (2017).
- [3] Ertugrul, A. M. and Karagoz, P.: Movie Genre Classification from Plot Summaries Using Bidirectional LSTM, *Proceedings of IEEE 12th International Conference on Semantic Computing (ICSC 2018)*, pp. 248–251 (2018).
- [4] Harris, Z. S.: Distributional Structure, *WORD*, 10:2-3, pp. 146–162 (1954).
- [5] Le, Q. and Mikolov, T.: Distributed Representations of Sentences and Documents, *Proceedings of the 31st International Conference on International Conference on Machine Learning (ICML 2014)*, pp. II–1188–II–1196 (2014).
- [6] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J.: Distributed Representations of Words and Phrases and Their Compositionality, *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS 2013)*, pp. 3111–3119 (2013).
- [7] Zhou, H., Hermans, T., Karandikar, A. V. and Rehg, J. M.: Movie Genre Classification via Scene Categorization, *Proceedings of the 18th ACM International Conference on Multimedia (MM 2010)*, ACM, pp. 747–750 (2010).
- [8] シド・フィールド映画を書くためにあなたがしなくてはならないこと, フィルムアート社 (2009).
- [9] 吉高淳夫, 田中荘詩, 平嶋宗映画等を対象としたダイジェスト映像生成のための

映像特徴に関する検討, 情報処理学会研究報告 ヒューマンコンピュータインタラクション (HCI), pp. 79–86 (2007).

- [10] 櫛見圭司, 北山大輔 ユーザレビューを用いた全体的・部分的観点の類似に基づく映画推薦, *DEIM Forum 2017* (2017).
- [11] 速水素子 映画ジャンルの「民族分類」, 公民論集, Vol. 4, No. 6, pp. 59–85 (1996).
- [12] 林貴宏, 尾内理紀夫 Web 上のレビューを利用した映画推薦システム, 人工知能学会論文誌, Vol. 30, No. 1, pp. 102–111 (2015).