

感性計量による映画リコメンデーションシステムの実現 —映画鑑賞で用いられる意味空間の世界—

飯島 萌子[†] 清木 康[‡]

[†] 慶應義塾大学環境情報学部 〒252-0805 神奈川県藤沢市遠藤 5322

[‡] 慶應義塾大学政策・メディア研究科 〒252-0805 神奈川県藤沢市遠藤 5322

E-mail: [†] {t16046mi, kiyoki}@sfc.keio.ac.jp

あらまし 映画の好みや評価の基準は個々によって実にさまざまであるため、個人に合わせた映画レコメンデーションシステムが必要になる。本方式は、意味の数学モデルを利用し、映画に関する意味空間を定義し、その空間上で映画の特徴や個人の好みを部分空間として構築する。本方式は、両者の相関計算を行うことにより、一人ひとりの好みに合った映画を抽出し、推薦する仕組みを有する。本方式では、感性をもとに映画の価値を計量することで、感性を切り口にした新しいユーザーエクスペリエンスを提供する。また本方式により、短期的には効率的な映画検索の実現、長期的には映画ファンが増加するための基礎となる方式として発展させていきたい。

キーワード データベース, 推薦方式, 意味の数学モデル, 映画, 価値観, レビュー, 感性, 内積

1. はじめに

百貨店で服やコスメを買う若者が増えている。ネットで多種多様な商品が買えるなか、情報が多すぎて何をどう選んでいいのかわからず悩む若者が少なくないからだ。こうして若者の心を掴んでいるのが、専門知識に長けた百貨店のスタッフである。

映画をめぐる情報過多も深刻だ。日本国内で2016年に公開された映画は1149本で、単純計算すると1日3本のペースで公開されたことになる。娯楽にかかる時間が限られているなか、一本あたりの拘束時間が長い映画の場合は特に自分の好みに合った映画を探すことが重要だ。

昨今では好みの映画を探すのに、レビュー（口コミ）やスコア（評価点）が掲載された映画サイトを利用するのが一般的だ。私自身、レビューを読んで映画を選択する場合や、読んでいて「なるほど」と思ったレビューからそのユーザーのレビューサイトを閲覧し、そのユーザーが高く評価している映画を鑑賞対象とすることもよくある。

しかしながら、例えばある特定の映画に対して他人と似た感想を持ったとしても、その人が高く評価している他の映画に対しては互いに全く別の感想を持つ可能性は低くない。つまり感性とは、性格[13]や気分、思考の深さによって複雑にかたちづくられる個人に固有な概念なのである。

そこで本方式では、ひとによって映画の好みはさまざまであることを踏まえた上で、ひとが映画を評価するときに着目する観点をもとに映画の感性空間を構築した。また、同じジャンルでも毛色の違う映画は沢山あることから、今回はミュージカル映画に限定している。

本方式の目的は一人ひとりに合った映画をレコメンデーションシステムの実現であり、それと同時に、感性を切り口にした新しいユーザーエクスペリエンスを、映画レコメンデーションシステムを通して提供することだ。個人が楽しみながら短時間で良作の作品と出会える機会の提供を目指す。

2. 基本方式

個々の映画について、ひとはそれぞれオリジナルな物差しを持ち、あらゆる事象を認識している。興味深いのは、その物差しが「動的」指標である点だ。

例えば、何に対してどのくらい強い印象を受けるのかは、個人の性格によって異なる。加えて、コンテンツへの理解度やその日の気分次第で、個人の物差しには大きくブレが生じると考えている。

本方式は、数学モデル化された感性空間に映画を表現し、そこに個々の「動的」指標を重み付けすることで、個人と映画の相性を計量可能にする。

2-1. 映画鑑賞における意味空間

本方式は、データに意味の数学モデル[4]を適用して、映画を意味的に計量する。意味の数学モデルを用いることにより、特徴量付きベクトルとして定義したマルチメディアデータを、正規直交系をなす空間上でその時々文脈で選択された部分空間に射影し計量することで、文脈ごとに相関の強いデータを抽出することができる。

各映画の持つ特徴の種類が意味的に重複しない最小単位（以下「意味素」という） n 個に分類できると仮定すると、映画の感性空間 I は f の集合 $f_n = \{f_1, f_2, \dots, f_n\}$ を基底とする n 次元空間で構成されるといえる。また、

この n 個の意味素は互いに正規直交な特徴量を持っている。

意味素に関しては、映画サイトのレビューからデータを収集して形態素解析を行い、得られた結果を参考に定義付けを行った。もともとはコアな映画好きの感性が集積された映画雑誌[5][6]を対象にしていたのだが、コアな映画好きと一般的な映画ファンとでは映画に対する嗜好がかなり異なることから[7]、両者の感性が入り交じる映画サイトを活用することに方針を変更した。

2-2.取得データ概要

本方式では、国内最大級の映画レビュー数を誇る映画情報サービス・Filmarks [9]からデータを取得し、データ分析を行う。

Filmarks は映画専用 SNS アプリだ。①見た映画を記録したり、②見たい映画をチェックしたりできる[図 1]。また、③見た映画に 0.1 刻みに 0 点から 5 点まででスコアをつけたり、④コメントやメモを書き込んだりできる[図 2]。フォロー機能を使えば、映画の好みが似ているユーザーと繋がれる。もちろん Web サイトからのアクセスも可能だ。



図 1：アプリ画面



図 2：アプリ画面

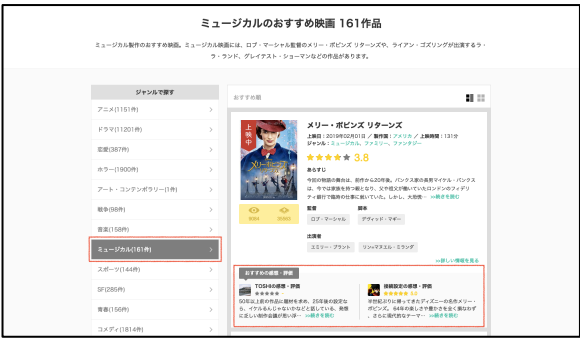


図 3：Web サイト

映画ジャンルでミュージカル映画を選択すると、ページごとに 20 作品ずつオススメ順に表示される[図 3]。ここでは 2019/02/16 の時点でトップ 20 の映画を対象にした。レビューは新しいものから順に集め、20 作品で合計 5,000 件のレビューを取得した[図 4]。

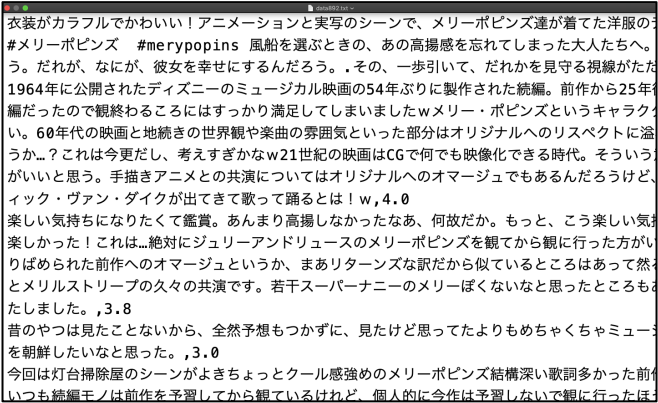


図 4：取得データの例

2-3.形態素分析と意味素候補の選定

映画のレビューとは、映画鑑賞を通して自分の感じたことや思ったことを文字で自由に表現することで、他人と感想を共有したり記録として記憶したりするためのものである。

このレビューデータから、ひとが映画を好んだり評価したりする要素を分析する。そこで意味素候補となるキーワードを探すために、レビューの形態素解析及び単語のカウントを行った。

形態素解析では、単語そのものに人の感性が含まれている品詞を対象にした[12,ChaSen 品詞体系参照]。具体的には名詞や形容詞、動詞-自立である。一方で、集計する必要がない品詞としては以下の 13 種類を設定した。

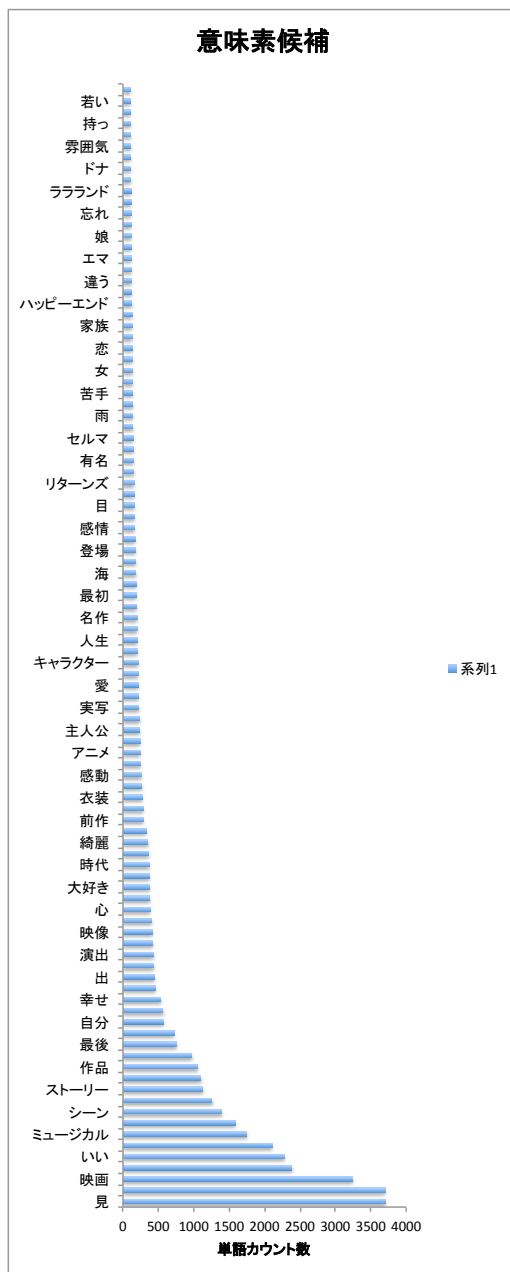
- ① 動詞,非自立- (例)「しまう」「ちやう」
- ② 動詞,接尾- (例)「られる」「させる」
- ③ 名詞,数- 数値
- ④ 名詞,非自立- (例)「もの」「かぎり」
- ⑤ 名詞,引用文字列- (例)「いわく」

- ⑥ 名詞, ナイ形容詞語幹- (例)「申し訳」「とんでも」
- ⑦ 名詞, 動詞非自立的- (例)「ごらん」「ちょうだい」
- ⑧ 名詞, 副詞可能- (例)「せんだって」「当分」
- ⑨ 名詞, 接尾- (例)「%」「み」
- ⑩ 名詞, 代名詞- (例)「そこ」「俺」
- ⑪ 名詞, 特殊- (例)「そう」「そ」
- ⑫ 形容詞, 接尾- (例)「ったらしい」「ぼい」
- ⑬ 形容詞, 非自立- (例)「つらい」「がたい」

次に、表記は異なるが意味は同じである同義語について補正調整を行った。補正の対象は、単語のカウント数が 100 回以上のもののみとし、意味素候補の数が合計 100 個になるまで調整を繰り返した。

(例)「見」「観」→「見」

「思っ」「思い」「思わ」→「思っ」



2-4.“絵画で”検索

既存の映画検索システムは、上映中の作品やジャンルの選択に始まり、製作国や製作年、キーワード語によるタグ検索など、非常に多岐にわたる。ただし自分の好みが明確でない場合、複数のアプローチを試しながら作品の目星を付けることになり、情報過多で困る私たちの根本的な問題解決にはなっていなかった。

また映画のレビューは誰でも自由に投稿できることが魅力である一方で、それによって生じる問題もある。つまり映画評論家のような専門家ではなく、大衆が主体となって各自の思考や嗜好を言語化することは、そのプロセスの途中で真の感性が抜け落ちるリスクの増加を意味すると考えている。

そこでミュージカル映画のレビューの頻出語（意味素候補）はあくまでも参考にしながら、自分のミュージカル映画に対する思考を交えながらイメージを膨らませ、自分のイメージに近い 8 つの絵画を意味素とし、検索自体を絵画経由で行うことにした[14]。感性の言語化をあえてあいまいなままにすることで、言語を超越した感性による映画検索を文字通り実現できると考える。尚“ ”の中が意味素である。

Picture 1 : “happy”



Pierre-Auguste Renoir. *Two Sisters (On the Terrace)*, 1881. The Art Institute of Chicago. This information, which is available on the object page for each work, is also made available under Creative Commons Zero (CC0).

Picture 2 : “cry”



Vincent van Gogh. *Weeping Woman*, 1883. The Art Institute of Chicago. This information, which is

available on the object page for each work, is also made available under Creative Commons Zero (CC0).

Picture 3 : "anime"



Henri de Toulouse-Lautrec. *L'Artisan Moderne*, 1896. The Art Institute of Chicago. This information, which is available on the object page for each work, is also made available under Creative Commons Zero (CC0).

Picture 4 : "dance"



Hilaire Germain Edgar Degas. *On the Stage*, 1876-1877. The Art Institute of Chicago. This information, which is available on the object page for each work, is also made available under Creative Commons Zero (CC0).

Picture 5 : "song"



Jan Steen. *The Family Concert*, 1666. The Art Institute of Chicago. This information, which is available on the object page for each work, is also made available under Creative Commons Zero (CC0).

Picture 6 : "old_days"



Gertrude Käsebier. *Blessed Art Thou Among Women*, 1899. The Art Institute of Chicago. This information, which is available on the object page for each work, is also made available under Creative Commons Zero (CC0).

Picture 7 : "laugh"



Smiling Figurine, 1000BC-300BC. The Art Institute of Chicago. This information, which is available on the object page for each work, is also made available under Creative Commons Zero (CC0).

Picture 8 : "story"



Jean Baptiste Camille Corot. *Interrupted Reading*, 1865-1875. The Art Institute of Chicago. This information, which is available on the object page for each work, is also made available under Creative Commons Zero (CC0).

2-5.意味的計量方法

本方式は、絵画から想起された感性によって検索するシステムである。絵画を選択した後に簡単な質問に答えてもらうことで性格や気分、思考の深さを推測し、一人一人に合ったミュージカル映画を推薦する。

例えば、疲れていたり落ち込んでいたりする人には深く心をえぐられるような映画を提案しないようにするなど、気軽に見られる映画なのか、かなり重い内容の映画なのかを考慮して提案できるようにした。

また映画に対する偏愛度を分析して、偏愛度が低いと判断される人には比較的最近公開されたアメリカ映画や邦画を優先的に提案したり、偏愛度が高いと判断された人には日米以外の製作国の映画や白黒映画を優先的に提案したりするようにした。これは、一般的な映画ファンが邦画やアメリカ映画を好む一方で、コアな映画好きは幅広い製作国の作品を好むという藤原氏の分析結果[7]を踏まえている。

3.実現方法

本章では、映画レコメンドシステムの概要や感性計量方法、使用データ、さらに実現環境を示す。

3-1.システム概要

まず、本システムの構成は以下の図のようになっている。

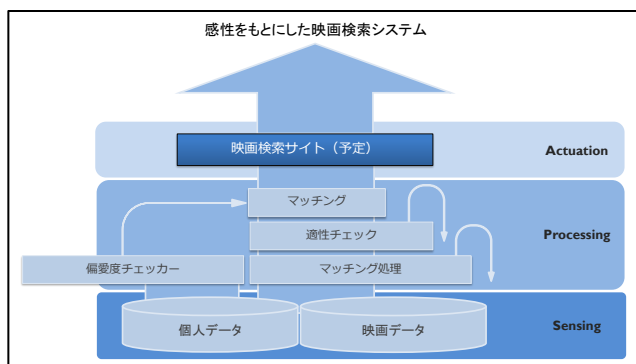


図 5：感性計量による映画レコメンドシステム

3-2.感性計量方法

検索者のデータが登録されると、データベースに登録された映画のデータが連動し、感性計量が開始され、各ステップを経て個々人に合った映画が推薦される仕組みになっている。

① マッチング処理 [図 7：青枠データについて]

1. 文字データの一致

項目「muke」と「hour」に書かれた両テーブルの文字データが合致している場合、次のステップへ進む。

2. 数値データの演算

残りの項目について、名前が同じ項目同士で数値を掛け合わせる。この計 8 項目の合計値が 1 以上ならば、次のステップへ進む。
(例) `movie.song * record.song`

② 適性チェック [図 7：緑枠データについて]

`record.mental` について、「tired」を 1、「good」を 0 とする。また `movie.energy` について、重い内容の映画ならば気合を入れて見る必要があるので「needed」とし 1 を、気軽に見られる内容ならば心の準備が不要であるため「not_needed」とし 0 を入力する。

このとき、`movie.energy * record.mental` の値が 0 以下ならば次のステップへ進む。

③ 偏愛度チェッカー [図 7：赤枠データについて]

年齢と生涯映画鑑賞数から、映画に対する検索者の思考の深さや偏愛度合いを分析する。

年齢 (`record.age`):

0 歳から 14 歳を 1、
15 歳から 24 歳を 2、
25 歳から 34 歳を 3、
35 歳から 45 歳を 4、
45 歳以上を 5 とする。

生涯映画鑑賞数 (`record.films_number`):

0 本から 199 本を 1、
200 本から 299 本を 2、
300 本から 399 本を 3、
400 本から 499 本を 4、
500 本以上を 5 とする。

なお、同じ映画を複数回鑑賞した場合でも 1 本としてカウントする。

こうして換算し直した「年齢」と「生涯映画鑑賞数」の数値を掛け合わせた値を偏愛度とする。

偏愛度:

10 未満を偏愛度低め (一般的な映画ファン)、
10 以上を偏愛度高め (コアな映画好き) とする。

④ マッチング [図 7：赤枠データについて]

`movie.country`:

japan, us ならば -1、
other ならば 1 とする。

`movie.color`:

color ならば -1、
blackwhite ならば 1 とする。

`made`:

latest, heisei ならば -1、

showa ならば 1 とする。

マッチングスコア：

movie.country, movie.color, movie.made の 3 項目の数値を足し算して求めた合計値 (③) に、①の合計値を掛け合わせたもの。なおマッチングの対象として考慮するのは、②のステップを通過した映画のみとする。

そのようにして、③で映画に対する偏愛度が高いと判断された人にはマッチングスコアの高い順に映画を推薦し、偏愛度が低いと判断された人についてはマッチングスコアの低い順に映画を推薦する。

movie	
title	
country	japan, us, other
muke	child, adult
hour	short, long
color	full, blackwhite
made	showa, heisei, latest
song	最も該当する要素を 2 つ選び, 1 を入力 その他は 0 を入力
story	
anime	
dance	
happy	
laugh	
cry	
old_days	
energy	-1: not_needed, 1: needed

record	
age	
films_number	
mental	tired, good
muke	child, adult
hour	short, long
song	最も該当する要素を 2 つ選び, 1 を入力 その他は 0 を入力
story	
anime	
dance	
happy	
laugh	
cry	
old_days	

図 6：テーブル構造

3-3. 使用データ

- ・IMDB（ウェブサイト）: 参考文献
- ・Filmarks（アプリ）: 参考文献

3-4. 実践

localhost で以下のようなウェブサイトを作成した。

localhost:81/query.php

なまえ

年齢

生涯映画鑑賞数（なお同じ映画を複数回観た場合も 1 回とカウントします）

現在の精神状態 ☐ 良好 ☐ 疲れ気味

色彩 ☐ フルカラー ☐ 白黒

鑑賞時間 ☐ 普通 ☐ 長め

映画のイメージ（2つ選択）

☐ 該当する ☐ 該当しない

Jan Steen. The Family Concert, 1666. The Art Institute of Chicago. This information, which Commons Zero (CC0).

上記のページに私自身の情報を入力すると、結果は以下のように返される。

localhost:81/recommend.php

映画を検索

もえこさんにお勧めの映画

あなたの映画に対する偏愛度は【高め】です。

Miss Saigon
-1点

Singing In The Rain
-1点

La La Land
-3点

The Greatest Showman
-6点

尚このシステムの裏側のデータベースには以下の図のように映画が 5 本分を登録されている他、検索者の入力する個人データも、検索ページがサーバーに送信されるのと同時に、自動的にデータベースに書き加

えられる仕組みになっている。

```
mysql> select * from movie;
```

PID	title	country	muke	hour	color	made	song	story	anime	dance	happy	laugh	cry	old_days	energy
1	La La Land	-1	adult	long	-1	-1	1	1	1	0	0	0	0	0	-1
2	The Greatest Showman	-1	adult	long	-1	-1	1	0	0	1	0	0	0	0	-1
3	Dancer In The Dark	-1	adult	long	-1	-1	0	1	0	0	0	0	1	0	1
4	Singing In The Rain	-1	adult	long	-1	1	1	0	0	0	1	1	0	0	-1
5	Miss Saigon	1	adult	long	-1	-1	1	1	0	0	0	0	0	0	1

5 rows in set (0.00 sec)

```
mysql> select * from record;
```

PID	age	films_number	mental	muke	hour	song	story	anime	dance	happy	laugh	cry	old_days
1	2	5	0	adult	long	1	0	0	1	0	0	0	0

以下、命令文である。

```
SELECT m.title,
m.song*r.song+m.story*r.story+m.anime*r.anime+m.dance*
r.dance+m.happy*r.happy+m.laugh*r.laugh+m.cry*r.cry+m.
old_days*r.old_days AS maru_ichi, m.energy* r.mental AS
maru_ni, r.age*r.films_number AS maru_san, m.country +
m.color + m.made AS maru_yon,
(m.song*r.song+m.story*r.story+m.anime*r.anime+m.dance
*r.dance+m.happy*r.happy+m.laugh*r.laugh+m.cry*r.cry+m.
old_days*r.old_days) * (m.country + m.color + m.made)
AS match_score FROM movie AS m, record AS r WHERE m.muke
= r.muke AND m.hour = r.hour AND
(m.song*r.song+m.story*r.story+m.anime*r.anime+m.dance
*r.dance+m.happy*r.happy+m.laugh*r.laugh+m.cry*r.cry+m.
old_days*r.old_days) >= 1 AND m.energy * r.mental <= 0
ORDER BY
(m.song*r.song+m.story*r.story+m.anime*r.anime+m.dance
*r.dance+m.happy*r.happy+m.laugh*r.laugh+m.cry*r.cry+m.
old_days*r.old_days) * (m.country + m.color + m.made)
DESC;
```

4-1.考察と今後の方針

学会での議論を通して、今後に向けた方針が大きく5つ定まった。

まず、絵画の扱い方についてである。本方式では、それぞれの意味素にふさわしい絵画を、それぞれ一点ずつ選定していたが、これでは絵画の解釈がただ1通りであると認めているに等しい。今後は、各意味素に相応する絵画を複数以上選定すると同時に、各絵画につき対応する意味素は一つという原則を廃止する方針だ。

次に、本ウェブサイトを利用するターゲットの細かな状況設定に関してである。映画を一人で見るのか、複数人で見るのか。複数人の場合、恋人なのか、友達なのか。或いは、家族なのか。それぞれの場合に合わせて、映画の好みの基準は変動すると考えられる。現時点では一人で見ることを想定していたが、こうした視点も踏まえて、演算における重み付けや係数の仕方について考えを深めたい。

また生涯映画鑑賞数について、答えにくいという意見がよく聞かれたことに関しても深刻に捉えている。

一般的な映画ファンとコアな映画好きの分類分けをする上で、高精度の演算を行うにはどのような種類のデータが必要なのか今一度考えたい。

パーソナライズの精度向上に関しても課題がある。本方式では、コアな映画好きは映画に対する偏愛度が高いとして、日米以外の製作国の映画や白黒映画などをあえて推薦する仕組みを取ったが、コアな映画好きの中にはそういった映画を好まない人がいることも予測される。ウェブサイトの質問事項を増やす等、検討していきたい。

最後に、検索結果の表示方法についてである。絵画による感性検索だけの結果も知りたいという意見があったため、今後はウェブサイトのシステムを一新させ、絵画検索だけの検索結果を表示したり、アンケート調査も踏まえた結果を表示したり、検索者の調査進行度合いに合わせてリコメンドされる映画がどのように変わるのかを可能な限りで可視化していきたい。

参 考 文 献

- [1] Ching-I Cheng , Damon Shing-Min Liu, “An Intelligent Clothes Search System Based On Fashion Styles”, pp.5, 2008
- [2] 川崎 彩子, “知的価値を有する観光対象を訪問する個人観光客のための価値観計量による観光地推薦方式”, 2018
- [3] 深澤 宏樹, “多次元空間計量によるユーザ・コンテキスト対応型アトラクション選択・推薦システムの実現”, 2018
- [4] Kiyoki, Y. Kitagawa, T. and Hayama, T. (1994) “A metadatabase system for semantic image search by a mathematical model of meaning”, ACM SIGMOD Record , vol. 23, no. 4, 1994, pp.34-41.
- [5] 『CINEMA Handbook 2018』2018年4月20日発行, 株式会社 TSUTAYA
- [6] 『僕の好きな映画。』2018年11月20日発行, 株式会社マガジンハウス
- [7] 藤原 武弘, “映画の内容分析(1)* -映画評論家と映画雑誌読者の外国映画への嗜好の違い-”, 2011
- [8] [online] <https://www.imdb.com> (参照 2018-1-9)
- [9] [online] <https://filmmarks.com> (参照 2018-1-9)
- [10] David Frankel / Aline Brosh McKenna (screenplay), Lauren Weisberger (novel) (2006): The Devil Wears Prada [DVD], France, USA: Fox 2000 Pictures
- [11] [online]<http://news.livedoor.com/article/detail/13056452/>, (参照 2018-2-16)
- [12] 形態素解析ツールの品詞体系 [online] <http://www.unixuser.org/~euske/doc/postag/>, (参照 2018-2-16)
- [13] 三和 義秀, “読後感の個人差に性格が及ぼす影響に関する研究:感情による小説検索システム設計のための基礎研究”, 2014
- [14] [online]<https://www.artic.edu> (参照 2018-2-16)