

テキストマイニングによる レビューサイトの映画分析

立教大学 経営学部 山口和範研究室

足立 一磨 池田 奈々子 大澤 智美

鎌田 綾香 坂田 麻奈 柴田 俊介

永井 沙里 中野 皓司 枡 卓人 村瀬 彩花

目次

- ◆ 研究の目的と背景
 - ◆ 現状分析
 - ◆ 研究目的
- ◆ 分析方法とデータ
 - ◆ 4108件のコメントをテキストマイニング
- ◆ 映画業界への提案

映画業界の現状分析

- ◆ 年々減少する映画館の数と入場者数

➡ 推移をグラフから検証。

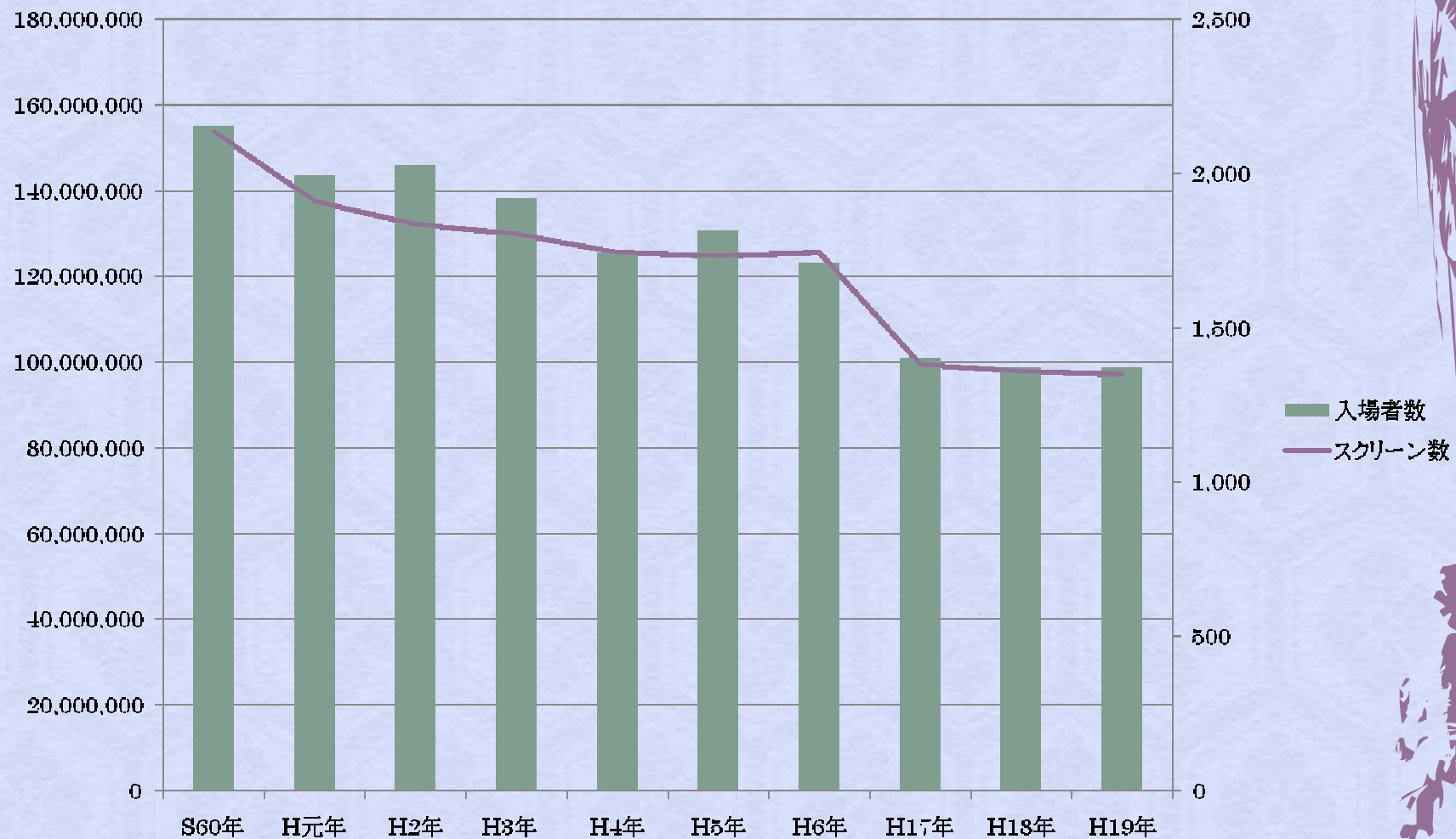
経済産業省

<http://www.meti.go.jp/statistics/>

全国生活衛生営業指導センター

<http://www.seiei.or.jp/advice/doukou1996/03.html>

S60～H19の 入場者数・スクリーン数推移



研究目的

- ◆ このような厳しい状況下で勝ち抜くには・・・



- ◆ 消費者が望む映画を作る



- ◆ 消費者どんな映画を好むのか？

歴代人気映画へのコメントから
人気が出る映画の要素を探る

分析方法

- 歴代(1980～2007)興行成績上位3位を各年から抽出。
- amazon、Google、YAHOO!のインターネット映画レビューサイトから視聴者の付けた得点・コメントを分析。
- 全28年度各上位3位(計84タイトル)をジャンル分けし、ジャンル別の人気指標を作る。
ジャンル:SF、アクション、ヒューマン、ドラマ、アニメ、ファンタジー、ミステリー、サスペンス、ホラー

使用したサイト & データ

◆ 調査対象サイト

- ◆ Amazon.co.jpホームページ <http://www.amazon.co.jp/>
- ◆ YAHOO!映画-映画情報 <http://movies.yahoo.co.jp/>
- ◆ 映画情報-goo映画 <http://movie.goo.ne.jp/>

◆ 入力データ

- ◆ 映画タイトル ex)スターウォーズ 帝国の逆襲
- ◆ 評価 1=100点 2=80点～100点未満 3=60点～80点未満 4=60点未満
- ◆ ジャンル ex)SF ファンタジー アクション etc.
- ◆ 日付
- ◆ レビューコメント(全文)

データの詳細

- ◆ 1980年～2007年 の興行収入TOP3の映画84タイトルを対象

参照サイト <http://www.eiga-ranking.com/>

- ◆ 84タイトルについての4108件のコメントを、下記サイトからランダムに抽出

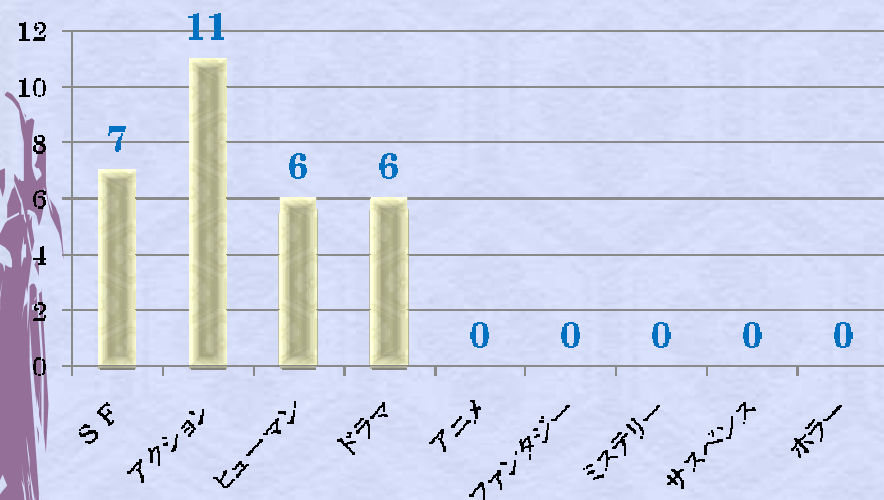
- ◆ Amazon.co.jpホームページ <http://www.amazon.co.jp/>

- ◆ YAHOO!映画-映画情報
<http://movies.yahoo.co.jp/>

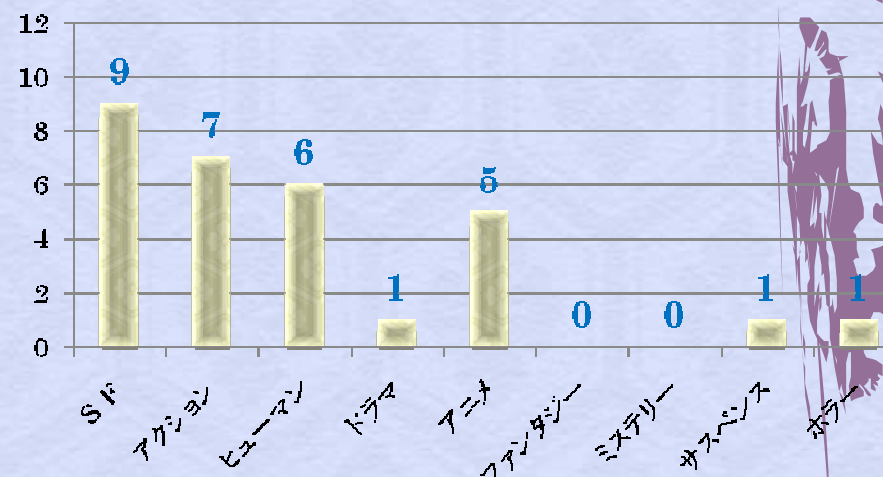
- ◆ 映画情報-goo映画
<http://movie.goo.ne.jp/>

年代別人気映画ジャンルの推移

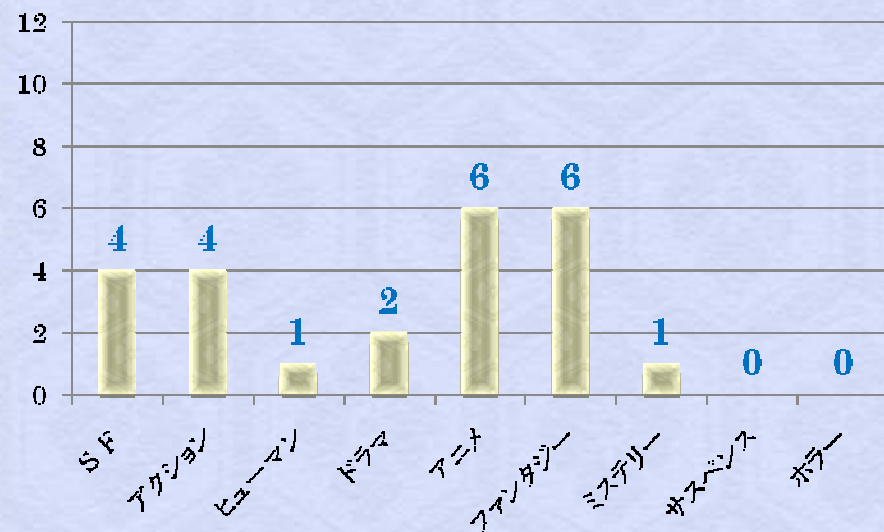
80年代



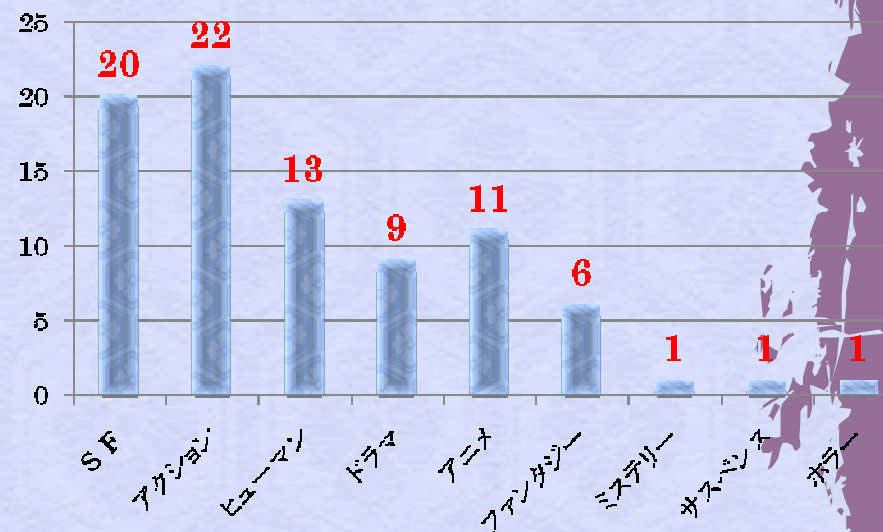
90年代



2000~2007



1980~2007



年代別人気映画ジャンルの推移

- ◆ 通年を通し、アクション・SFが人気である。
- ◆ 近年はアニメ・ファンタジーの人气が上昇。

＝ 近年の消費者の趣向

- ◆ さらにそのジャンルを見た人が
映画のどの様なところに着目しているか
レビュー内のキーワードから分析。

分析内容

- ◆ 単語頻度解析：
 - ◆ 上位頻出200単語の「評価」別頻度解析
 - ◆ 「評価」間の単語発生率のばらつき(標準偏差)が大きいキーワードを抽出。評価の高低によるキーワードの違いを探る
- ◆ 係り受け頻度解析：
 - ◆ 係り受けについても同様の分析を行う

頻度解析結果

	1	2	3	4	標準偏差
原作	5.1%	9.5%	12.2%	15.2%	0.043
心	13.3%	10.3%	6.6%	4.0%	0.041
ストーリー	12.8%	14.3%	14.4%	19.4%	0.029
涙	6.5%	4.2%	1.7%	1.0%	0.025
前作	2.6%	5.4%	8.1%	7.4%	0.025
音楽	9.0%	7.6%	4.9%	4.3%	0.022

- ◆ 青: 評価の低いレビューと結びついている単語
- ◆ 赤: 評価の高いレビューと結びついている単語

◆※評価 1=100点 2=80点～100点未満 3=60点～80点未満 4=60点未満

◆※標準偏差は、単語ごとに4つの比率から算出

係り受け頻度解析結果

]	係り先単語	1	2	3	4	標準偏差
シーン	多い	0.54%	0.37%	1.02%	1.90%	0.0096
原作	面白い	0.11%	0.28%	0.85%	1.38%	0.0090
テンポ	良い	1.03%	1.10%	1.02%	0.35%	0.0048
音楽	良い	1.19%	0.92%	0.00%	0.69%	0.0035
映像	美しい	0.38%	0.83%	0.51%	0.17%	0.0015

- ◆ 青: 評価の低いレビューと結びついている単語
- ◆ 赤: 評価の高いレビューと結びついている単語

分析結果まとめ

◆ 単語頻度解析結果

◆ 評価が低い場合

- ◆ 「原作」、「前作」

◆ 評価が高い場合

- ◆ 「心」、「涙」、「音楽」

◆ 係り受け頻度解析結果

◆ 評価が低い場合

- ◆ 「シーン—多い」、「原作—面白い」

◆ 評価が高い場合

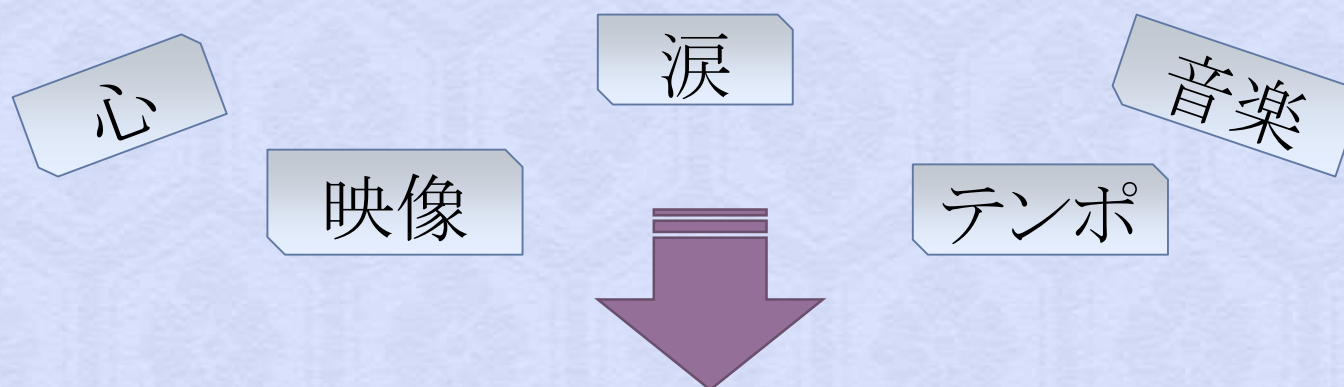
- ◆ 「テンポ—良い」、「音楽—良い」、「映像—美しい」

考察

- ◆ 原作を元にした作品ほど、原作と比較され批判的な評価が多くなる
 - ◆ ただし、今回の分析では、ハリーポッターやロードオブザリングの影響が強いため、上記の考察が導かれた可能性が高い
 - ◆ 原作によって話題性があるため見る前の期待値が高すぎるという外部要因も考えられる
- ◆ ストーリーだけでなく、視覚や聴覚など五感に訴えた映像効果を評価する人が多い

提案

- ◆ 1980年代以降不動の人気 ...SF・アクション
- ◆ 2000年以降人気急上昇 ...ファンタジー・アニメ



視覚や聴覚など五感に訴えたSF・アクションムービーを制作すれば売れるのではないか

「映画館でしか味わえない臨場感」を求めた映像効果でDVD鑑賞との差別化を図る