

映画の興行収入に関する統計的分析

2011SE139 近藤友樹

指導教員：木村美善

1 はじめに

私は、映画が好きだ。何百何千とある映画は、年代や気分、性別によっても映画一つ一つの観た時の感想は変わってくるものであろう。私は、沢山の映画を観る中で、どのような映画が一般的に評価され、劇場に足を運んでみたくなるのか興味を持った。

本研究では、広告費などの外的な側面からではなく、どういう特徴があるのかという内的な側面から研究を行った。作品のサンプルは、興行収入ランキング [1] を基に厳選した。映画の本質部分を検証する為に、Movie Walker[2] を基に考察を行った。尚、ある程度のジャンルに絞る為、本研究に用いるサンプルは邦画（アニメーションを除く）作品のみとした。

2 データ

作品サンプルとして、国内邦画興行収入ランキングより上位 200 作品（アニメーションを除く）を用いた。目的変数 y を興行収入として、説明変数は、あらすじやジャンルを基に映画の本質を問う要素 x_1, \dots, x_{16} とする。説明変数は、 x_1 : 上映時間 (短い), x_2 : 上映時間 (長い), x_3 : 主役 (子供), x_4 : 主役 (動物), x_5 : 原作, x_6 : シリーズ物/リメイク, x_7 : ノンフィクション, x_8 : 時間軸 (過去), x_9 : 時間軸 (未来), x_{10} : テーマ (アクション), x_{11} : テーマ (SF), x_{12} : テーマ (歴史), x_{13} : テーマ (サスペンス), x_{14} : メイン (カップル), x_{15} : メイン (複数人), x_{16} : 年齢制限とする。尚、数量化 III 類と判別分析の際には、興行収入を (高い/普通/低い) のカテゴリーに分けて分析を行う。

3 分析方法

本研究では、サンプルを全 200 作品、上位 100 作品、下位 100 作品の 3 つに分けて分析を行った。それぞれのサンプルに対して、重回帰分析、数量化 II 類、数量化 III 類、クラスター分析を行った。本論旨には、主に上位 100 作品の結果を記載し、まとめには全ての分析結果を総合した考察を記載する。

4 重回帰分析

最初の分析では、寄与率 0.278 自由度調整済み寄与率 0.139 となった。VIF、固有値を確認し、変数選択を行った結果、 x_1 (上映時間:短い), x_4 (主役:動物), x_5 (原作), x_6 (シリーズ物/リメイク), x_{11} (テーマ:SF), x_{13} (テーマ:サスペンス) という 6 個の変数が選択された。6 個の変数のみを用いて分析を行うと、寄与率 0.256 自由度調整済み寄与率 0.209 となり、改善された。

次に、てこ比とクックの距離を確認すると、てこ比から影響力の強いサンプルとして、2 番, 41 番, 50 番, 65 番

が挙げられた。クックの距離から影響力の強いサンプルとして、1 番, 2 番, 38 番が挙げられた。

外れ値としては、1 番と 3 番が挙げられた。ただし、3 番は、僅かに外れているサンプルであり、特異では無かった為、言及しないこととする。外れ値として検出された 1 番に関しては、他の作品と比べると非常に高い興行収入となっており、クックの距離からも影響力の強いサンプルとして検出されている為、1 番のサンプルを除いて再度分析を行う。分析結果としては、寄与率 0.254 自由度調整済み寄与率 0.205 となり改善されたことが分かる。ただし、外れ値を除く前よりも値は悪くなっている為、必ずしも全体のデータとして改善されたとは言い難い。故に、外れ値を外す前の結果を表 1 に示す。

表 1 重回帰分析結果

変数	係数	標準偏差	p 値
(Intercept)	45.272	4.796	3.13×10^{-15}
x_1	-11.554	5.355	0.034
x_4	13.979	9.092	0.128
x_5	-7.864	4.597	0.090
x_6	17.626	4.354	0.001
x_{11}	-7.590	4.919	0.126
x_{13}	-16.153	6.398	0.013

表を基に考察を行うと、主役が動物である事、シリーズ物/リメイク作品である事が興行収入にプラスに働き、上映時間が短い事、原作がある事、テーマが SF やサスペンスである事がマイナスに働いていることが分かった。 p 値を考慮して興行収入の高い作品の特徴をまとめると「上映時間の短くない前作のある作品」であると言える。

5 数量化 II 類

興行収入を 2 つのカテゴリー (高い/普通) に分け、数量化 II 類による分析を行う。外的基準を興行収入 y とし、説明変数を映画の本質を問う 16 個の要素 x_1, \dots, x_{16} を用いた。結果を表 2 に示す。

表 2 数量化 II 類による分析結果

興行収入	普通	高い	判別率
普通	20	9	69%
高い	21	50	70%

偏相関の値を考慮しつつ、要素の値を検証すると、上映時間 (短い), 時間軸 (過去), メイン (複数人) がプラスの値を示しており、シリーズ物/リメイク, ノンフィクション, 主役 (動物), テーマ (SF), 年齢制限がマイナスの値を

示している。上記の中で最も影響力の強い要素はシリーズ物/リメイクであるかどうかであり、最も影響力の弱い要素はテーマ (SF) であるかどうかであった。判別結果では、判別関数の値がプラスになると普通に分類され、マイナスになると高いに分類されている。故に、興行収入の高い作品はシリーズ物/リメイクであり、上映時間が比較的短くない作品だと考察出来る。

6 数量化 III 類

興行収入を (高い/普通) のカテゴリーに分け、10 個の変数を用いて数量化 III 類を行う。分析方法や考察を行うに当たり [3][4] を参照した。レンジの大きな変数を 3 つ選択し、それぞれのカテゴリーの中で最も値の大きな正負の要素から各軸を解釈すると次のようになる。

第 1 軸：

長くなく過去を舞台としない動きの少ない作品

第 2 軸：

見る人を選ぶフィクションの穏やかではない作品

第 3 軸：

未来を舞台としない派手な演出の恋愛要素のある作品

第 4 軸：

人が主役で未来を舞台にしない恋愛要素の少ない作品

7 クラスタ分析

数量化 III 類で得たサンプルスコアを用いて分析を行う。分析方法はウォード法を用いて、距離 20 で 4 群に分けた。左から順に第 1 群、第 2 群、第 3 群、第 4 群とすると次の結果を得る。

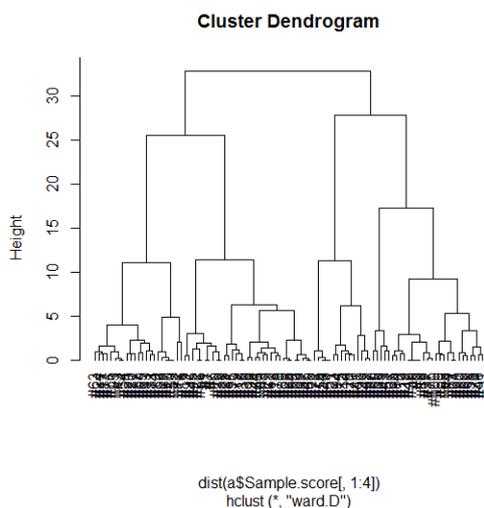


図 1 ウォード法

第 1 群：過去を舞台にした原作のある作品

第 2 群：現代を舞台にした歴史と SF 以外の作品

第 3 群：複数人にスポットを当てた SF 作品

第 4 群：現代を舞台にした原作のあるヒューマン作品

全ての群に共通しているのは、ノンフィクションでは無い事である。故に、ノンフィクションであるかどうかで群分けはされていないと判断し、言及しないこととした。数量化 III 類のカテゴリースコアによる考察と比べると、第 1 群と第 2 群からカテゴリースコアの第 1 軸と第 3 軸の特徴が見られる為、それぞれの得点が高い作品が多く含まれている事が考察出来る。

8 まとめ

3 つの重回帰分析の傾向から興行収入の高い作品には、上映時間が短くない事、メインがカップルである事、シリーズ物/リメイクものである事が特徴として見られた。上位 100 作品ではテーマがアクションである事も見られた。

3 つの数量化 II 類による分析の傾向からシリーズ物/リメイクものである作品は、200 作品の中でも特に興行収入が高いか逆に低い作品に多い傾向が見られる。上映時間の要素で判別すると、平均的な長さが高いに判別されやすいと考察出来る。また、主役が動物や子供である作品が下位に多く見られているのは、他の分析からも同様の結果を得ていることから信憑性が高いと言える。

全ての分析から興行収入に影響力が強い要素は、シリーズ物/リメイクであることが分かった。上映時間に関しては、平均的な長さの作品が興行収入が高い傾向が見られた。テーマに関しては、アクションや男女をメインとした作品の方が興行収入が高い傾向が見られた。故に、全体の分析結果からは、「前作のある丁度良い長さの男女をメインとしたアクション作品」が興行収入の高い映画の特徴であると考察出来る。

9 おわりに

本研究を通して、人によって感じ方や伝わってくるものが違う言葉で説明できないものを、色々な視点から見ることで傾向を掴む事が出来た。テーマや主役などが大きく関わってくるわけではないが、それぞれの質が重なることで内容からもある程度の傾向が見る事が出来た。今後、このような不明瞭なものを分析する際に有効な方法が更に充実、発展していくことを強く願う。

参考文献

[1] 映画ランキングドットコム

<http://www.eiga-ranking.com/> 2014 年 9 月

[2] 映画-Movie Walker

映画作品・映画館・上映時間・映画ランキング
<http://movie.walkerplus.com/> 2014 年 9 月

[3] 青木繁伸：おしゃべりな部屋

(プラネタリウム, 星, 植物, 熱帯魚, 統計学)
<http://aoki2.si.gunma-u.ac.jp> 2014 年 12 月

[4] シリウス先生の心理統計学

<http://homepage2.nifty.com/nandemoarchive/>
2014 年 12 月