

ブログの言及数とゲーム機の売上との相関について

2005MM051 西 義生

指導教員：田中 豊

1 はじめに

ゲームメーカーと消費者の双方にとって、ゲーム機の売上台数をより早く知ることは大きな関心事である。ある週の売上台数は、その2週間後の金曜日になれば知ることができるが、より早く知れることが望ましい。そこで、私はインターネットを通じて手軽に情報を得られるブログと検索サイトを使って、日本国内のゲーム機の売上台数を予測することを試みた。また、各ゲーム機の特徴を知るためにゲーム機毎のゲームソフトのジャンル別の売上本数を調査し、分析した。

2 データについて

ゲーム機の売上台数予測の対象とするゲーム機はDS、PSP、Wii、PS3、PS2、XBOX360の6機種とする。2006年12月から2008年8月までを調査期間とし、はてなダイアリー [1] における1日毎の言及数とGoogle Trends[2] における1日毎の検索数と週刊ファミ通 [3] が発表する月曜日から日曜日までの1週間毎の売上台数データを使って調査した。各ゲーム機のゲームソフトのジャンル別売上本数データは週刊ファミ通が調査した2006年と2007年それぞれの売上本数TOP500のデータを使った。ゲームソフトのジャンルはアクション、RPG、シミュレーション、スポーツ、アドベンチャー、その他の6種類に分けた。

3 分析方法

はてなダイアリーを用いた単回帰分析とGoogle Trendsを用いた単回帰分析と自己回帰分析と重回帰分析を行った。はてなダイアリーを用いた単回帰分析とGoogle Trendsを用いた単回帰分析ではより予測精度の高い手法を見つけるために、1週間の実際のゲーム機の売上台数に対して、言及数データもしくは検索数データの7日間、8日間、9日間、10日間、11日間の5つの範囲の分析方法についてそれぞれ3日前、2日前、1日前、ずらし無し、1日後、2日後、3日後の7通りのずらし方の分析方法を用いた。例えば、9日間と1日後の組み合わせの場合、ゲーム機の売上台数と今週の火曜日から来週の水曜日までの言及数を用いて、ゲーム機の今週の月曜日から日曜日までの売上台数との相関を調べることになる。この35通りの分析方法の中から、6機種のゲーム機の決定係数を平均した値が最も高い分析方法を各節における理想的な手法とする。また、カイ2乗検定による分割表の独立性の検定と対応分析を行ってゲーム機毎のゲームソフトのジャンルの特徴を調べた。

4 分析結果 (ゲーム機の売上台数について)

4.1 はてなダイアリーを用いた単回帰分析

全ての組み合わせについて、6機種のゲーム機の決定係数の平均を表1に示す。6機種の決定係数の平均が最も大き

かったのは10日間と1日後の組み合わせである。よって、この組み合わせが本節における理想的な手法である。この組み合わせの相関係数を表2に示す。DSとWiiは他のゲーム機と比べて決定係数が大きいことが分かる。DSとWiiは共に任天堂のゲーム機であり、なおかつ売上台数が多いという特徴がある。

表1 決定係数の平均

	7日間	8日間	9日間	10日間	11日間
3日前	0.1802	0.1838	0.1816	0.1929	0.2052
2日前	0.1807	0.1781	0.1907	0.2041	0.2105
1日前	0.1835	0.1973	0.2113	0.2179	0.2150
ずらし無し	0.1843	0.2004	0.2085	0.2060	0.2072
1日後	0.1870	0.1969	0.1951	0.1968	0.1923
2日後	0.1937	0.1905	0.1917	0.1864	0.1787
3日後	0.1898	0.1892	0.1823	0.1734	0.1800

表2 10日間と1日前での相関係数と決定係数

	DS	PSP	Wii	PS3	PS2	XBOX360
相関係数	0.5524	0.4430	0.5278	0.3884	0.4963	0.3606
決定係数	0.3051	0.1962	0.2786	0.1509	0.2463	0.1301

4.2 Google Trends を用いた単回帰分析

全ての組み合わせについて、6機種の決定係数の平均を表3に示す。6機種の決定係数の平均が最も大きかったのは8日間とずらし無しの組み合わせである。よって、この組み合わせが本節における理想的な手法である。この組み合わせの相関係数を表4に示す。はてなダイアリーの言及数での理想的な分析方法とは対照的な結果となった。Google Trendsの検索数ではPSPとPS2の決定係数がかなり高い。PSPとPS2はソニー株式会社のゲーム機である。はてなダイアリーの言及数での理想的な分析方法ではDSとWiiという任天堂株式会社のゲーム機が高い決定係数だったので、Google Trendsとはてなダイアリーではかなり結果が異なる。これはGoogle Trendsとはてなダイアリーの性質の違いのせいだろう。Google Trendsとはてなダイアリーの性質の違いは主に2つある。1つ目の違いはGoogle Trendsは非常に幅広くインターネットユーザーに利用されているGoogleの検索数データをカウントしているのに対し、はてなダイアリーはGoogleと比べれば狭い範囲のインターネットユーザーにしか利用されていないことだ。2つ目の違いはGoogle Trendsの検索数の多さは情報の受信側の興味の高さによって左右されるのに対し、はてなダイアリーの言及数の多さは情報の発信側の興味の高さによって左右されていることだ。

4.3 自己回帰分析

自己回帰の相関係数と決定係数を表5に示す。自己回帰の決定係数の平均は、はてなダイアリーの言及数による

表 3 決定係数の平均

	7 日間	8 日間	9 日間	10 日間	11 日間
3 日前	0.4179	0.4278	0.4280	0.4313	0.4402
2 日前	0.4348	0.4338	0.4362	0.4450	0.4458
1 日前	0.4502	0.4514	0.4593	0.4589	0.4574
ずらし無し	0.4636	0.4704	0.4686	0.4657	0.4615
1 日後	0.4686	0.4661	0.4623	0.4572	0.4510
2 日後	0.4532	0.4493	0.4440	0.4376	0.4283
3 日後	0.4333	0.4284	0.4224	0.4134	0.4056

表 4 8 日間とずらし無しでの相関係数と決定係数

	DS	PSP	Wii	PS3	PS2	XBOX 360
相関係数	0.5496	0.7908	0.5822	0.6915	0.8611	0.5801
決定係数	0.3021	0.6253	0.3389	0.4782	0.7415	0.3365

理想的な分析方法の決定係数の平均より大きく、Google Trends の検索数による理想的な分析方法の決定係数の平均と同じくらいである。

表 5 自己回帰の相関係数と決定係数

	DS	PSP	Wii	PS3	PS2	XBOX 360
相関係数	0.7314	0.6490	0.7101	0.6313	0.7191	0.5486
決定係数	0.5349	0.4212	0.5042	0.3985	0.5170	0.3009

4.4 重回帰分析

「自己回帰」と「はてなダイアリーの 10 日間と 1 日前の組み合わせ」と「Google Trends の 8 日間とずらし無しの組み合わせ」の 3 変数を説明変数として、重回帰分析を行った。その相関係数と決定係数を表 6 に示す。PS2 の決定係数は 0.7033 とかなり高いが、XBOX360 の決定係数は 0.3818 と低い。このことからゲーム機によって決定係数がかなり違うことが分かる。決定係数の平均は 0.5744 と高い。

5 分析結果 (ゲームソフトのジャンルについて)

5.1 カイ 2 乗分析

ゲーム機の種類とゲームソフトのジャンルを 2 変数に取って、表 7 にある分割表を作った。この分割表に対して、自由度 20 のカイ 2 乗検定を行った。カイ 2 乗検定を行った結果、カイ 2 乗値は 35655481 となった。カイ 2 乗値が 45.31 を超えているので、有意水準 0.1% で分割表は独立であるという帰無仮説は棄却される。

5.2 対応分析

各ゲーム機毎のゲームソフトのジャンル別に対応分析を行ったところ、第 2 固有値までの累積寄与率が 93.68% となるので第 1 固有値と第 2 固有値に対応する得点のみを分析する。第 1 軸と第 2 軸の得点の散布図は図 1 にあるような結果になった。ジャンルの配置から、この図の意味するところを考えた。この図は右端にその他があり、左端にスポーツがあることから分かるように、左方向に行くほど勝敗がはっきりするジャンルになっている。また、上端にスポーツがあり、下端にシミュレーションと RPG

表 6 3 変数での重回帰分析の相関係数と決定係数

	DS	PSP	Wii	PS3	PS2	XBOX 360
相関係数	0.7806	0.8066	0.7578	0.7262	0.8386	0.6179
決定係数	0.6094	0.6505	0.5742	0.5273	0.7033	0.3818

表 7 ゲームソフトの売上台数

	DS	PSP	Wii	PS3	PS2
アクション	14042238	5573766	4030217	1920884	11518943
RPG	15549119	2101418	203683	87817	8716737
シミュレーション	2238187	1100987	237919	141767	2376146
スポーツ	1672595	998851	3374711	714709	5400465
アドベンチャー	4285473	345741	137460	154410	2027191
その他	37087485	1575299	5177607	38305	3229139

があるので、下方向に行くほど頭を使うジャンルで上方向に行くほど体を使うジャンルになっている。

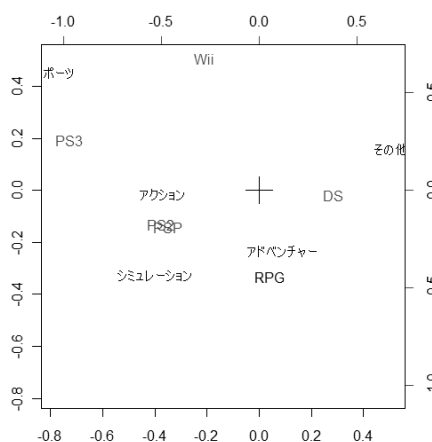


図 1 対応分析の散布図

6 おわりに

ゲーム機の売上台数の予測に最も適している分析方法は 3 変数での重回帰分析であった。一方で、はてなダイアリーの言及数による単回帰分析はゲーム機の売上台数の予測にはあまり適していなかった。また、自己回帰分析では DS と Wii の売上台数の予測の精度が高いが、Google Trends の検索数による単回帰分析と 3 変数での重回帰分析では PSP と PS2 の売上台数の予測の精度が高いことが分かった。予測に最も適している 3 変数での重回帰分析でも XBOX360 の決定係数は 0.38 程度であるため、売上台数の予測は難しい。しかし、XBOX360 を除く 5 つのゲーム機では中程度の精度なら売上台数の予測は可能だ。

参考文献

- [1] はてなキーワード : <http://d.hatena.ne.jp/keyword/>
- [2] Google Trends : <http://www.google.co.jp/trends>
- [3] TV ゲームニュース - ファミ通.com <http://www.famitsu.com/game/>