

デジタルゲームレビュー 評価尺度の提案

佐藤宏輝[†] 菱沼千明^{††}

Web上に投稿されたレビューは、その評価尺度はばらばらであり、客観性に乏しい。しかし、多数のレビューを集めて特徴づける複数の項目をもとに分類すれば価値観を集約させた客観性のある評価尺度を見出すことができる。本文では、デジタルゲームを対象に本手法を適用し、その有効性を示す。また、デジタルゲームの企業別、ジャンル別の特徴も見出すことができた。

An objective calibration for digital game reviews

Hiroki Satou[†] and Chiaki Hishinuma^{††}

The reviews on the web site are useful for evaluating the products and the services. However many reviews are written by different calibrations, therefore they lack in objectiveness. This paper presents the new classification method for the elements of reviews and shows the application results to the digital games. It is shown that this idea also applicable to the decision making of the suitable game selection.

1. まえがき

インターネットの普及により誰もが情報発信できる環境を生み、レビューの書き手が大量に生まれた。レビューを書くという行為自体は、読書感想文と同じであり、個人サイトのコンテンツとして多くの人々が取り入れやすいものであった。ネットショップのコンテンツとしても商業側の情報では無く、購入したユーザの率直な感想や評点を掲載できるレビューは有益なものであった。投稿されたレビューはユーザの購入判断材料に使用され、売れ行きを左右する力を持つようになったことからレビューを分析し、評価やニーズを抽出する手法が研究されるようになり、語句や表現を統一する辞書の作成手法、語句や表現を分類する手法が研究されている[1]。これらの研究はWeb上に公開されているレビューから評価情報を抽出することを目的していた[2]。

本研究ではユーザが製品評価を行う際にユーザの背景となる価値観が独自に考慮されている点に着目し、多くの事前に公開されているエンターテインメント分野のレビュー(以下レビュー)を分析することで価値観を形成する要素を明らかにし、要素を分類することで様々な価値観を集約させた評価尺度を見出す手法について提案する。

2. 評価尺度の提案

2.1 極大区分と評価関数

一般的に、ユーザのレビュー投稿内容は、個人の主観に基づいており、読み手はそのことをある程度理解した上で読み、場合によっては他人のレビューと比較して自らの評価を下している。本文におけるレビュー評価法はこの考え方に基づいており、大量に投稿されたレビューからその内容を特徴付ける複数の語句を中心に関連する語句をグループ化する。さらに、各グループが持つ特徴を基準とし、これと比較することにより、評価を行う。以下では、一般性を論ずるために、レビューを意味のある語句の集まりで捉え、集合論で表現する。

集合 X のすべての要素 x_{ij} が集合 $S_i(i=1, \dots, n)$ に対して次式の関係があるとき、部分集合 S_i は $S = (\cup_{i=1}^n S_i)$ に関して極大(minimum spanning)であるという。また、 S は S_1, S_2, \dots, S_n に極大区分できるという。

$$S_i = \{x_{ij} \mid f(S_i, x_{ij}) \geq f(S_k, x_{kj}), \text{ for all } x_{ij} \in X, k \neq i\} \quad (1)$$

[†] 東京工科大学大学院 コンピュータサイエンス専攻
Graduate School of Computer Science, Tokyo University of Technology

^{††} 東京工科大学 コンピュータサイエンス学部
School of Computer Science, Tokyo University of Technology

ここで、評価関数 $f(S_i, x_{ij})$ は、集合 S_i の各要素 s_{kj} と x_{kj} の関係の強さを表し、各区分 S_i から見た x_{ij} の評価値を表す関数である。例えば、レビューに含まれるジャンル S_i に関するある話題 x_{ij} の数が 100 個というように計算することができる。問題によってはより適切な評価関数を与えることができるであろう。この評価値は、 S_i の各要素 s_{ij} から見た x_{ij} の評価値の総和に等しいとする。すなわち、

$$f(S_i, x_{ij}) \geq \sum_{s_{ij} \in S_i} f(s_{ij}, x_{kj}) \quad (2)$$

とする。

式(1)、式(2)を言い換えると、レビューが評価対象とする製品やサービスに関する代表的な複数の語句と、それらに強く関連する語句を区分として集め、それぞれの集合が持つ意味あるいは重要性を基準にして他の要素との関係の強さを区分ごとの評価値とする。評価方法は対象によって異なるので、一律に示すことは困難であるが、例えば、出現頻度や共起頻度などが挙げられる。

また、基準となる語句の選定、「極大区分」の生成を効率よく、問題最適に構築することも重要な研究課題であるが、ここでは省略する。

式(1)において、評価値がある一定値 L_i を超えたときに区分 S_i に属するとする場合には、次式で表わされる。

$$S_i = \{x_{ij} \mid f(S_i, x_{ij}) \geq L_i, \text{ for all } x_{ij} \in X, k \neq i\} \quad (3)$$

この場合には、同一の要素 x_{ij} が複数の区分に属することになり、評価値の総和は式(1)による場合に比べて大きい。

2.2 多レベル評価要素

一般的に、レビューで示される評価項目は、細部であったり、全体を対象としたり様々である。すなわち、対象とする要素の次元は必ずしも同じとは限らない。そこで、本モデルでは、異なる次元の要素が混じっても評価出来るように、 X が例えば下記のように部分集合 Y_k で表わされている場合、最下位のレベルまで探索せずに、指定されたレベルで評価する。

すなわち、レビュー X が次式のような形で与えられたときは、式(1)は次式に示されるとおり式(5)に替わり、要素 x_{ij} よりも上位のレベルにある Y_k に対して評価が行われる。

$$X = \{\dots, x_{ij}, \dots\} \cup Y_1 \cup \dots \cup Y_k \cup \dots \quad (4)$$

$$S_i = \{x_{ij} \text{ or } Y_k \mid f(S_i, x_{ij}) \text{ or } f(S_i, Y_k) \geq f(S_k, x_{kj}) \text{ or } f(S_i, Y_k) \dots\} \quad (5)$$

ただし、

$$f(S_i, Y_k) = \sum_{x_{kj} \in Y_k} f(S_i, x_{kj}) \quad (6)$$

である。

図1に式(1)から式(6)に基づいた分類、集約方法について概要を示す。

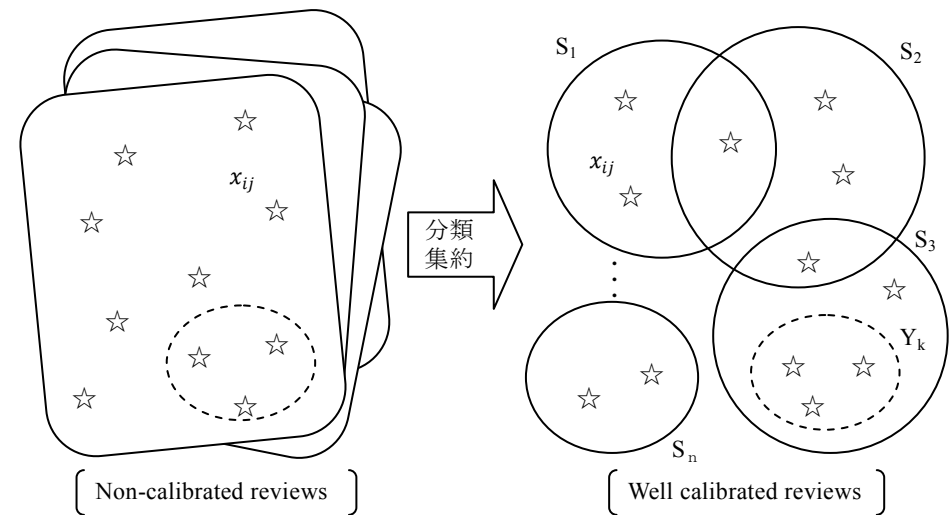


図1 本提案手法の概要

多レベルの評価要素によって比較することで、価値観の違いやユーザ像を明らかにする。比較するには上位集合からその部分集合を評価すればよい。例えばある製品群全体の「極大区分」を上位集合と生成し、その製品群のある観点(企業毎、カテゴリ毎)によって分けし、それらの「極大区分」を部分集合として生成する。この二つを比較することで上位集合から見た部分集合の特徴を見いだすことができる。

3. ゲームレビューへの適用

3.1 適用製品の選択

本研究ではエンターテインメント分野の製品について提案手法を適用する。この分野は評価者の主観に差が出やすいため、本手法を適用するのに最適と判断した。製品に

については Playstation2 向けのゲームソフトを選択した。ゲームはレビューを投稿するユーザーが多く、また Playstation2 向けのゲームソフトは全ハードで最もレビュー数が多い。投稿数の多さについてはネットショップ最大手 Amazon.com にてエンターテインメント分野のレビュー投稿数を調査した結果である。

3.2 ゲームレビューの収集

Web 上に投稿されているレビューを収集する。レビュー数が少ないと出現する要素が片寄るので、出来るだけ多くのレビューを収集することが望ましい。またレビュー内容の質が悪いサイトは避けて収集する必要がある。

収集したレビューは約 5 万件、文字数は約 3 0 0 0 万字、収集元は playstationmk2[3] という個人サイトを選択した。このサイトでは管理人やユーザにより掲載されているレビューが審査されており、不適切な内容が掲載されにくいのでレビューの質は高いと判断できる。レビューの量も個人サイトの中で掲載数が最も多い。

3.3 要素の選定

レビューに登場する語句から出現頻度とレビューの本文を参考にして要素を選定する。出現頻度とは文章にどれだけその語句が含まれているかを示しており、出現頻度が高いほど文章に登場する回数が多いので、要素になる可能性が高い語句である。

レビューより名詞のみを抽出し、出現頻度の昇順に整理する。出現頻度の高い名詞から順に要素として採用するかをレビューの本文を参考にして区分けした。すべての名詞を確認するのは困難なので、レビュー 1 0 0 0 件に 1 語句あることを最低条件とした。今回は 5 万件なので出現頻度 5 0 以上の語句から要素となる語句を選定した。

3.4 同義語処理

要素となる語句について同義語を探し出して、語句を統一する。同義語処理をする要素は誤字と判断される語句、同じ意味を持つ語句など一般的な同義語、その他には広義の意味で同義語としては扱われるものでも、レビューの本文を確認することで、同じ要素と判断される語句についても同義語処理を行う対象とする。

3.5 評価項目の作成

選出された要素を集合させることによって評価項目を作成する。レビューで示される評価項目は細部であったり、全体を対象としたり様々であるので、本研究で示す「極大区分」における評価項目は要素の集合体における表現例として、一番下の階層はすでに要素を集合させて 2 階層による評価項目で表現している。

第 1 階層は大評価項目(以下大項目)、その「大項目」の部分集合にあたる第 2 階層を小評価項目(以下小項目)とする。今回作成された評価項目は「小項目」2 5 個、「大項目」6 個によって構成されている。

要素の集合には、要素同士の共起頻度、レビューの本文を用いた。共起頻度は文章中で近くに記述されている語句数を調べた数値であり、数値が高いほど関係性の高い語句であり、部分集合と上位集合の関係が形成されやすいと判断した。また要素の選

定と同様にレビューの本文を調べることで裏付けを取る。

3.6 評価尺度の集約

ゲームレビューにおける評価の要素、「小項目」、「大項目」が決定したので、レビューに出現する要素を集約できる条件が整った。収集した 5 万件のレビューに出現する全要素を「極大区分」として全て集約させることで、上位集合にあたる「極大区分」が生成される。次に 5 万件をある観点から部分集合させれば、その観点における部分集合の「極大区分」が生成され、比較評価を行うことができる。今回部分集合させるのはジャンル毎と企業毎である。

3.7 ジャンルの区分け

5 万件のレビューをジャンル毎に区分けする。Playstationmk2 で定義されているジャンルの中でレビュー投稿数が多いものを採用する。採用したジャンルは二つのジャンルが統合されている(アクションロールプレイングなど)、外部的要素でジャンル付けされているもの、(スポーツなど)を除外した上位 5 ジャンルである。対象ジャンルはロールプレイングゲーム(以下 RPG)、アクションゲーム(以下 ACT)、シミュレーションゲーム(以下 SLG)、アドベンチャーゲーム(以下 ADV)対戦格闘ゲーム(以下 BACT)の 5 種類となった。

3.8 企業の区分け

5 万件のレビューを企業毎に区分けする。企業の選定基準としては Playstation2 向けのソフトでミリオンセラーを記録したソフトを発売した企業とした。選出された企業はカプコン、コーエー、コナミデジタルエンタテインメント(以下コナミ)、スクウェアエニックス(以下スクエニ)、ソニーコンピュータエンタテインメント(以下ソニー)の 5 社となった。

4. 極大区分の評価

4.1 使用する評価関数

今回の評価で使用する評価関数は相関率である。今回の評価では要素の関係の強さを相関関係の強さと同義であると仮定して、評価値を算出し、評価するものとする。

4.2 評価値の計算

要素の集合体である「極大区分」から見た、「大項目」、「小項目」の相関率を計算し、それを評価値とする。以下に相関率の計算方法を論ずる。

レビューの本文一つに対してスコア(重み)を割り振っていく。スコアの割り振り方としては、要素が文章に登場した時に、登場した要素を集合させている「小項目」と「大項目」にスコアを 1 割り振り振り、また要素の集合体である「極大区分」にもスコアを 1 割り振る。また複数回登場する場合も同様にスコアを 1 加算するものとする。

この作業を全レビューに対して行うことでレビュー毎に「極大区分」のスコア数、

「大項目」のスコア数、「小項目」のスコア数が求まる。「極大区分」のスコア数と「大項目」のスコア数について相関関係を計算することで「大項目」の評価値が求まる。同様に「極大区分」のスコア数と「小項目」スコア数について相関関係を計算することで「小項目」の評価値が求まる。

評価値の計算について、5万件のレビューを集約させた「極大区分」に行うことによって上位集合の評価値が基準値として与えられる。次に5万件のレビューをジャンル毎、企業毎に区分けし、区分けしたレビューを集約させた「極大区分」に同様の計算を行い、評価値を求める。この5万件のレビューにおける評価値と、ジャンル毎、企業毎における評価値の数値を比較することによって、ジャンル毎、企業毎のユーザの評価尺度が明らかになり、またその裏にある価値観やユーザ像が見えてくる

4.3 ジャンルによる評価尺度の比較

評価値を比較することで本手法の分類法、評価項目に付与された評価値の正確性を評価する。この評価検証のためにゲーム全体と5つのジャンルについて「極大区分」を作成した。以下に生成された「極大区分」の評価項目の内容を示す。

【登場人物】

登場するキャラクターについての「大項目」である。単純に「キャラクター」という語句を使って全体を評価している場合が多い。キャラクターの外見については「デザイン」、キャラクターの性格については「内面性」などの語句が用いられている。

【物語性】

物語部分についての「大項目」である。単純に「ストーリー」と評価している場合や、「演出」といったワンシーンについて評価していることもある。また「世界観」など舞台設定について評価していることもある。

【ゲーム性】

ルールについての「大項目」である。ゲーム毎に様々なルールが設けられるので評価項目が多い。まず「システム」はルールそのものを差して全体を評価している。「アクション性」、「思考性」、「育成」はゲーム毎に設けられたルールの特色を評価したものである。「バランス」はルールがうまく機能しているかの評価、「自由度」はルールの汎用性について評価している。「対人」は数人で対戦した時について評価している。

【映像性】

映像についての「大項目」である。単純に「グラフィックス」として全体を評価している場合が多い。「デザイン」は映像の美しさでは無く、ユーザの映像に対する好みを現しており、その好みについては3D技術を用いたリアルな映像を「リアル」ドット絵やデフォルメした映像を用いる場合は「アニメ」と評価している。

【音楽性】

音楽についての「大項目」である。単純に「音楽」として全体を評価している場合が多い。「BGM」は声ははいていないバックミュージックの評価であり、「歌」は声

が入っている曲についての評価である。

【ユーザビリティ】

快適に遊べるかを評価した「大項目」である。快適性について全体的な印象を評価したのが「快適さ」である。細部の評価は誤動作について評価した「バグ」。読み込みによって待たされた時間を評価した「ローディング」。ゲームのルールや世界観などの理解しやすさを評価した「わかりやすさ」。コントローラのレスポンスを評価した「操作性」キャラクターの視点など画面に関する評価として「カメラワーク」などがある。

これらの評価項目についてジャンル毎に評価値を比較したものを「大項目」について図2に、「小項目」について図3に示す。

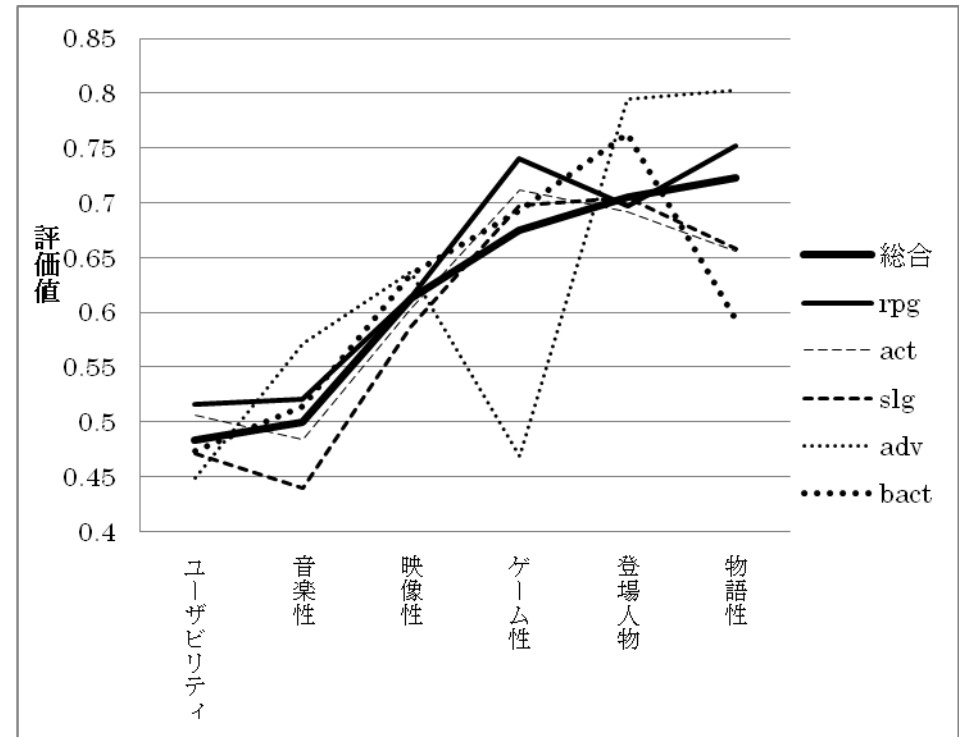


図2 ジャンル別大項目の評価値

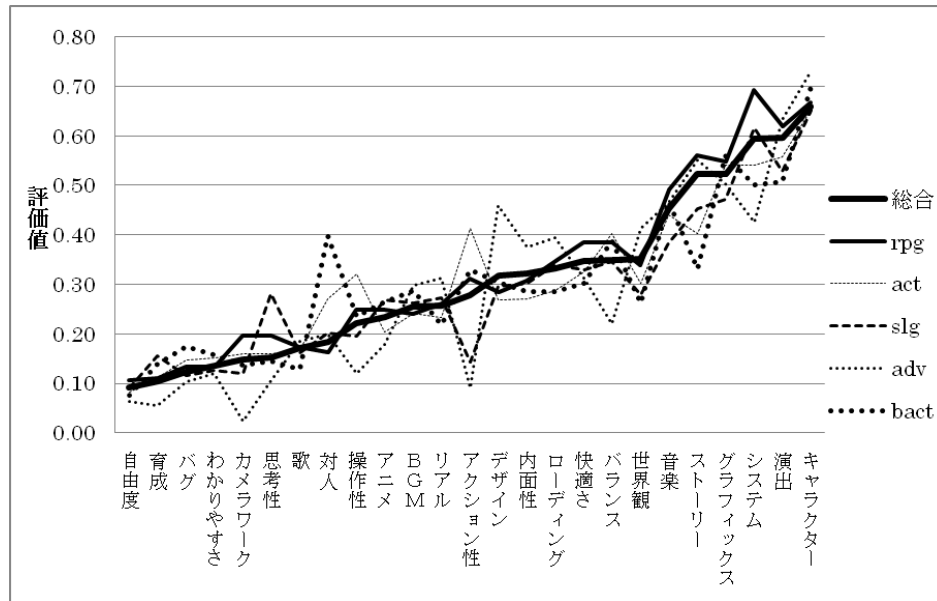


図 3 ジャンル別小項目の評価値

図 2 と図 3 は評価項目について適用した評価値の変化をジャンル毎に調べたものである。評価項目はゲーム全体の特徴として良く知られている語句である。図 2 では「物語性」が最も評価値が高く、図 3 では「キャラクター」が最も評価値が高かった。

図 2 において最も評価値の変化が見られるのは「ゲーム性」であった。RPG が最も高く、ADV が最も低い。また ADV 以外のジャンルにおいては RPG とさほど開きがない。これは ADV というジャンルが物語を楽しむことに特化しているジャンルであり、ゲーム性を求めているユーザが少ないという事象と評価値の値が連動していると言える。また ADV の物語性に着目すると最も大きい数値を示している、このことも裏付けになると言える。

図 3 において最も評価値の変化が見られたのはアクション性であった。ACT が最も高く、ADV が最も低い。ACT はその名の通りアクション性を重視しているので当然ユーザもアクション性を求めている、その事象が一致していると言える。また ADV は先ほど述べたように物語を楽しむことを特化している、アクション性を求めているユーザは低く、評価値も低い値を示している。

このように「小項目」で捉えたり「大項目」で捉えたりするところで、異なるレベルにおける評価を行うことが可能となっているので、本手法の分類法はうまく機能する

と言える。また評価値の数値変化も現実世界の事象と一致しており、相関率によって計算された評価値の変化によってジャンル毎の特徴について見出すことに成功している、ジャンル毎の評価値の比較について成功したと言える。

4.4 企業による極大区分の比較

ゲーム全体と 5 つの企業について評価値を比較する。企業は抱えているユーザ像が認知されているので、そのユーザ像と「極大区分」における評価値を比較した結果を比較することでユーザ像が抽出できるかを検証する。この 5 社の「大項目」の評価値を図 3 に、「小項目」の評価値を図 4 に示す。

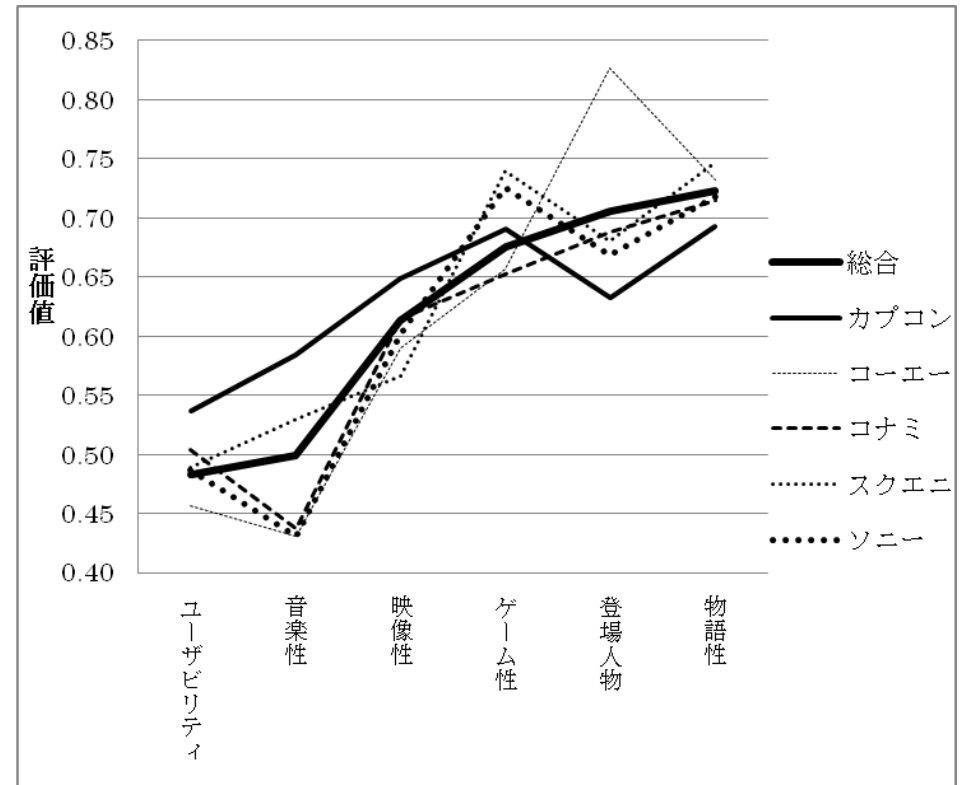


図 4 企業別大項目の評価値

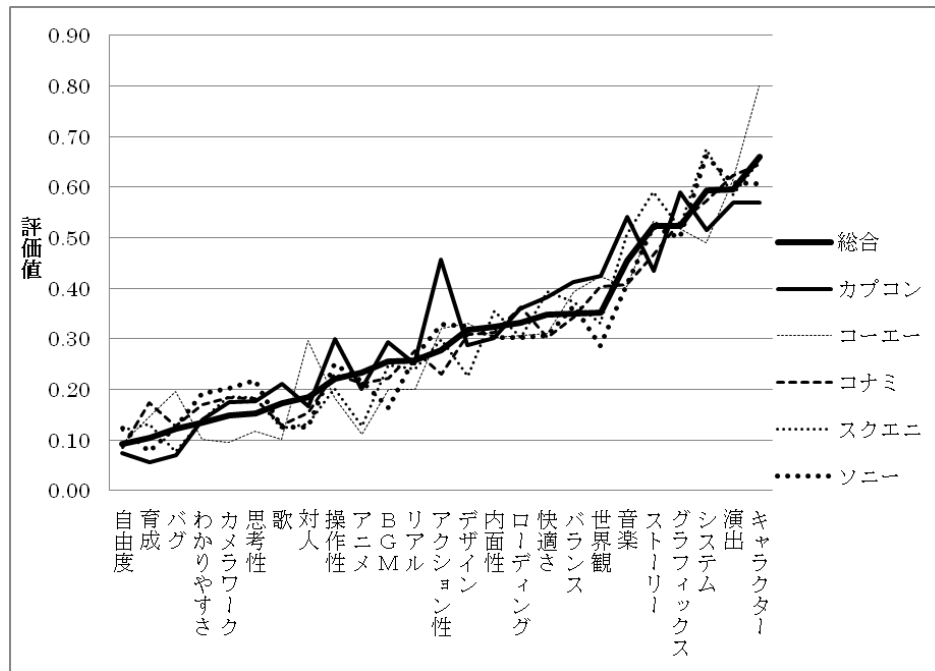


図 5 企業別小項目の評価値

【カプコン】

会社の方向性としては ACT を主軸としている。図 4 においてはその特徴について見出すことができないが、図 5 では「ゲーム性」の部分集合である「アクション性」について高い数値を示しており、これによって特徴を見いだすことに成功している。

【コーエー】

会社の方向性としては SLG と ACT を主軸としている。この二つのゲームは違う性質を持っているので、お互いが打ち消しあう形で価値観が表れにくい。評価値を見ると図 4 において「登場人物」の数値が高く、また図 5 においてその部分集合である「キャラクター」の数値も高いことが解った。この特徴については三国志、戦国時代などを舞台とし、実際の歴史上の人物を登場させていることからユーザの関心が高いのではないかと推測できる。

【スクエニ】

会社の方向性としては RPG を主軸としている。RPG は物語のみに特化したアドベンチャーゲームに次いで物語性を重視されるジャンルであり、またプレイ時間も長

くかかり、腰を据えてやることからゲーム性についても最も重視されている。図 4 において「物語性」、「ゲーム性」の数値が高く、図 5 においても「物語性」の部分集合である「ストーリー」、「ゲーム性」の部分集合である「システム」の数値が共に高いことから特徴を見い出すことに成功している。

【その他】

コナミとソニーについてはこれといったユーザ像が浮かび上がらなかった。コナミについては野球とサッカーをモチーフにしたゲームが多く、ソニーについてはゴルフと車をモチーフにしたゲームが多かった。このことから現実世界で体験できるものをモチーフとしたゲームは、現実世界の価値観がまじるため、ユーザ像が浮かび上がりにくいものであると推測できる。

むすび

本研究では「極大区分」によって要素を集合させ、個々の特徴を見い出す評価尺度について提案を行った。その結果「極大区分」のみで全体的に特徴把握するだけでは不十分であり、多レベルの要素は上位集合、部分集合に分け階層的に集合させることによって隠されている特徴を見出し、評価できることが解った。

参考文献

[1] 奥村学, 乾孝司「blog マイニングと評価分析」NLP2006 チュートリアル資料
 [2] 中村健二, 田中成典, 高橋亨輔, 寺口敏生, 田中裕一, 瓜谷芳樹, 関西大学大学院総合情報学研究科, 関西大学総合情報学部, CGMからの評価表現辞書の自動作成に関する基礎研究
 [3] Playstationmk2, <http://www.psmk2.net/>