

# レビュー文を考慮したゲーム推薦システムの実現に向けた 単語の類似度調整の取り組み

大山 浩暉<sup>1,a)</sup> 竹川 佳成<sup>1,b)</sup> 平田 圭二<sup>1,c)</sup>

**概要:** 本研究の目的は、ユーザの要望と合致したゲームの情報を推薦するシステムの設計と実装である。本システムは、システムとユーザがインタラクティブに質問応答を繰り返すことで、探索的にゲーム情報を抽出することができる検索エンジンである。本システムの実現には、ユーザの経験的価値が多く含まれているレビュー文を学習コーパスとした、言葉の足し引きによる意味処理とゲームのベクトル化表現が必要である。そのため本論文では、そのマイルストーンの1つとして、ゲーム辞書の作成と表記ゆれ対策を行い、それらを学習コーパスに適応する。そのコーパスを用いて Word2Vec の精度を確認する実験を行った。実験結果は適合率が高いジャンルと低いジャンルが同程度存在していた。今後は適合率を高めるため、ジャンルと単語の関連づけ、否定文の考慮を行う予定である。

## Activities of Word Similarity Adjustment Towards Game Recommendation System Considering Review sentences

KOKI OYAMA<sup>1,a)</sup> YOSHINARI TAKEGAWA<sup>1,b)</sup> KEIJI HIRATA<sup>1,c)</sup>

**Abstract:** The purpose of this research is to design and implement a system that recommends game information that matches the user's request. This system is a search engine which allows the system and the user to interactively repeat question responses and extract game information exploratory. In order to realize this system, semantic processing by addition of words and vectorized expression of games are required, with review texts containing a lot of empirical value of users as learning corpus. Therefore, in this paper, as one of the milestones, we prepare game dictionaries and countermeasures against notation wording, and adapt them to learning corpus. Experiments were conducted to confirm the accuracy of Word2Vec using the corpus. Experimental results showed that genres with high relevance rates and low genres were similar. In the future, we plan to associate genres with words and negate sentences in order to increase the relevance rate.

### 1. はじめに

ゲーム市場は年々広がっており、ジャンルの種類、機種の違いなど多種多様に存在する。そのため、ゲームに関する情報はインターネット上で膨大となっており、その中からユーザにとって最適な情報を自ら探索するのは極めて難しい。一般的な情報検索手段である Web でのキーワード検索では、概略の情報や一般的な知識などといった期待し

ない不要なページを多く含んでしまう。そのため、「アクションゲームがプレイしたい」といった曖昧な検索では、ユーザの要望と合致したゲームを得ることは困難である。例えば、謎解き要素を含んだアクションゲームをプレイしたい場合、ユーザは謎解き要素とアクション要素を同時に包含するゲームの情報を求めている。このような場合、ユーザは「謎解き AND アクション」とキーワード検索することが一般的であろう。しかし、実際に既存の検索エンジンで検索を行うと、表1の結果となる。表1の合致の有無とは、検索結果が「謎解き」と「アクション」を包含するゲームであり、そのゲームの概要が存在するか否かを示す。有は示す、無は示していないとなる。表1から、10件

<sup>1</sup> 公立はこだて未来大学  
Future University Hakodate

a) g2117010@fun.ac.jp

b) yoshi@fun.ac.jp

c) hirata@fun.ac.jp

中3件しか条件を満たしていない。このことから、謎解きとアクション要素を網羅したゲーム情報の出力は少なく、ユーザの要望に適した情報を見つけるのは困難であると思われる。

上記の問題を解決するため、本研究ではユーザの要望は言葉の足し引きでの表現が可能であることに着目した。言葉の足し引きによりユーザの要否が直接的に伝わるため、言葉での表現が困難であるゲームを簡潔に表現することが可能である。そこで本研究では、ユーザの要望と合致したゲームの情報を推薦するシステムの設計と実装を目的とする。提案システムを図1に示す。(1)では、ユーザの要望(自然文)をゲーム推薦システムに投げかける。前処理として、ゲームのレビュー文を学習させた Word2vec[1] にユーザの要望を入力する。(2)では、Word2vec で得た入力と最も類似度が高いゲーム名を出力し、ユーザに提示する。(3)では、その推薦がユーザにとって合致したものかを問い、合致しない場合、(4)で要望の再入力を行う。このように、ユーザと推薦システムがインタラクティブに質問応答を繰り返し、ユーザにとって、より良いゲーム推薦を目指す。

本論文では、上記の目的を達成するための1つのマイルストーンとして、Word2vec を用いたユーザの要望に対する言葉の足し引きによる意味処理の実現を目指す。例えば「謎解き要素を含んだアクションゲームをプレイしたい」という要望は「謎解き + アクション」と表現できる。また「ゼルダの伝説からアクション要素を無くして、ストーリー性があるゲームをプレイしたい」という要望は「ゼルダの伝説 - アクション + ストーリー性」と表現できる。このような表現によって、ユーザの要望を直接的に捉えることができる。本論文では、ゲームプレイヤーの経験的価値が多く含まれているレビュー文を学習コーパスとした Word2vec を用いた検索システムのプロトタイピングと、その性能評価のための実験結果について議論する。本実験では、言葉の足し引きを行う際、最も多く扱われるであろう要素、ジャンルに焦点を当て、適合率、要望とゲーム間の類似度を調査した。

本論文の構成を以下に示す。次章では、Word2Vec やゲーム推薦に関する関連研究を示す。3章では、本研究で作成したプロトタイプシステムの実装方法について述べる。4章では、実験結果とその分析について述べ、5章でその考察を述べる。6章では、本稿で得られた成果と課題をまとめる。

## 2. 関連研究

Word2Vec を用いた研究は様々存在する [2], [3]. その中に、野沢らの代替え食材の発見手法の提案の研究がある [4]. 近年、レシピ投稿検索サイトに大量の料理レシピが投稿されており、その中には、一部の食材を入れ替えた類似レ

表1 「謎解き AND アクション」の Yahoo! JAPAN 検索結果  
Table 1 Yahoo! JAPAN research result of "solve mystery AND action"

検索結果	合致の有無
『ゼルダの伝説』みたいな謎解きがあるゲーム...	無
謎解きアクションゲームに関する Q&A ...	無
64 ゼルダみたいなアクションと+謎解きが...	無
恋人に会いに行く謎解きアクション Pretentious...	有
PS4: 気分は探偵! 推理& 謎解きがあるゲーム...	有
謎解きアクション好きにオススメの『ロゼと...	有
【厳選】未だ遊んでないかもしれない PS3 の...	無
アクションアドベンチャーゲーム - Wikipedia...	無
PS4 おすすめゲームソフトランキング!? RPG、...	無
謎解きゲーム—無料ゲーム/フリーゲームの...	無

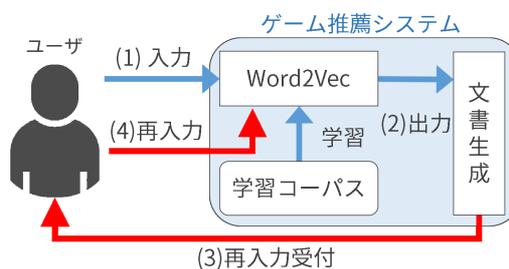


図1 提案システムの概要

Fig. 1 Summary of proposal system

シピが数多くある。ユーザが要望するレシピを見つけ出したとしても、記載されている食材をそろえることができない、あるいは、必要な調理器具を所持していないなどの場合は、調理することができない。野沢らの研究では、大量のレシピデータの中から、特定の食材に対する代替可能な食材を発見する手法を提案する。調理手順の言語モデルを4つ作成し、それらのコーパスを用いて Word2Vec で語彙空間を学習する。それで得られた結果と正解データを比較する実験、考察を行った。その結果として、代替食材を発見できる精度は、モデルの複雑さには、ほとんど影響されず、食材や調理法の表記ゆれを解決することの重要性が明らかとなった。本論文では、表記ゆれも考慮した上でゲームの類似度を分析し、さらに向上する方法を検討する。

推薦に関する研究は様々存在する [5], [6], [7]. その中に、中谷らのユーザの経験的価値を元にゲーム推薦を行う研究がある [8]. 中谷らは、個人向け電子商取引のウェブサイトでは、推薦システムが導入されているが、ゲームの重要な要素である経験的価値を考慮していないことを問題点として指摘している。ここでの経験的価値とは、ゲームをプレイした経験がある人にしか語りすることができない感想を意味する。中谷らは、経験的価値を感覚的知覚、創造的思考、身体性、社会性、衝動的感情の5つに分類した。その後、辞書を用いてレビューサイトからレビュー文を抽出し、テキストマイニングでゲームの分類を行った。そして、ゲー

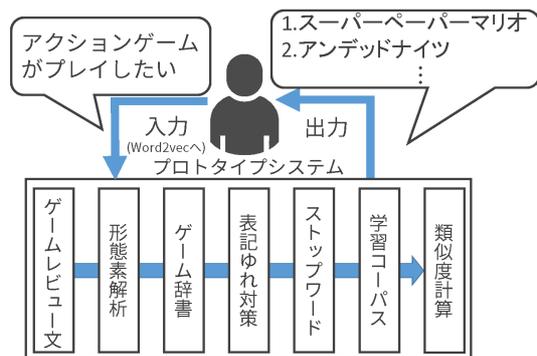


図 2 プロトタイプシステムの概要  
Fig. 2 Summary of protptype system

ムの経験を定量化し経験値へと変換させ、個々のユーザの経験に適するゲームを推薦するシステムを提案した。本研究も同様に、ゲームプレイヤーの経験的価値としてゲームレビューサイトのレビュー文を用いる。しかし、本研究では言葉の足し引きに着目した Word2vec を用いてゲームを定量化し推薦を行う。

### 3. プロトタイプシステムの詳細

本研究で実装したプロトタイプシステムの概要を図 2 に示す。学習コーパスには、ゲームレビューサイトを用い、ゲーム辞書の作成や表記ゆれに対処した上で Word2vec に学習させ、入力データとゲーム間の類似度が高いゲーム名を出力する。提案システムの開発手順は、以下の通りである；(1) ゲームレビューサイト mk2 におけるレビュー文の抽出、(2) MeCab を用いた形態素解析、(3) ゲーム辞書の作成と表記ゆれへの対処、(4) Word2vec を用いた類似度計算。

まず、(1) のレビュー文の抽出では、ゲームレビューサイト mk2<sup>\*1</sup> を採用した。本サイトを採用した理由は、他のレビュー文が記述されているサイトと比較した際に、ゲームそのものに対する記述が多く盛り込まれているなど、1 つのレビューの情報の質が高くデータ量が多いためである。ゲーム推薦におけるゲームの経験的価値は考慮する必要があり [8]、さらに、一般的にゲーム制作者とゲームプレイヤーはゲームに対する捉え方が異なる。そのため [9]、本研究ではゲーム制作会社が記述したゲーム概要を用いなかった。(2) に関して、MeCab<sup>\*2</sup> を用いて、レビュー文の形態素解を行う。ここで、ゲームの特徴を表している単語は名詞と形容詞に多く含まれると考えたため、名詞と形容詞のみを抽出する。(3) に関して、ゲーム辞書の作成は mk2 で取得した全てのゲーム名を記述した。レビュー文中にゲーム名が出現した場合、形態素解析による分割を防ぐため用いる。また、表記ゆれへの対処は重要なため [4]、ステミング、シソーラスに加え、ストップワードリストを適応した。

\*1 <http://switchmk2.net/>

\*2 MeCab: <http://taku910.github.io/mecab/>

表 2 ゲームジャンルとその定義  
Table 2 Game genres and definition

ID	ジャンル名	定義
(1)	シューティング	敵自身や敵の攻撃を避けながら射撃武器で敵的を狙い撃つ
(2)	アクション	ゲーム内でプレイヤーがキャラクタをコントローラで操作し、リアルタイムに起こる様々な課題や障害を攻略する
(3)	アドベンチャー	物語上に配置された謎を解く
(4)	ロールプレイング	課題や障害を解決した報酬としてプレイヤーの操作キャラクタを強化し、物語のエンディングを目指す
(5)	パズル	出題された問題を試行錯誤によって解く
(6)	レース	乗り物を操作し、他のプレイヤーやコンピュータと競争する
(7)	シミュレーション	プレイヤーがある役割となって戦略的に目標を達成するあるいは、自分の駒を動かして相手の駒を倒す
(8)	音楽	リズムや音楽に合わせてボタンを操作する

ストップワードリストの作成では、単語単体で意味を持たない単語とゲームの特徴を表していない単語を追加した。具体的には、記号や数字、さらには、ゲーム、クリア、ソフトなど全 697 単語を人出で作成し、追加した。最後に、(4) に関して、上記で作成した学習コーパスを Word2Vec に学習させる。Word2Vec は gensim<sup>\*3</sup> で公開されているものを用いた。予備実験の結果、次元数は 300、ウィンドウサイズは 5 に設定した。予備実験とは、あるゲームとそのゲームのシリーズの類似度は高いという仮定のもと、その類似度が最も高くなる設定を明確にする実験である。次元数は 100-800、ウィンドウサイズは 1-15 を予備実験対象とした。本システムの出力は、入力データとの類似度が高いゲーム名となっておりユーザに提示される。

### 4. 実験結果と分析

本章では、前章で述べたプロトタイプシステムの性能評価のための実験について述べる。本実験では、システムにおける入力と Word2Vec の出力結果をジャンルの判断基準と評価基準をもとに調査した。

ジャンルの判断基準は、著者を含めた 2 名が Wikipedia の当該ページ<sup>\*4</sup> を参考に定義した (表 2)。全 8 項目のジャンル数は、ほとんどの種類のゲームを網羅している。ロールプレイングはアドベンチャーゲームと比較し、キャラクタの成長要素が強いことが挙げられる。成長要素が強いは、操作キャラクタが成長によりレベル、体力、攻撃力といったステータスが上昇し、数値として存在するものである。ここで、ゲームは様々なジャンルの特性を所有しているため、1 つのジャンルのみに絞れるということは少ない。

\*3 <http://radimrehurek.com/gensim/models/word2vec.html>

\*4 <https://ja.wikipedia.org/wiki/コンピュータゲームのジャンル>

表 3 ゲームジャンルの適合率

Table 3 Precision of game genre

ジャンル ID	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
適合率	0.3	0.9	0.7	0.4	0.5	0.9	0.4	1.0

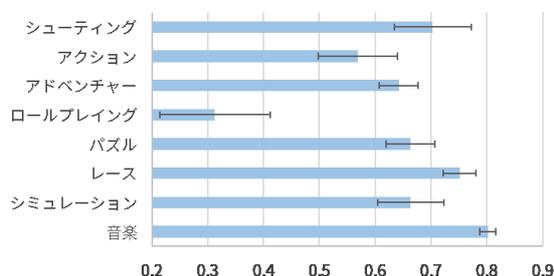


図 3 ゲームジャンルの類似度

Fig. 3 Similarity of game genre

入力と出力結果の評価基準は、著者を含めた 2 名が以下の基準に基づき判断した。

- ゲームが入力の要素を含む、あるいはゲームが入力の要素を一部含む
- ゲームが入力の要素をほとんど含まない、あるいはゲームが入力の要素を全く含まない

具体的な評価方法として、入力内容とコサイン類似度が高い出力結果上位 10 件を比較する。評価基準を用いて分類し、合致したゲームを正解とした適合率をジャンルごとに調査する。各ゲームジャンルに対する適合率の結果を表 3 に示す。ここで、ジャンル ID とは、表 2 で指定した各ジャンルに対応する ID を意味する。表 3 の結果から、全ジャンルにおける平均適合率は 0.64 であった。また、各ゲームジャンルの類似度を図 3 に示す。ジャンル ID は上記と同様である。

## 5. 考察

表 3 の実験結果から、プロトタイプシステムの出力では、ゲームジャンル間の適合率にばらつきがあり、特に、シューティング、ロールプレイング、パズル、シミュレーションは低い適合率（平均 0.64 以下）であることがわかった。また表 3 から、類似度にもばらつきがあり、特にロールプレイングは低い類似度（0.5 以下）であることがわかった。つまり、シューティング、シミュレーション、パズルは類似度が 0.5 を超えているにも関わらず、適合率は低い。この原因として、我々は (1) ジャンルの不明確さ、(2) 否定文の不考慮、(3) 予備実験と本実験の不一致、の 3 点に着目した。

(1) に関して、表 2 で定義したジャンルを比較すると、シューティングはアクション要素に加えて、避けながら射撃武器で狙撃つと捉えることができる。つまり、シューティングとアクションは包含関係にあると捉えることが

できる。実験から、シューティングで出力されたゲームのジャンルを調査すると、上位にアクションが多く含まれていることがわかった（表 4）。しかし、図 3 から、シューティングは類似度が比較的高い傾向にある。つまり、プロトタイプシステムにおいて、シューティングはアクションに包含されるジャンルと捉え、アクションと誤って判断されたと考えられる。適合率が低かった他ジャンル（ロールプレイング、パズル、シミュレーション）においても同様の調査を行った（表 5）。表 5 は各ジャンルで出力されたゲーム中で最も多く含まれていた不正解ジャンルを表している。この結果から、パズルはシューティングと同様に、アクションと判断されたと考えられる。シミュレーションは、アクション、パズルの適合率と類似度の関係に近似しているが、不正解ジャンルが分散していた。つまり、アクション、パズルとは逆に、シミュレーションは多ジャンルを包含するジャンルと捉えられ、不正解ジャンルが分散したと考えられる。ロールプレイングは適合率が低く、類似度も低いが、最も割合が高い不正解ジャンルはアクションであった。この原因は、ロールプレイングにはアクションロールプレイングといった、アクションとロールプレイングの要素を包含するゲームが存在する。そのため、アクション要素とロールプレイング要素がプロトタイプ内で近いものと判断され、アクションが不正解ジャンルとして多く混在したと考えられる。以上を解決するため、ジャンルに出現しやすい単語と対応するジャンルを強く関連付ける必要がある。例えば、Word2vec では Skip-gram モデルが用いられ、文中の単語と単語の距離が近いとそれらの類似度が高く学習される [10]。そこで、シューティングゲームに表れやすい単語「弾幕」や「スクロール」などを明瞭化し、それらの単語の近くに「シューティング」を配置することで強い関連付けをする。

(2) に関して、「自由度がない」や「アクション要素が皆無」といったレビュー文においても、自由度、アクション要素があるものとして学習コーパスに付与している。そのため、アドベンチャーらしい要素がないにも関わらずアドベンチャーゲームと認識されることや、パズルゲームではないにも関わらずアクションゲームと認識されると考えられる。解決策は、係り受け解析を行い、どの単語に否定がかかっているかを明瞭化する。その単語を学習コーパスから削除するなどの処理を施す。

(3) に関して、入力は予備実験ではゲーム名とし、本実験ではジャンル名とした。予備実験と本実験で入力内容が異なるため、ゲーム名ではうまくいったが、ジャンル名ではうまくいかなかった可能性がある。つまり、Word2vec のパラメータ設定が不確実である可能性がある。そのため、予備実験と本実験に一貫性がある実験にするべきである。

表 4 入力「シューティングゲーム」で出現したゲームジャンル

Table 4 Game genres appeared by input "shooting game"

出力項目	ジャンル名
1	アクション
2	アクション
3	横スクロールゆるキャラシューティングゲーム
4	アクション
5	3D アクション
6	アクション
7	シューティング
8	アクションアドベンチャー
9	パズル
10	サードパーソンシューター

表 5 不正解ジャンルとその割合

Table 5 Incorrect answer genres and their rate

ジャンル	不正解ジャンル (割合)
(1) シューティング	アクション (60%)
(4) ロールプレイング	アクション (40%)
(5) パズル	アクション (30%)
(7) シミュレーション	バラバラ (10%)

## 6. おわりに

本研究では、ゲームレビュー文を学習コーパスとした Word2Vec を用いたプロトタイプシステムを構築し、その出力結果と入力間の合致度と類似度について調査した。実験結果から、ゲームジャンル間で適合率に差があり、特に、シューティング、ロールプレイング、パズル、シミュレーションのジャンルに関して精度が低いことがわかった。また類似度においては、特に、ロールプレイングが低いことがわかった。この原因として、ゲームジャンルの不明確さ、学習コーパスにおける否定文の不考慮、予備実験と本実験の不一致が挙げられる。そのため、今後はゲームの各ジャンルに表れやすい単語とジャンル名間の関連を強くする工夫を学習コーパスに施す必要がある。さらに、係り受け解析からどの単語に否定語が関係しているかを明瞭化し、その単語を学習コーパスから削除するなどの処理を施す。さらに、予備実験と本実験を一貫性がある実験にし、上記の課題に対処する。

### 謝辞

本稿執筆にあたり、実験と評価手法に対する助言や、正解データの作成、論文の推敲方法に関して多くの指導をいただいた公立はこだて未来大学の三浦寛也氏、牧野孝史氏、庵愛氏、芳野洸太氏に心より感謝いたします。

## 参考文献

- [1] T.Mikolov, K. Chen, G. Corrado, and J. dean: Efficient Estimation of Word Representations Vector Space. In ICLRWorkshop (2013).
- [2] 松本和幸, 土屋誠司, 芋野美紗子, 吉田稔, 北研二: 感性を考慮した日本語俗語の標準語変換, 人工知能学会論文誌, vol.32, no.1, pp.1-12 (2017).
- [3] 名渡山夏子, 岡本一志: Word2Vec に基づく購買履歴からのアイテムベクトル学習, 知能と情報 (日本知能情報ファジィ学会誌), Vol.29, No.3, pp.579-585 (2017).
- [4] 野沢健人, 中岡義貴, 山本修平, 佐藤哲司: word2vec を用いた代替食材の発見手法の提案, 電子情報通信学会 データ工学研究会, Vol.114, No.204, pp.41-46 (2014).
- [5] 林貴宏, 尾内理紀夫: Web 上のレビューを利用した映画推薦システム, 人工知能学会論文誌 一般論文/特集論文「Web インテリジェンスとインタラクションの新展開」, Vol.30, No.1, pp.102-111 (2015).
- [6] 大久保和訓, 土方嘉徳: 情報推薦におけるユーザプロフィール提示とユーザの気づきに関する調査, 知能と情報 (日本知能情報ファジィ学会誌), Vol.25, No.1, pp.511-523 (2013).
- [7] 渡邊亮彦, 笹野遼平, 高村大也, 奥村学: Web ページ推薦システムにおけるユーザ指向型スニペット生成, 人工知能学会論文誌, Vol.31, No.5, pp.C-G41.1-12 (2016).
- [8] 中谷知博, 星野雄一, 経験的価値の分類に基づくゲーム推薦システム, 社会法人 情報処理学会 研究報告, 2008-EC-011, Vol.2008, No.129, pp.49-56 (2008).
- [9] Hua Wang, Cuihua Shen, Ute Ritterfeld: Enjoyment of Digital Games What Makes Them "Seriously" Fun?, Serious Games: Mechanisms and Effects, pp.25-47 (2009).
- [10] 西尾泰和: word2vec による自然言語処理, オライリー・ジャパン (2014).