

D-37 ジャンルごとの関係性とレビュー内容に基づくゲーム推薦

松谷隆成*, 馬場隆寛*

(久留米工業大学大学院工学研究科電子情報システム工学専攻)

1. はじめに

PCゲーム販売サイトSteamにおけるゲームタイトル数[1]は、2023年には2007本のリリースを行っている。消費者が、自分に合ったゲームを正確に選び出し、さらにその選ぶ時間を短縮出来れば、選ぶ時間の短縮と自分に合わなくてすぐにそのゲームをやめてしまい後悔する機会を減らせると考えた。そのような問題を解決する方法として情報推薦システムの利用が期待されている。情報推薦システムとは、情報収集を助けるアプローチ方法の一つであり、ユーザに対して情報を推薦する手法である。またユーザに合うゲームを正確に素早く知ることができる。

そこで、本研究では、ゲーム販売サイトから得たユーザの遊んだことのあるゲームのジャンルの情報とレビュー内容のテキスト情報を用いた情報推薦システムによって、ゲーム推薦を助けることを解決するシステムを開発し検証を行う。

2. 提案手法

ゲーム販売サイトのユーザのレビュー内容データを収集し、取り出したレビュー内容から2値分類を行う。次に、同じくゲーム販売サイトのユーザのプレイしているゲームタイトルのデータ収集を行い、取り出したユーザのプレイしているゲームのジャンルを抽出する。その上でユーザのプレイしているゲームのジャンルを分類し、ジャンルとレビュー内容の関係からゲームジャンルの組み合わせの推薦を行う。最後に、ジャンルの組み合わせからゲームタイトルの推薦を行う。

2.1 データセット

今回の研究では、Steamのレビュー内容のデータを用いる。レビューデータは、高評価したものと低評価したものをそれぞれ集める。表1は、収集したデータの一例である。これらに対して、おすすめできるものは1、おすすめしないものには、0としてアノテーションを行った。

表1：元データ例

コメント	評価
面白かった	おすすめ
すぐにあきた	おすすめしません

2.2 実験方法

レビュー文から高評価と低評価の2値分類による予測を行う。層化5分割交差検証を用いてモデルの性能の評価を行う。混同行列を用いて評価を行った結果、表2のような性能になった。

今回の実験では、混同行列から分かる、正解率、適合率、再現率、F値の4つの評価指標を用いる。正解率は、モデルが正しく予測できた割合を表すため、予測するのにどれだけの精度があるかを明確に示すことができる。他の適合率、再現率、F値を用いる理由は、適合率と再現

率は、トレードオフの関係にあるため、そのバランスをF値を用いることで、データの不足している要素を明らかにするためである。

表2：層化5分割交差検証による性能評価

正解率	適合率	再現率	F値
0.861	0.888	0.916	0.901

3. 実験結果

表2の結果から分かることは、正解率は非常に高い数値であり、F値が高く出ているので、適合率と再現率のバランスに優れ、予測性能もよいことが分かった。

4. 考察

実験の結果、レビューデータのみを用いた場合、提案手法によるモデルの正解率は0.861と高い、しかし、適合率が低いためのデータの偏りが考えられる。本研究では評価の高いゲーム上位10タイトルを対象としたが、これを拡張し、より多くのタイトルとレビューを含めることでモデルの汎化性能を向上させられると考える。また、特定のジャンルに偏ったデータ収集ではなく、幅広いジャンルをカバーするデータを収集することでより新規のデータに対しても精度を高く分類できると考える。

5. まとめ

今回の研究では、ゲームタイトルを推薦する新たなシステムについて提案した。Steamのレビューから高評価と低評価のレビューデータを収集し、モデルを作成し、評価を行った。モデルの性能は高かったが、適合率を高めることで、さらに精度を向上させられることも期待できる。今後は、レビューを行っているユーザが他にどのゲームをプレイしたことがあるかのデータと合わせてジャンルのデータを収集し、レビューとジャンルの情報を組み合わせてレビューから他のジャンルのゲームを推薦できるシステムの開発を行っていききたい。

参考文献

- [1] SteamDB . Steam game releases by year . <https://steamdb.info/stats/releases/?tagid=0>. 2024年8月7日アクセス.