

Steamにおけるタグ付与支援を目的としたゲームタグ推薦に向けて

玉置 文人[†] 梶本 真佑[†] 楠本 真二[†]

[†] 大阪大学大学院情報科学研究科

E-mail: [†]{f-tamaki,shinsuke}@ist.osaka-u.ac.jp

あらまし 世界中のPCゲームが公開されるプラットフォームとしてSteamがある。Steamでは開発者やプレイヤーはゲームに対し、ゲームの概要を表すゲームタグを付与できる。タグはゲーム内容の把握や検索に利用されるため、無数のゲームが存在する昨今において適切なタグの付与は必須と言える。しかしながら適切なタグの付与は容易ではない。その要因としてSteamに存在するタグが豊富である点や、一貫したタグ付与基準がない点が挙げられる。本研究ではSteamのゲームタグ付与支援を目的とし、機械学習によるタグ推薦手法を提案する。ゲーム内容の説明文やゲーム画面画像からタグを予測するマルチラベル分類により、ゲームを跨いだ判断に基づくSteamのタグ推薦を目指す。

キーワード 機械学習, マルチラベル分類, タグ付与支援, Steam

1. はじめに

世界中のPCゲームが公開されるプラットフォームとしてSteam¹がある。Steam上のゲームには複数のゲームタグが付与される。タグはゲームの概要を表す情報の1つであり、ゲーム内容の把握に利用される。タグはゲーム検索やレコメンドにも利用される。Steamでリリースされているゲームは年々増加しており、2024年には約19,000タイトルものゲームがリリースされている²。その中から好みのゲームを探すには適切なゲーム概要の把握や検索は必須であり、そのためにも適切にタグを付与する必要がある。

しかしながら適切なゲームタグの付与は容易ではない。その難しさの1つはタグの豊富さにある。Steamでは13種類のカテゴリーについて400種以上のタグが用意されており、その全体を把握することすら困難である。またタグはゲームに対する一種の属性であり可能な限り多く付与すべきであるが、多種多様なタグの存在が網羅的なタグ付与を妨げる要因になる。客観的なタグ付与が要求される点も難しさの1つである。タグ利活用の観点からは、タグ付与基準は可能な限り開発者の主観を排除しつつSteam全体で一貫しているのが望ましい。Steamでは集合知に基づく、すなわちプレイヤーによるタグ付与もサポートされている。しかしゲームリリースの時点では当然ながら適用できないため、開発者の主観に頼らざるを得ない。

ゲームタグの付与を支援する手法はいくつか提案されている。Steamではゲームタグを推薦するTag Wizard³が提供されている。Tag Wizardはタグの共起関係を学習し、開発者が付与

したタグから関連する他のタグを推薦する。しかしTag Wizardはタグの共起関係のみに基づいており、ゲームの内容を直接学習しているわけではない。ゲーム内容を直接学習してタグやジャンルを予測する手法も提案されている[1][2]。RubeiとSipioはゲームの説明文を学習データとしたタグ推薦システムAURYGAを提案している[1]。またJiangとZhengはゲームの説明文に加えカバーイラストも学習データとしたマルチモーダルな手法を提案し、ゲームのメインジャンル予測を行っている[2]。しかしこれらの手法は予測するタグやジャンルの候補が制限されている。また予測するタグやジャンルの数も高々5個と少ない。

本研究の目的は、Steam上でのゲームリリース時における客観的なゲームタグ付与の支援である。そのため本研究ではマルチモーダルなタグ推薦手法を提案する。提案手法ではゲームリリース時に存在する情報を機械学習し、適するタグを推薦する。また既存手法からの新規性として、新たなモダリティの追加と推薦内容の拡張、および新たな損失関数の設計を行う。提案手法では既存手法で用いられたゲームの説明文と画像に加え、要求スペック情報を学習するモデルも導入する。加えて推薦するタグの範囲に制限を設けず、Steamに存在する全てのタグから、Steamで付与できる最大数の20個のタグ推薦に取り組む。さらに推薦タグ候補の拡大のために、タグの偏りや排他性を考慮した損失関数を設計し導入する。

本稿では提案手法の一部であるゲームの説明文のみを用いたゲームタグの推薦を実装し、評価実験でその推薦性能と傾向を確認した。実験の結果、ゲームの説明文のみの推薦精度は最大でも27%と、十分とは言えない予測精度であった。予測精度が低い原因を調べるため提案手法の推薦傾向を分析した結果、テキストでは判断が困難なカテゴリーの存在と排他的なタグの存在が推薦結果に悪影響を与えていることが分かった。

(注1) : <https://store.steampowered.com/>

(注2) : <https://steamdb.info/stats/releases/>

(注3) : <https://partner.steamgames.com/doc/store/tags>



図1 Steamのゲームタグの一例

表1 ゲームタグのカテゴリーとゲームタグ例

カテゴリー	種類数	ゲームタグ例
Top-Level Genres	11	Action, Racing, RPG
Genres	68	Action-Adventure, Fighting, Soccer
Sub-Genres	62	3D Fighter, FPS, Survival Horror
Visuals & Viewpoint	30	2D, Realistic, Pixel Graphics
Themes & Moods	110	Fantasy, Military, Steampunk
Features	64	Conversation, PvP, Team-Based
Players	10	Co-op, Multiplayer, Singleplayer
Other Tags	24	Indie, Mod, Soundtrack
Software	13	GameMaker, Photo Editing, Utilities
Assesments	20	Beautiful, Difficult, Emotional
Ratings etc	7	Blood, Mature, Violent
Hardware / Input	7	Controller, Mouse only, Voice Control
Funding etc.	4	Crowdfunded, Early Access, Kickstarter

2. 準備

2.1 Steamのゲームタグ

Steamでは各ゲームにゲームタグ⁴を付与できる。図1にゲームタグの一例を示す。ゲーム開発者はSteamでゲームをリリースする際に最大20個のタグを付与できる。またリリース後はゲーム開発者だけでなくプレイヤーも適するタグを提案できる。

本稿執筆時点でSteamには全部で447種類のゲームタグが用意されている。これらのタグはジャンルや視覚属性などの13種類のカテゴリーで構造化されている³。表1にSteamのタグのカテゴリー、タグの種類数、および該当するタグの一例を示す。表に示すようにSteamでは多様なカテゴリーのタグが提供されているため、ゲーム開発者はゲームジャンルだけでなく、世界観やグラフィックなど様々な視点から適切なタグを付与する必要がある。

ゲームタグは主にゲームの検索で利用される。Steamではゲームを検索する際、特定のタグが付与されたゲームで絞り込む検索が可能である。そのため例えばレースゲームに興味があるプレイヤーなら、Racingのタグが付いたゲームを検索してその中から気になるゲームを探す、といった使い方が可能である。また、ゲームタグは Recommendにも利用される。Steamはあるプレイヤーが過去に遊んだゲームの傾向から、そのプ

レイヤーが興味を持ちそうなゲームをRecommendする。このRecommendの際にゲームの傾向を分析するために利用しているのがゲームタグである。このようなタグの性質から、無数のゲームが存在するSteamでプレイヤーが自分の好みに合うゲームを探すために、適切なタグの付与が必須である。

2.2 適切なゲームタグ付与の難しさ

しかし適切なゲームタグの付与は困難であると考えられる。本稿では困難である要因を3つに大別して考える。

1つ目はゲームタグの豊富さである。2.1節で述べた通りSteamには本稿執筆時点で447種類のタグが用意されている。またタグは毎年追加されており、今後もこの種類数は増加すると考えられる。さらにSteamでは最大20個のタグを付与できるが、適切な範囲内で数多くのタグを付与すると、ゲームの検索や推薦に該当しやすくなる。そのため、多くのプレイヤーの目に触れるようにするには、ゲーム開発者はタグとして適切な範囲内で可能な限り多くのタグを付与する必要がある。

2つ目は一貫したタグ付け基準が存在しない点である。SteamのゲームタグはSteamにリリースされている全てのゲームで共通して使用される横断的な情報である。そのためタグ付けをする際には、Steamにリリースされているゲーム全体で一貫した基準のもとでタグ付けをすべきと考えられる。しかしSteamでリリースされているゲーム全体をゲーム開発者が把握するのは現実的ではなく、それゆえ一貫性のあるタグ付与基準を作るのは難しいと考えられる。

3つ目は、リリース前のタグ付けはゲーム開発者の主観に頼るしかない点である。ゲームタグはゲーム開発者やプレイヤーの主観をもとに選ばれるが、ゲームリリースから十分時間が経てば集合知により客観的なタグに収束していく。しかしリリース前には当然ながらプレイヤーは存在しないため、リリース前のタグからは開発者の主観が排除できない。特にインディゲームに関しては、日々多数のゲームがリリースされている昨今で注目を集めるためにもリリース直後の動きは重要であると考えられる。そのためインディゲーム開発者はセーブルスに影響しうるリリース直後のタグを、限られた情報と開発者の主観のみで決めなければならない。

2.3 既存手法

Steamではゲームタグ付与支援を目的とした機能としてTag Wizard³が提供されている。Tag Wizardはゲームに付与されたタグの共起関係を学習し、関連する他のタグを推薦する。この機能を利用すれば、ゲーム開発者はいくつかタグを決めることでそこから関連するタグを推薦機能で絞り込んで探せる。しかしTag Wizardはあくまでタグの共起関係のみに基づいており、実際のゲームの内容を直接学習してタグ推薦をしているわけではない。

ゲーム内容を学習してゲームのタグやジャンルを推測する機械学習の手法はいくつか提案されている[1][2]。RubeiとSipioはSteamのタグ推薦を行うシステムAURYGAを提案している[1]。AURYGAは学習データとしてゲームの説明文を利用する。与えられたゲームの説明文から、テキストに対する特徴量

(注4) : <https://store.steampowered.com/tag/steam>

抽出手法である TFIDFVectorizer⁵を用いて特徴量抽出する。そして抽出した特徴量に対して CNB (Complement Naive Bayes) [3] モデルを適用し、タグの予測結果を得る。しかし AURYGA では予測候補が主要なタグ 100 種類に限定されていた。また、予測するタグの数も最大 5 個に留まっていた。

Jiang と Zheng は IGDB.com⁶に掲載されているゲームについて、ゲームのメインジャンルを機械学習で予測する手法を提案している [2]。Jiang と Zheng の手法では学習データとして、ゲームの説明文とカバーイラストの 2 種類を利用する。与えられたゲームの説明文にはテキストベースモデルを、与えられたカバーイラストには画像ベースモデルを用いてそれぞれ特徴量を抽出する。そして抽出した特徴量を連結し、連結した特徴量と活性化関数を用いてメインジャンルを予測する。Jiang と Zheng は実際にテキストベースモデルとして Universal Sentence Encoder [4] を、画像ベースモデルとして ResNet-50 [5] を用いて提案手法の評価をしている。その結果、説明文やイラストを単体で用いてジャンル予測をするよりも、説明文とイラストの両方を組み合わせて予測する方が予測精度が高くなったと報告している。しかし、Jiang と Zheng は予測候補となるメインジャンルを 15 種に限定していた。また Jiang と Zheng の手法はシングルラベル分類を行う手法であり、予測するジャンルは 1 個のみであった。

3. 提案手法

3.1 概要

本研究の目的は、Steam 上でのゲームリリース時における客観的なゲームタグ付与の支援である。そのために、機械学習を用いたマルチモーダルなタグ推薦手法を提案する。

提案手法の概要を図 2 に示す。提案手法は既存手法 [1][2] と同様に、公開されているゲームの情報をモダリティとして利用しタグ推薦を行う。まず提案手法ではゲーム説明文、スクリーンショット、スペック情報の 3 つの情報を独立に学習しそれぞれのモデルを得る。次に各モデルから得られた 3 つの特徴量を連結して、1 つの特徴量にする。最後に連結した特徴量に活性化関数を適用すると、各タグが入力データに適する予測確率が出力される。この予測確率の値が事前に定めた閾値以上であればゲームに適するタグであると見なし、そのタグを推薦する。もし予測確率の値が閾値以上のタグが 20 個を超える場合、予測確率の値が高い方から 20 個のタグを、提案手法が推薦するタグとする。

3.2 既存手法との違い

既存手法との違いは、モダリティの種類、推薦するゲームタグの範囲、および損失関数の種類である。表 2 に既存手法 3 種と提案手法の違いを示す。提案手法では既存手法のモダリティだけでは予測が困難なタグの推薦精度向上を目指して、タグの推薦に利用するモダリティを拡充させる。また提案手法

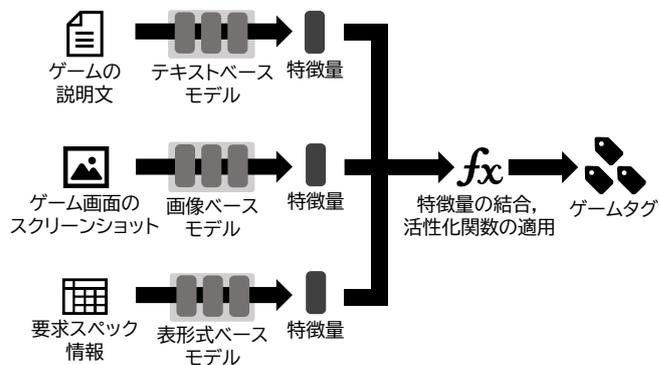


図 2 提案手法の概要

では推薦するタグに制限を設けず、全てのタグの中から適切な 20 個のタグの付与を目指す。さらに推薦するタグの拡張に伴いカテゴリーを加味した損失関数を設計し、タグの偏りと排他性を考慮したタグ推薦を試みる。以降では、モダリティの種類と推薦するタグの範囲、カテゴリーを加味した損失関数についてそれぞれ説明する。

3.2.1 モダリティの種類

提案手法ではゲームタグ推薦のためのモダリティとして、ゲームの開発者がリリース時に設定するゲームの説明文とゲーム画面のスクリーンショット、要求スペック情報の 3 種類を利用する。以降では、それぞれのモダリティの内容やそれがどのようにタグ予測に寄与するかについて順に説明する。

ゲームの説明文：ゲーム開発者によって記述される、ゲームの特徴や内容を記したテキストデータである。Steam ではゲームの説明文として、数百文字でゲームの主となる要素を端的に説明する Short Description と、長い文章でゲームの詳細を具体的に説明する About This Game の 2 種類がある。ゲームの説明文では主にゲームのジャンルや世界観が説明される。ゲームの説明文はプレイヤーがゲームの概要を知るために利用する基礎的なデータである。このゲームの説明文はゲームの特徴を決定づける重要な情報であり、既存手法 [1][2] でも採用されている。

ゲーム画面のスクリーンショット：ゲームのプレイ画面を写した画像データである。スクリーンショットでは主にゲームの世界観やグラフィック、プレイヤーの視点が見える。Jiang と Zheng [2] はゲームの説明文に加えてカバーイラストもモダリティとして取り入れ、予測精度を向上させている。そこで提案手法でも画像素材としてゲームのスクリーンショットを利用し予測精度の向上を図る。なお Steam のストアページでは IGDB.com とは異なり各ゲームのスクリーンショットが複数枚公開されている。1 つのゲームに対し複数の画像があるため、1 枚のカバーイラストのみを利用する Jiang と Zheng の手法よりも予測精度が向上すると考えられる。

要求スペック情報：ゲームをプレイするために必要な、PC のプロセッサやメモリ等の性能情報である。Steam のストアページでは各ゲームに推奨スペックと最低スペックが公開されており、提案手法ではこれらのスペック情報もモダリティとして利用する。スペック情報を取り入れた狙いは、一部のゲー

(注 5) : https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

(注 6) : <https://www.igdb.com/>

表2 既存手法と提案手法の違い

	モダリティの種類	推薦タグの候補	推薦タグ数	カテゴリーの考慮
Tag Wizard ³	ゲームタグの共起関係	全てのゲームタグ	最大 20 個	なし
AURYGA [1]	ゲームの説明文	主要なゲームタグ 100 種類	最大 5 個	なし
Jiang & Zheng [2]	ゲームの説明文 / カバーイラスト	メインジャンル	1 個	なし
提案手法	ゲームの説明文 / ゲーム画面画像 / 要求スペック情報	全てのゲームタグ	最大 20 個	あり (損失関数)

表3 カテゴリー別タグ付与回数

カテゴリー	タグ付与回数	カテゴリー	タグ付与回数
Top-Level Genres	59,001	Other Tags	23,325
Genres	22,163	Software	2,411
Sub-Genres	13,365	Assesments	13,832
Visuals & Viewpoint	18,525	Ratings etc	6,398
Themes & Moods	27,327	Hardware / Input	1,103
Features	14,051	Funding etc.	6,090
Players	14,879		

ムタグの予測に寄与すると考えられるためである。リアルな挙動を再現するために複雑な物理計算が求められる Simulation (シミュレーションゲーム) や、僅かな遅延が大きく影響する PvP (対戦型ゲーム) など、一部のゲームジャンルのタグには処理負荷と関連性があると考えられる。また Steam のタグには Realistic (リアル) や Pixel Graphics (ドット絵) など、グラフィックに関するタグも存在する。ゲームのグラフィックは処理負荷に関係する。そのためゲームの説明文だけでは予測が難しいグラフィックに関するタグが、ゲーム画面画像とパソコンのスペック情報を利用すれば推測できると考えた。スペック情報はプロセッサやメモリなど、どのゲームでも共通する複数の項目から構成されるため、表形式データとして扱う。

3.2.2 推薦するゲームタグの範囲

推薦するゲームタグの範囲について、まず推薦するタグの候補を既存手法 [1] [2] から拡張する。既存手法 [1] [2] は推薦するタグの候補に制限を設けていた。提案手法では推薦するタグの候補を Steam が提供している全てのタグに拡張して取り組む。

また推薦するゲームタグの数も既存手法 [1] [2] から拡張する。1 つのゲームで予測するタグの数について、Rubei と Sipio [1] は最大 5 個のタグ、Jiang と Zheng [2] は 1 個のメインジャンルを予測していた。提案手法では 1 つのゲームにつき Steam で設定可能なタグの最大数である 20 個のタグ推薦を行う。

3.2.3 カテゴリーを加味した損失関数

推薦するゲームタグの範囲を拡張するには、カテゴリーによるタグの偏りと排他性が課題となる。表 3 に Steam Store Games⁷で付与されているタグの数をカテゴリー別に集計した結果を示す。表 3 にある通り付与されているゲームタグの数はカテゴリーによって大きく異なる。特に Top-Level Genres のタグはゲームのメインジャンルを示す性質上、付与される回数が多くデータセットに出現する頻度が高い。このような不均衡

なデータセットで機械学習を行うと、推薦結果には Top-Level Genres のように出現頻度が高いタグが多く現れる。

一方で同じカテゴリーに属するタグは同じ側面からゲームの概要を表現している。この性質上、同じカテゴリーのタグが複数個付与されるゲームは数少ない。例えば、表 3 に示したように Top-Level Genres に属するタグは Steam Store Games で最も多く出現する。しかし 1 つのゲームに付与される Top-Level Genres のタグは平均 2.03 個である。このためカテゴリーを考慮せずに学習すると、特定のカテゴリーのタグに偏った不自然な推薦になってしまう。

全てのゲームタグの適切な推薦を実現するには、タグの偏りや排他性を踏まえた学習が必要である。Roth と Wen-tau はラベルの排他性を取り入れる方法として、排他的なラベルの組が予測に含まれると損失が ∞ になる項を損失関数に導入している [6]。本研究では Roth と Wen-tau が提案した方法を参考に損失関数を設計する。Roth と Wen-tau が提案する損失関数との違いとして、ゲームタグに合わせた制約の緩和を行う。Roth と Wen-tau の損失関数は排他的なラベルが 1 組でも存在すると損失が ∞ になる関数であり、排他性に対して極めて厳しい。しかし Steam のタグの場合、2,3 個程度であれば同じカテゴリーからタグが付与されるケースも少なくない。そのため Roth と Wen-tau の損失関数をそのまま提案手法に適用すると制約が過度に厳しく、かえって不自然なタグ推薦になると考えられる。この点を踏まえ、2,3 個程度の同一カテゴリータグの付与は許容し、かつ損失も過度に大きくならない損失関数の設計に取り組む。

4. 実験

4.1 実験の全体像と手法の一部実装

提案手法のタグ推薦性能を評価するために、Steam で公開されているデータを用いて提案手法によるゲームタグ推薦を行った。なお本稿では 3. 章で述べた提案手法の一部実装に取り組む。3.2.1 節で紹介したモダリティのうちゲームの説明文のみを用いた。また 3.2.3 節の偏りと排他性を考慮した損失関数は実装せず、マルチラベル予測問題でよく用いられる Binary Cross Entropy Loss⁸のみを用いた。

4.2 実験設定

データセットには Steam Store Games⁷ を利用した。このデータセットには 27,329 タイトルのゲームの情報が含まれる。ま

(注7) : <https://www.kaggle.com/datasets/nikdavis/steam-store-games>

(注8) : <https://pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html>

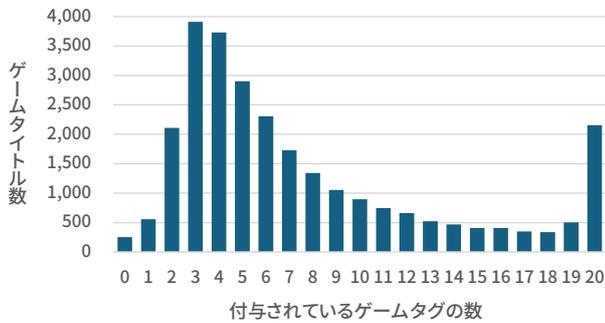


図3 データセットの1ゲーム当たりのタグ数の分布

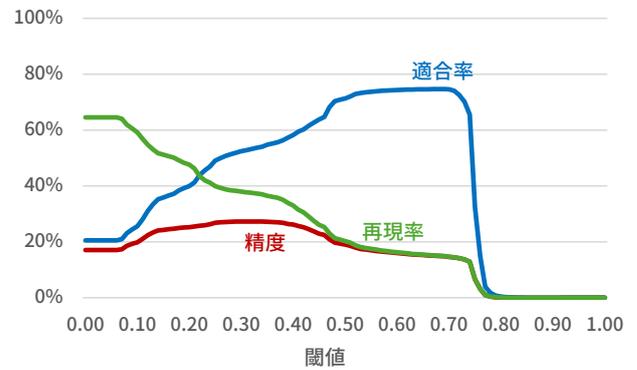


図5 閾値ごとのタグ予測性能



図4 SteamのShort Description

表4 カテゴリー別タグ推薦回数(閾値0.32)

カテゴリー	推薦回数	正解データに出現する回数
Top-Level Genres	74,975	57,167
Genres	275	21,340
Sub-Genres	132	12,799
Visuals & Viewpoint	117	18,432
Themes & Moods	331	26,291
Features	186	13,591
Players	552	14,185
Other Tags	27,116	22,702
Software	57	1,433
Assesments	102	13,428
Ratings etc	65	6,147
Hardware / Input	39	1,061
Funding etc.	44	5,708

たこのデータセットには全部で371種類のゲームタグの情報が記載されている。図3に各ゲームに付与されているタグの数の分布を示す。このデータセットでは、3~5個のタグを付与されているゲームが多い。一方で、データセットの半分は6個以上のタグを付与されているゲームである。また最大数である20個のタグを付与されたゲームも、データセット全体の1割程度存在する。データセットには10分割交差検証を適用して、モデルの学習と予測性能の評価を行った。

提案手法へ入力するゲームの説明文として、本実験ではShort Descriptionを利用した。図4にShort Descriptionの一例を示す。テキストベースモデルにはBERT[7]を転移学習して用いた。

本稿では2つの実験を行った。1つ目の実験は、提案手法の予測性能の評価である。提案手法が推薦するゲームタグがどの程度正解データと合致しているか確認する。予測精度の評価指標には、GodboleとSarawagi[8]が提唱したマルチラベル分類用の精度と適合率、再現率を用いる。GodboleとSarawagiはマルチラベル分類問題において、正解集合 T と予測集合 S に対する精度を $|T \cap S| / |T \cup S|$ 、適合率を $|T \cap S| / |S|$ 、再現率を $|T \cap S| / |T|$ で定義している。

2つ目の実験は、提案手法により推薦されたゲームタグの分析である。提案手法の推薦結果をカテゴリー別に集計し、タグの推薦頻度や傾向を確認する。

4.3 実験結果

4.3.1 実験1: 予測性能の評価

まずは提案手法の予測性能について確認する。図5は、提案手法による予測結果の精度、再現率、適合率を、閾値別に

示したグラフである。精度は閾値が0.32前後の時で最大27%と、ゲームタグ推薦という目的に対して十分とは言えない結果であった。適合率は閾値が0.48~0.75の間の時は70%以上と高い値を保っているが、閾値が下がるにつれて低下し、最終的には20%まで低下した。この結果から、提案手法では1,2個程度のタグは高い確率で的中させられるが、それ以上の数のタグを同時に的中させるとすると予測性能が著しく悪くなると分かる。再現率は閾値が0.06未満の時で最大64%であったが、閾値が0.06未満の時は提案手法は全てのゲームに対して20個のタグを推薦していた。そして本実験で使ったデータセットは3~5個のタグが付いたゲームが多い点を考えると、再現率64%という値は十分な数値とは言えない。

4.3.2 実験2: 推薦されたゲームタグの分析

続いて、提案手法が推薦するゲームタグの傾向について確認する。ここでは実験1で精度が高かった閾値0.32の時に提案手法が推薦したタグについて確認する。

表4に提案手法が推薦したタグの数と正解データに出現するタグの数をカテゴリー別に示す。表を見ると正解データに出現する回数が多いTop-Level Genresのタグは、推薦された回数も著しく多い。またOther Tagsも推薦された回数が顕著に多いカテゴリーであるが、これはIndie(インディゲーム)というタグによるものである。Indieが推薦された回数は27,009

回、正解データに出現する回数は 20,270 回と、Other Tags の大部分を占めている。この Indie が Top-Level Genres のタグに匹敵する頻度で正解データに出現したため、カテゴリーとして Other Tags も Top-Level Genres と同じような結果になった。

一方で、Visuals & Viewpoint のタグは出現頻度に対する推薦回数が最も少なかった。これはテキストベースモデルのみを利用した推薦では視覚属性を示すタグの付与が困難であることを示している。

4.4 考 察

本実験の結果から、全てのゲームタグ予測においてはテキストベースモデルのみを用いたタグ推薦には限界がある点があった。特に、グラフィックに関するタグの推薦が困難である点があった。全てのタグの中から推薦を行うにはグラフィックに関するタグ推薦も必須であり、そのためにも画像やスペック情報など異なるモダリティからのアプローチが必要と言える。

また、精度を高めるにはカテゴリーを踏まえたゲームタグ推薦が必要である点も分かった。3.2.3 節で述べたように本研究で扱うタグには一般的なマルチラベル分類のラベルとは異なり、カテゴリーによる偏りや排他性が存在する。そしてその偏りが推薦内容に影響しているが、排他性があるため偏った推薦では精度が上がっていない。Steam のタグ推薦においては偏りと排他性を考慮した損失関数の設計により、より自然なタグ推薦が求められる。

5. 今後の課題

5.1 予測性能改善に向けた課題

今後の課題として、まずは図 2 で示したようなマルチモーダルな推薦手法の実現が挙げられる。本稿で評価できていないゲーム画面のスクリーンショットと要求スペック情報を利用したゲームタグ推薦を実装し評価する。また単純に各モデルを導入するだけでなく、説明文と画像、要求スペックから得られる特徴量の結合方法についても検討する。

カテゴリーを反映した評価指標の導入も、推薦性能を向上させるために必要な課題の 1 つである。4.4 節で述べたように、カテゴリーによる偏りと排他性は推薦精度の低下を招く。そこで同一カテゴリーのタグばかりを推薦すると評価値が大幅に下がるような評価指標を導入し、カテゴリーを考慮した自然なタグ推薦の実現を試みる。

5.2 予測性能改善以外の課題

提案手法をより実用的な手法にするための課題として、タグの希少性を反映した評価指標の導入が挙げられる。Steam のレコメンドアルゴリズムには、使用頻度が低いタグを重視するアルゴリズムが存在する³。そのため適切な範囲内で希少性の高いタグを付与すると、頻出タグのみを付与しているゲームよりもレコメンドされやすくなる。そこで希少なタグを適切に予測できると評価値が上がるような指標を導入し、適切な範囲内でおすすめされやすくなるゲームタグ推薦を目指す。

また同様の狙いで、タグの順序を反映した学習方法の導入も検討する。Steam では開発者はゲームにタグを付与する際、

単にタグを選択するだけでなく、付与したタグの順序も決定する。そしてタグの順序は、ストアページ上に表示されるタグやゲーム説明に使われるタグの決定基準に利用される³。しかし本稿で紹介したような単純なマルチラベル分類手法では、学習時にラベルの比重や優先順位は考慮されない。今後はタグの順序を考慮した学習方法を導入し、提案手法がより実用的な手法になるよう試みる。

6. おわりに

本研究では Steam のゲームタグ付与支援を目的とした、機械学習によるゲームタグ推薦手法を提案した。また本稿では提案手法の一部としてゲームの説明文を用いたタグ推薦の実装と評価を行った。結果、機械学習でタグを推薦する際に、データの不均衡性と排他的なタグへの対応が課題となることが分かった。

今後の課題として、まずはゲーム画面のスクリーンショットと要求スペック情報を利用したマルチモーダルな推薦手法の実現が挙げられる。単に画像モデルや表形式モデルを導入するだけでなく、各モデルから得られる特徴量の結合方法についても検討が必要である。次に、カテゴリーの性質を反映した損失関数の設計と実装が挙げられる。本稿の実験でタグの偏りや排他性を考慮したタグ推薦の必要性が確認できたため、これらを反映した損失関数を設計し自然なタグ推薦を実現する。また新たな評価指標の導入も今後の課題である。タグの希少性や順序などを反映した評価指標も検討し、推薦手法をより実用的な手法にする。

謝辞 本研究の一部は、JSPS 科研費 (JP24H00692, JP21H04877, JP21K18302) による助成を受けた。

文 献

- [1] R. Rubei and C. Di Sipio, "AURYGA: A Recommender System for Game Tagging," Italian Information Retrieval Workshop, 2021.
- [2] Y. Jiang and L. Zheng, "Deep learning for video game genre classification," Journal on Multimedia Tools and Applications, vol.82, no.14, pp.21085–21099, 2023.
- [3] J.D.M. Rennie, L. Shih, J. Teevan, and D.R. Karger, "Tackling the Poor Assumptions of Naive Bayes Text Classifiers," In Proceedings of International Conference on Machine Learning, pp.616–623, 2003.
- [4] D. Cer, Y. Yang, S. yiKong, N. Hua, N.L.U. Limtiaco, R.S. John, N. Constant, M. Guajardo-Céspedes, S. Yuan, C. Tar, Y.-H. Sung, B. Strope, and R. Kurzweil, "Universal Sentence Encoder," 2018. <https://arxiv.org/abs/1803.11175>
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), pp.770–778, 2016.
- [6] D. Roth and W.-t. Yih, "A linear programming formulation for global inference in natural language tasks," In Proceedings of Conference on Computational Natural Language Learning, pp.1–8, 2004.
- [7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," In Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp.4171–4186, 2019.
- [8] S. Godbole and S. Sarawagi, "Discriminative methods for multi-labeled classification," In Proceedings of Advances in Knowledge Discovery and Data Mining, pp.22–30, 2004.