

# 修士論文

## Pixiv のタグを利用したアニメ作品 推薦方法の研究

平成 27 年度

三重大学大学院工学研究科  
博士前期課程情報工学専攻  
人間情報学研究室

福田凌平

## 目次

第1章	はじめに	3
1.1	研究背景	3
1.2	研究目的	3
1.3	従来研究の問題点	4
1.4	推薦方法	4
1.5	はてなキーワードを使った作品タグの自動削除	5
第2章	実験方法	6
2.1	タグの自動分類	6
2.2	作品タグの類似度の計算	7
2.3	推薦システム利用者の興味度の計算	9
2.4	推薦システム利用者の推定興味度の計算	10
第3章	はてなキーワードを用いた誤分類の削減	11
3.1	はてなキーワードとは	11
3.2	はてなキーワード API を使った方法	12
3.3	はてなキーワード自動リンク API を使った方法	13
3.3	はてなキーワードに登録されている単語	13
第4章	評価	18
4.1	評価方法	18
4.2	評価結果	19
第5章	考察	22
5.1	全体の結果の考察	22
5.2	はてなキーワードを用いることによる差の考察	22
5.3	今後の課題	23
	謝辞	24
	付録	25
A	各種学会発表時の質疑応答	25
B	評価結果の順位まとめ	28
C	実験の作業の手順と注意点	30
D	各種発表時に使用したパワーポイント	32
	参考文献	33

# 第1章 はじめに

## 1.1 研究背景

Pixiv とはピクシブ株式会社が運営する、イラストの投稿に特化したソーシャルネットワーキングサービスである。Pixiv は誰でも付与することができる「タグ」によって、柔軟かつ重層的な分類と閲覧を実現している。2012年1月時点でアカウント数400万・イラスト総数2450万枚・一日の投稿数2万5000枚という膨大な情報を持っているにも関わらず、それらに付与されるタグは整理されていないため、閲覧の補助以外に利用することは難しい。

図1の左の写真はPixivのホーム画面で、ユーザーが投稿したイラストが新着順に表示されている。イラストをクリックすると右のようにイラストが拡大され、イラストの上部分には描かれたキャラクター名などのタグが付与されている。



図1 Pixivのホーム画面

## 1.2 研究目的

日本の新規アニメ放送数は年間130本を超えており、全てのアニメを視聴することは難しく、その中で個人の嗜好に合った作品だけを選ぶことは困難である。また、その時期に放送しているアニメの中でオススメのアニメを紹介するサイトはあっても、好みのアニメのジャンルは人によって変わるため全ての人が楽しめるとは限らない。

Pixivはアニメや漫画の二次創作が中心であり、大規模なアニメや漫画の情報を持っている。本研究ではPixivに存在するタグ情報を使って個人の嗜好に合わせたアニメ・漫画作品を推薦する推薦システムを作る。

### 1.3 従来研究の問題点

Pixiv に関する研究は、竹渕らの”Pixiv におけるキャラクタータグを検出する手法の検討[1]”や”Pixiv の二次創作イラストに含まれるジャンルタグの自動分類[2]”がある。竹渕らの研究では作品タイトルを表す作品タグと、そのキャラクターを表すキャラクタータグという区別を行い、イラストのタグ群からキャラクタータグと作品タグを検出した。しかし、検出した後の応用については提案されていない。

### 1.4 推薦方法

本研究ではタグを分類した後、アイテムベースの協調フィルタリングという手法を使ってアニメや漫画作品を推薦する。協調フィルタリングとは、多くのユーザーの嗜好情報を蓄積し、あるユーザーの情報を用いて自動的にリコメンデーション(推薦)を行う方法である。その中でもアイテムベースの協調フィルタリングはユーザーの行動履歴からアイテム間の類似度を計算し、類似するアイテムをおすすめする手法である[3][4]。例として大手の通販サイトである Amazon.com の「おすすめ商品」にも応用されている。本研究ではアニメや漫画の作品をアイテムとして、作品同士の類似度を計算し、類似する作品を推薦している。

**この商品を買った人はこんな商品も買っています**



図 2 Amazon のおすすめ商品

## 1.5 PixivAPI とその特徴

本研究では PixivAPI[5]を使ってタグの収集を行っている。以下はタグの例である。左に\*がついているタグは作者が付与したタグで、ついていないタグは閲覧者が付与したものである。この場合、キャラクターが白ソックスを履いて描かれたため白ソックスのタグが付与されている。右に P がついているタグは Pixiv 百科事典に登録されているタグ、その右に赤いラクダのマークがついているタグはそのアニメやキャラクターのグッズなどの関連商品があることを表す。



これらのタグを含むイラスト情報を PixivAPI で収集すると以下のようにになる。タグの情報は赤字の部分のみで、タグの単語以外の情報は削除されてしまっているため使用することができない。

```
{  
  "id": "51291292",  
  "title": "【のんのんびよりりぴーと】放送開始記念",  
  "caption": "入学式れんちゃん",  
  "tags": ["のんのんびより", "の  
  のんのんびよりりぴーと", "れんちゃん", "宮内れんげ", "白ソックス"],  
  "tools": ["Photoshop", "IllustStudio", "CLIP STUDIO  
  PAINT"],  
  "image_urls": {"px_128x128": "http://i1.pixiv.net/c/128x128/img-  
  master/img/2015/07/07/11/38/15/51291292_p0_square1200.jpg", "px_480mw": "http://i1.pixiv.net/c/480x960/img-  
  master/img/2015/07/07/11/38/15/51291292_p0_master1200.jpg", "large": "http://i1.pixiv.net/img-  
  original/img/2015/07/07/11/38/15/51291292_p0.jpg"},  
  "width": 1021, "height": 1431, "stats": null, "publicity": 0, "age_limit": "all-  
  age", "created_time": "2015-07-07 11:38:15", "reuploaded_time": "2015-07-07 11:38:15"  
}
```

## 第2章 実験方法

### 2.1 タグの自動分類

タグとはイラストのキーワードのようなもので、投稿時には最低1つのタグを付ける必要がある。タグは1つのイラストに複数登録することができ、投稿者以外のユーザーも付与することができる。Pixivに投稿されるイラストのタグは、描かれたキャラクターの名前、描かれたキャラクターの登場するアニメやゲームの名前、イラスト作成に使われた道具や、閲覧者の感想などがある。下の図は付与されたタグの例である。



マウスとピクシアはイラストの作成に使われた道具やソフト、ポケットモンスターはアニメの名前、ピカチュウとサトシはキャラクターの名前である。本研究ではタグを大きく三つに分類した。

- ① 作品タグ … アニメ・漫画のタイトル
- ② キャラクタータグ … アニメ・漫画に登場するキャラクター
- ③ 不要タグ … イラストの作成に使った道具など Pixiv 特有のタグ

Pixivのタグはカテゴリ分けされていないため自動で分類する必要がある。本研究では、タグのカテゴリ分けに、

1. 作品タグとキャラクタータグは共起する確率が高い
  2. 作品タグはキャラクタータグより出現する頻度が高い
- という二つの特徴を使うことで、確率的に行う。

具体的には、タグ*i*とタグ*j*の共起率が0.7を超え、タグ*i*がタグ*j*より出現回数が多い場合、タグ*i*を作品タグ、タグ*j*をタグ*i*に属するキャラクタータグとし、どちらにも当てはまらないタグを不要タグとする。これを繰り返すことで、図3のように作品タグとキャラクタータグ、不要タグに分類でき、作品タグが複数の子のキャラクタータグを持つ図4のようなデータ構造の辞書が得られる。

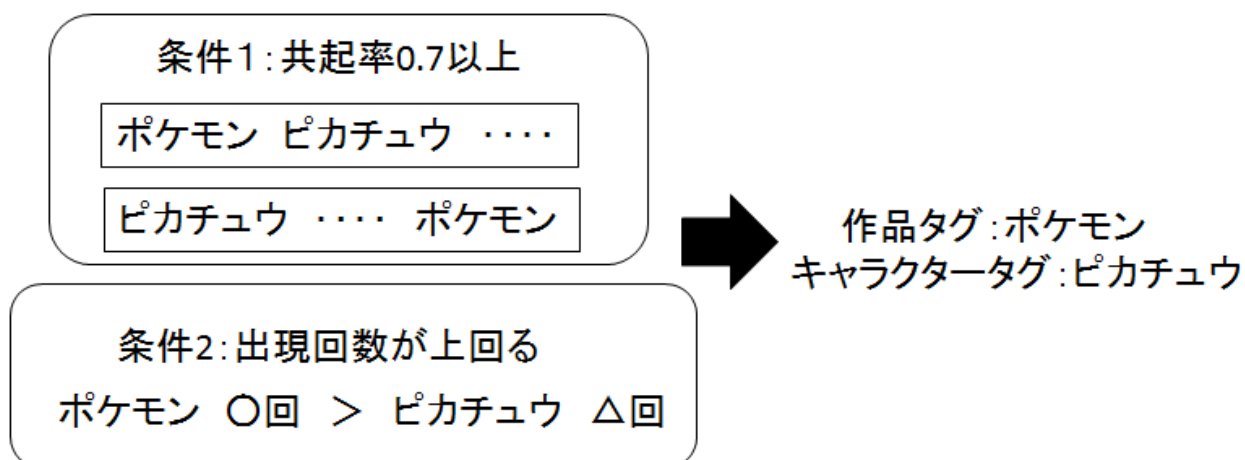


図3 作品タグとキャラクタータグの生成

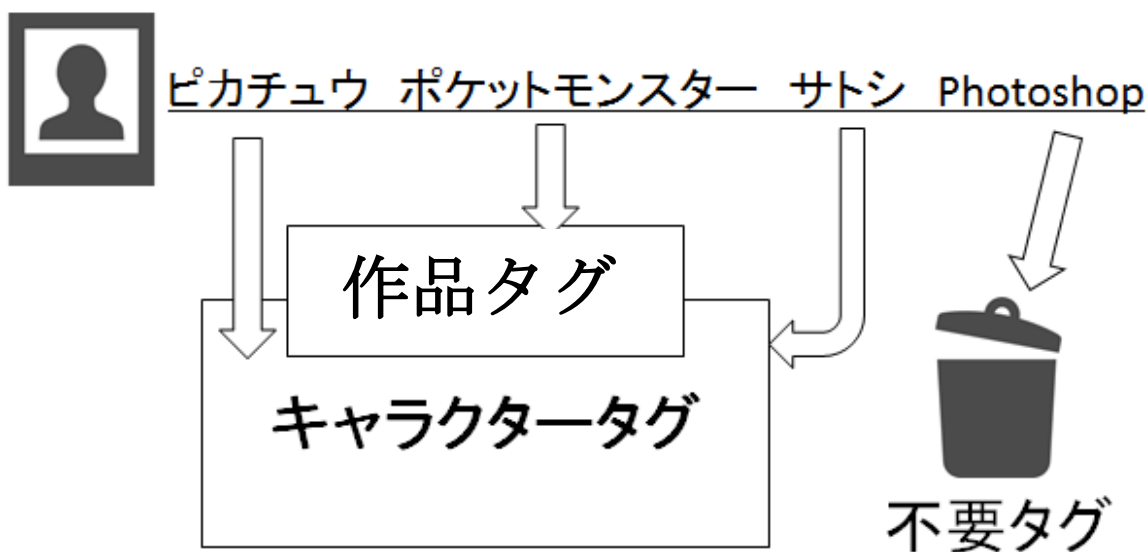


図 4 タグのカテゴリ分け

6 ページのタグは実際にカテゴリ分けのアルゴリズムを使ってタグを分類した例である。ポケモンという作品タグの下に複数のキャラクタータグが存在している。

ポケモン -> **Cattleya** DPPt ORAS TeddyJack XY(アニポケ) teentitans それいけ!アンパンマン  
 なにこの仔かわいい アポロ アルファサファイア イブキ イーブイ ウィッシュルーム ウィンデ  
 ィ エンプオー オイチ オオタチ オニゴーリ オメガルビー カイオーガ カイル・ハイド カスミ  
 カトレア カビゴン カルセレ カンベエ ガンダム以外 クチート クルマユ ケロマツ ゲッコウガ  
 ゲンシカイオーガ ゲンシグラードン コジョンド ゴクリン ゴシル ゴース シルクリ シーク ジ  
 ャイアン ジャローダ ジョウト組 ジラーチ スイクン スピアー セレナ セレナ(トレーナー) ゴ  
 ロアーク ダイアン ダイアン(ポケモン) ダイゴ チョンチー

赤字のタグは正確に分類できたキャラクタータグ、黒字のタグは作品タグや不要タグが誤ってポケモンの子のキャラクタータグに分類されてしまったものである。それ以外に、作品名でないタグを誤って作品タグと分類してしまったものもあった。これらの解決方法については 3 章で説明している。

## 2.2 作品タグの類似度の計算

類似度とは、作品タグ同士の関連度の高さを表し、類似度が高い作品タグ同士は同じユーザーが共に好む確率が高い。本研究ではユーザーが投稿したイラストの投稿履歴に基づいて類似度の計算を行う。過去の投稿履歴はユーザーの嗜好を表すと考えられるからである。例として、図 5 では同じユーザーがドラゴンボールとセーラームーンのイラストを投稿している。すなわちドラゴンボールとセーラームーンの類似度は高いと考えられ、ドラゴンボールに興味がある人にセーラームーンを推薦すれば嗜好に合う確率が高い。

まず、ランダムでユーザーを選び、投稿したイラストのタグ一覧を収集してくる。そして2.1節でのカテゴリ分けに従い、キャラクタータグは上位の作品タグに置き換える。このとき、不要タグと重複は削除する。これを複数のユーザーに対して行う。 $n$ 種類の全作品タグの集合を $\alpha = \{1, \dots, n\}$ とし、作品タグ $i$ と $j$ の類似度 $\rho_{ij}$ はJaccard係数と呼ばれる以下の式で求める。出現回数の中の共起数が多いほど類似度は大きくなる。

$$\rho_{ij} = \frac{|i \cap j|}{|i \cup j|}$$



図 5 Pixiv の投稿例





**ポケットモンスターとドラゴンボールの類似度は $1/3=0.333333\dots$**

図 6 類似度の計算例

## 2.3 推薦システム利用者の興味度の計算

興味度とは作品タグに対する興味の大きさである。興味度が高いほど利用者はその作品を好む。ユーザーにイラストを複数枚選んでもらい、そのイラストに付与されたタグを全て上位の作品タグに置き換えた出現回数を興味度とする。このとき不要タグは削除する。ユーザーが選んだイラストに付与されたタグが

ポケットモンスター ヒトカゲ ゼニガメ  
ルフィ ワンピース

ならば、2.1 節で作った辞書を活用してこれらを作品タグに置き換える。すると、

ポケットモンスター ポケットモンスター ポケットモンスター  
ワンピース ワンピース

となり、ポケットモンスターの興味度が 3、ワンピースの興味度が 2 となる。

## 2.4 推薦システム利用者の推定興味度の計算

2.3 節から利用者が興味を持つ作品は明らかになった。2.2 節で求めた類似度を組み合わせることで、利用者の興味を持つ作品に類似した作品を推薦する。 $s_j$ を選択されたイラストによる作品  $j$  に対する興味度とすると、選択されなかった作品  $i$  に対する推定興味度  $t_i$  を以下の式で予測する。ここでの、 $\alpha$  は  $n$  種類の全作品タグの集合、 $\rho_{ij}$  は 2.1 節で求めた作品タグ  $i$  と  $j$  の類似度である。このとき最も推定興味度  $t_i$  が高かった作品  $i$  を推薦する

$$t_i = \sum_{j \in \alpha} s_j \rho_{ij}$$

推定興味度の計算例を図 7 に示す。

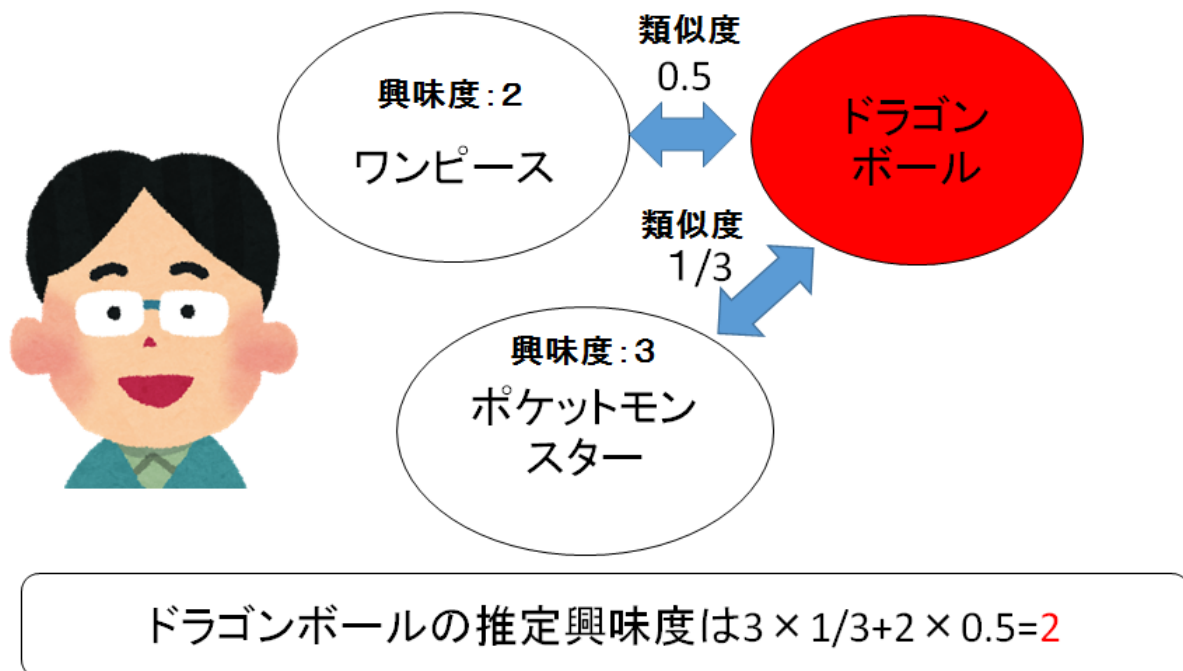


図 7 推定興味度の計算例

## 第3章 はてなキーワードを用いた誤分類の削減

### 3.1 はてなキーワードとは

アニメ、ポケットモンスターを例とすると、図8のようにカテゴリ分けされるべきである。しかし、2.1節で示したように機械的に分類してしまうと、図9のように作品タグとキャラクタータグに、本来は不要タグに分類されるはずのアニメや漫画に関係ないタグが多く入ってしまう。このような誤分類をなくすための方法を考える。

本研究では、はてなキーワードと呼ばれる百科事典のAPIであるはてなキーワードAPIを用いる[6]。はてなキーワードAPIの使用目的は、その単語がアニメやゲーム、漫画に関連する語かどうか調べることで図10のように正しい分類をすることである。はてなキーワードのAPIには以下の2種類がある。

1. はてなキーワードAPI
2. はてなキーワード自動リンクAPI

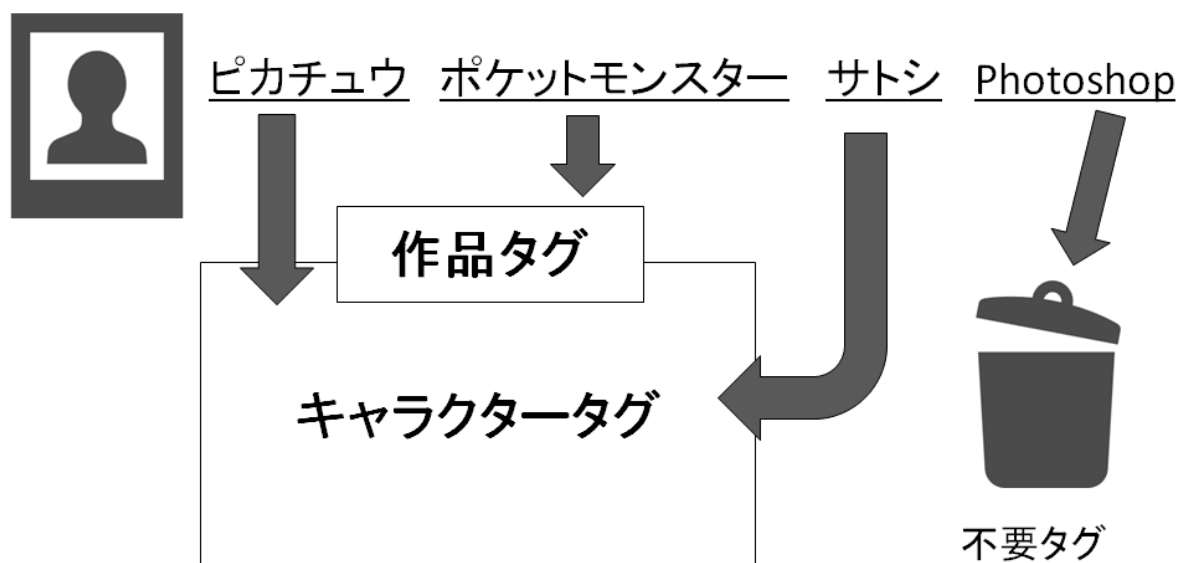


図8 カテゴリ分けの成功例

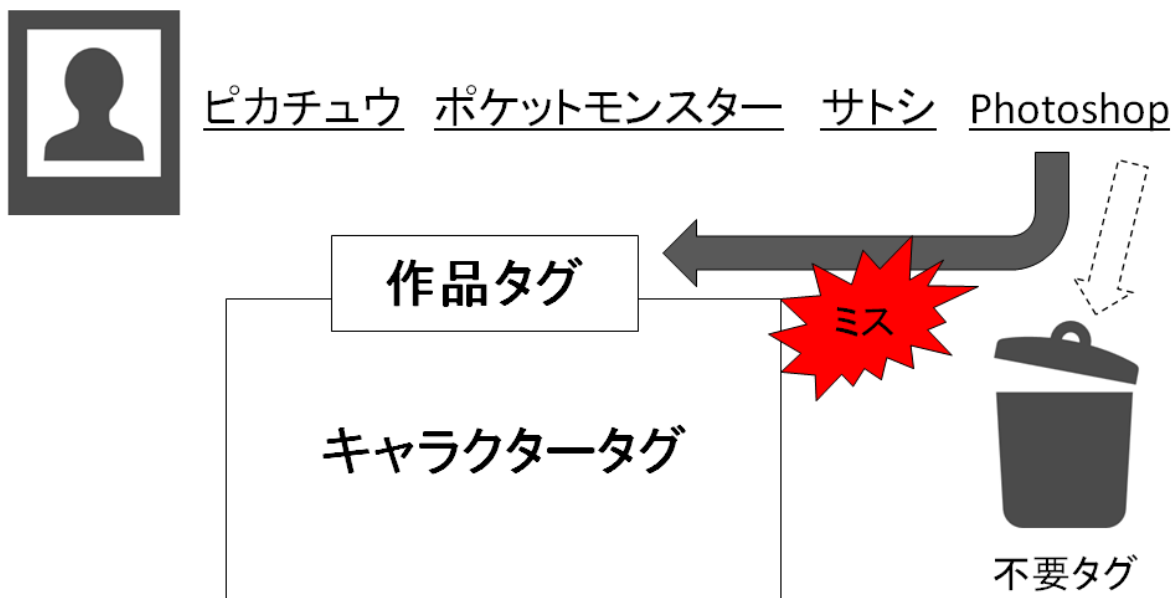


図 9 カテゴリ分けの失敗例

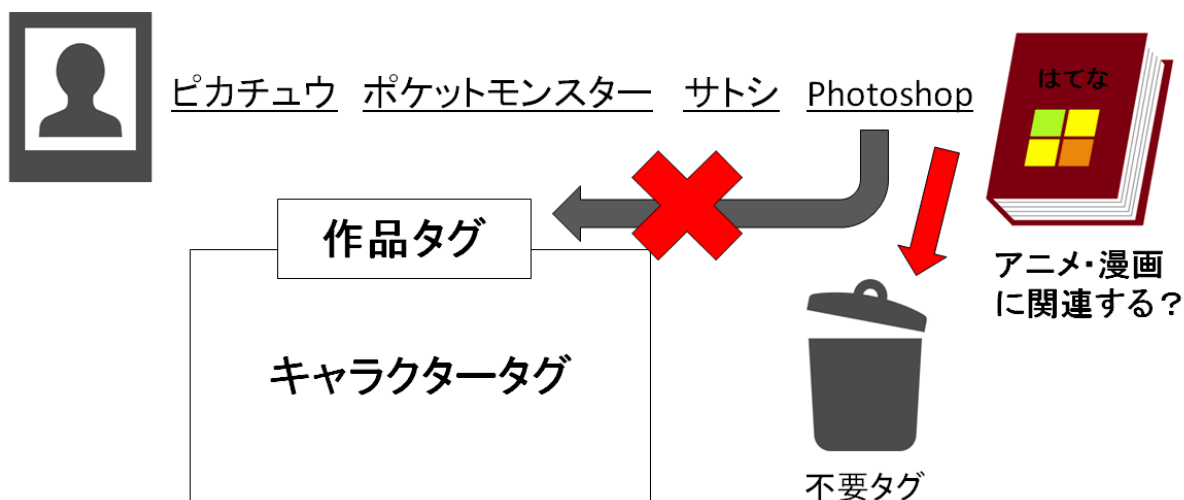


図 10 はてなキーワードを使ったカテゴリ分けの修正

### 3.2 はてなキーワード API を使った方法

キーワードを入力すると、はてなキーワード上の関連語が出力される。アルゴリズムは不明で、出力される数も一定ではない。

例として、ラブライブ！を入力すると

サンライズ、2013 年、シリーズ構成、音楽、雑誌、東條希、キャスト、絢瀬絵里、佐倉綾音、大橋歩夕、まれ、ランティス、公野櫻子、西田亜沙子、矢澤にこ、tv アニメ、ミューズ  
原案、制作、飯田里穂、6 月、プロジェクト、監督、小泉花陽、南條愛乃、2014 年、スタッフ、花田十輝、三森すずこ、徳井青空、1 月、クリップ、英玲奈、作画、電撃 g' s magazine、アニメ、μ' s、高坂、西木野真姫、理事長、あんじ、ライブ、原作、星空凜、放送、Pile、ミカ、久保ユ

リカ、松永真穂、楠田亜衣奈、3月

が出力される。アニメという文字が含まれているため、ラブライブ！はアニメに関連する単語と予想され、消去しない。

艦これを入力すると

黒歴史::その他、艦隊これくしょん～艦これ～

が出力され、一回ではアニメや漫画という単語は含まれていない。

そこで、更に出力された艦隊これくしょん～艦これ～を入力することで、

日高里菜、2014年、RJちゃん、ゲーム、運営、形態、商品、竹達彩奈、睦月、10月10日、image、イベント、井口裕香、旧日本軍、巡洋艦、大和、八八資源セット、dmm.com、アクティブ、ユーザ、基本無料、種田梨沙、操作、突破、8月、まり、ブラウザ、艦隊、最上、声優、東山奈央、5月23日、！すでのな、スタート、概要、航空巡洋艦、榛名、天龍、連装砲ちゃん、2015年、tvアニメ、コミック、運営側、軽巡洋艦、上げ、中島愛、睦月型、1月、pc、キーワード、育成シミュレーションゲーム

と関連語にアニメ、ゲームが出力された。省略された名称でも繰り返し検索することで判別することができる。

### 3.3 はてなキーワード自動リンク API を使った方法

作品タイトルで検索すれば、以下のように所属カテゴリが出力される。ほとんどのアニメや漫画は登録されている。

```
{score=62, cname=anime, refcount=6, word=ラブライブ！}
```

```
{score=35, cname=game, refcount=72, word=艦これ}
```

```
{score=63, cname=comic, refcount=2, word=黒子のバスケ}
```

```
{score=63, cname=anime, refcount=2, word=黒子のバスケ}
```

score とは閾値でこの値を高く設定すればノイズを減らすことが出来る。閾値の具体的な説明は3.4節の最後で述べる。cname が所属するカテゴリである。作品によっては、comic と anime など2つ以上のカテゴリに所属している場合もある。本研究の場合 anime,game,comic のいずれかに所属していれば良い。

本研究ではこちらの、はてなキーワード自動リンク API を用いる。理由は、はてなキーワード API の関連語の出力アルゴリズムが不明なため、アニメ作品を入力しても関連語にアニメという単語が含まれないことが予想されるからである。

### 3.4 はてなキーワードに登録されている単語

○作品タイトル

アニメ作品 DB の中からはてなキーワードに登録されていない作品を赤、登録されている作品を黒で、放送年と Pixiv のイラスト総数を横に記述した。登録されている数が膨大なため、ア行から17作品を例として表示した。

あ行

ああっ女神さまっ 2005年 615件

アイアン・ジャイアント 1999年 13件

鉄の処女(アイアンバージン) 1992年 112件

アイアンマン 1996年 3179件

i-wish you were here- あなたがここにいてほしい 2001年 0件

あいうら 2013年 270件

AIKa 1998年 1168件

アイカツ! 2012年 20108件

アイ・シティ 1986年 0件

愛してナイト 1983年 13件

愛してるぜベイベ 2004年 32件

IGPX -Immortal Grand Prix- 2005年 18件

愛少女ポリアンナ物語 1986年 44件

アイシールド21 2005年 4767件

アイス・エイジ 2002年 47件

あいつとララバイ 1987年 3件

愛天使伝説ウェディングピーチ 1995年 181件

参考

艦これ 363456件

ラブライブ! 115082件

○登場キャラクター

有名な作品に登場する主要キャラクターは登録されているが、ほとんどのキャラクターは登録されていないと言ってもよい。

本研究では、3.3節でも述べたように、はてなキーワード自動リンク API を使って誤分類を削減する。作品タイトルに限った場合、はてなキーワードに登録されていない作品は Pixiv のイラスト総数が少ない作品、つまりほとんど登場しないマイナーな作品に集中しており、ご分類の削減に活用することができる。ただし、キャラクタータグはほとんど登録されていないためキャラクタータグの誤分類の削減に用いることはできない。

しかし、単純にはてなキーワードのカテゴリがアニメ、漫画、ゲームの単語を作品タグとした場合、ノイズが含まれることが予想される。そこで、どの程度ノイズが含まれているかを作品タグに分類されたリストに適用することで示した。15ページからの一覧が確率的に分類した作品タグの中で、はてなキーワードのカテゴリがアニメ、漫画、ゲームとなった単語である。色付きの単語は何らかの理由で誤って作品タグと見なしてしまった単語で、白の単語が正しい作品タグである。誤って作品タグと見なしてしまった作品を理由ごとに分類し、色分けした。このときの模

式図を図 11 に示す。

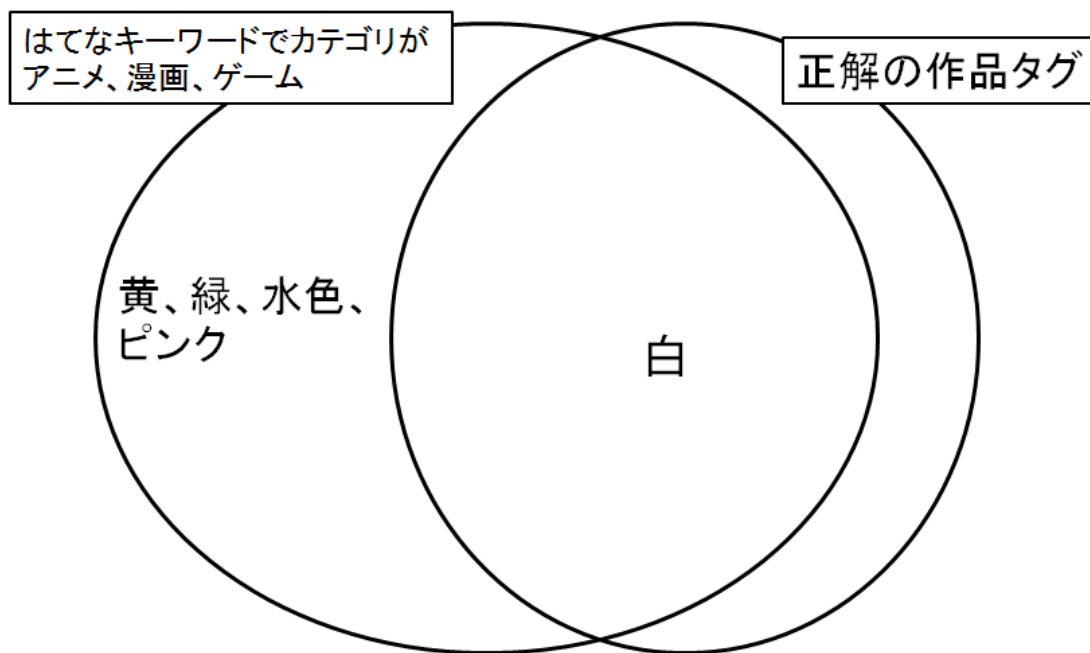


図 11 はてなキーワードと作品タグリストの比較

黄…キャラクタータグ

緑…アニメや漫画に関する用語

水色…タグとはてなキーワードの同名によるミス

ピンク…はてなキーワード特有のミス

白…作品タグ

{score=67, cname=comic, refcount=1, word=あだち充}

{score=31, cname=game, refcount=10, word=ブースター}

{score=51, cname=game, refcount=3, word=梨花}

{score=42, cname=comic, refcount=2, word=孫悟空}

{score=35, cname=game, refcount=29, word=AA}

{score=35, cname=anime, refcount=29, word=AA}

{score=33, cname=anime, refcount=2, word=アイカツ!}

{score=11, cname=comic, refcount=13, word=時雨}

{score=50, cname=comic, refcount=0, word=奈良シカマル}

{score=61, cname=anime, refcount=0, word=エウレカ}

{score=61, cname=game, refcount=0, word=エウレカ}

{score=90, cname=game, refcount=0, word=テイルズオブシンフォニア}

{score=90, cname=anime, refcount=0, word=テイルズオブシンフォニア}

{score=64, cname=comic, refcount=1, word=ラーメンマン}  
 {score=24, cname=anime, refcount=326, word=アニメ}  
 {score=33, cname=comic, refcount=0, word=桂小太郎}  
 {score=26, cname=comic, refcount=7, word=アメコミ}  
 {score=86, cname=game, refcount=0, word=世界樹の迷宮}  
 {score=48, cname=comic, refcount=5, word=スラムダンク}  
 {score=66, cname=anime, refcount=2, word=きんいろモザイク}  
 {score=66, cname=comic, refcount=2, word=きんいろモザイク}  
 {score=30, cname=comic, refcount=1, word=クラピカ}  
 {score=100, cname=comic, refcount=0, word=ラーメン三銃士}  
 {score=58, cname=comic, refcount=7, word=進撃の巨人}  
 {score=58, cname=anime, refcount=7, word=進撃の巨人}  
 {score=100, cname=anime, refcount=0, word=爆走兄弟レッツ&ゴー!!}  
 {score=100, cname=comic, refcount=0, word=爆走兄弟レッツ&ゴー!!}  
 {score=100, cname=comic, refcount=0, word=繰繰れ!コックリさん}  
 {score=50, cname=anime, refcount=5, word=魔法科高校の劣等生}  
 {score=85, cname=game, refcount=0, word=餓狼伝説}  
 {score=85, cname=anime, refcount=0, word=餓狼伝説}  
 {score=100, cname=game, refcount=0, word=BDFF}  
 {score=11, cname=game, refcount=0, word=ミツル}  
 {score=86, cname=anime, refcount=0, word=蟲師}  
 {score=86, cname=comic, refcount=0, word=蟲師}  
 {score=12, cname=game, refcount=0, word=ロードラ}  
 {score=56, cname=anime, refcount=1, word=コードギアス}  
 {score=77, cname=game, refcount=0, word=ファイナルファイト}  
 {score=32, cname=game, refcount=3, word=ポイズン}  
 {score=83, cname=game, refcount=1, word=ゴッドイーター}  
 {score=83, cname=anime, refcount=1, word=ゴッドイーター}  
 {score=35, cname=comic, refcount=1, word=フェイト}  
 {score=35, cname=anime, refcount=1, word=フェイト}  
 {score=35, cname=game, refcount=1, word=フェイト}  
 {score=57, cname=anime, refcount=7, word=ガンブラ}  
 {score=57, cname=comic, refcount=7, word=ガンブラ}  
 {score=53, cname=game, refcount=0, word=lol}  
 {score=64, cname=anime, refcount=0, word=魔人探偵脳噛ネウロ}  
 {score=64, cname=comic, refcount=0, word=魔人探偵脳噛ネウロ}  
 {score=50, cname=comic, refcount=0, word=悪魔のリドル}  
 {score=50, cname=anime, refcount=0, word=悪魔のリドル}



```
{score=28, cname=game, refcount=3, word=アーチャー}
{score=41, cname=anime, refcount=1, word=忍たま}
{score=57, cname=comic, refcount=0, word=ジョースター}
{score=70, cname=anime, refcount=1, word=らき☆すた}
{score=70, cname=comic, refcount=1, word=らき☆すた}
{score=21, cname=game, refcount=11, word=ECO}
{score=66, cname=game, refcount=0, word=AYAKASHI}
{score=31, cname=game, refcount=2, word=ファンタジア}
```

**AA、フェイト** AA は、はてなキーワードではアークエンジェルというアニメの略で検索されたが実際のタグはアスキーアートの意味で付与されている。カテゴリがアニメのフェイトはリリカルなのはに登場するフェイト・テストロッサ、カテゴリが漫画のフェイトは DC コミックのキャラクター名でいずれも他作品のキャラクター名でふさわしくない。

**時雨、ファイナルファイト、ファンタジア** はてなキーワードは完全一致でなくても勝手にタグを区切って部分一致で検索してしまう。時雨は付与されていたタグの時点では時雨改二という艦これのキャラクター名でとして付与している。しかし、時雨と検索されることで漫画家の名前と誤って判定されている。同じようにファイナルファイトは本来ならばポイズン(ファイナルファイト)、ファンタジアは pixiv ファンタジア FK であった。

Recall(白/正解の作品タグ)は 24/26、Precision(白/(黄,水色,緑,ピンク,白))は 24/47  
メルクストーリーとペルソナ Q の 2 つの作品タグを抽出しきれなかったが Recall は高めである。  
Precision を向上させたいければ、score の閾値を 40 程度まで上げればよいが、アイカツ! のような正しい作品タグも削除されてしまうため今回の実験では設定していない。

# 第4章 評価

## 4.1 評価方法

実際に Pixiv 利用者に対して本システムを使用してもらい、感想を求めることがより良い方法であるが、今回は Pixiv 利用者の過去の投稿履歴を用いることで機械的に精度を評価した。対象者は 2.2 節で類似度を計算したユーザー以外のランダムで選んだ 50 人である。

評価方法としては、対象のユーザーの投稿イラストからタグを収集、2.3 節の方法に従って作品タグの興味度を計算した。その中から、最も興味度の高かった作品タグを隠し、興味度が 2 位以下の作品タグから 2.4 節の推定興味度の予測を行った。そして、その隠された作品タグが何番目の推定興味度の大きさで推薦されたかをグラフにまとめた。推薦不能とは推定興味度が 0 となった作品である。評価方法の模式図を図 12 に示す。

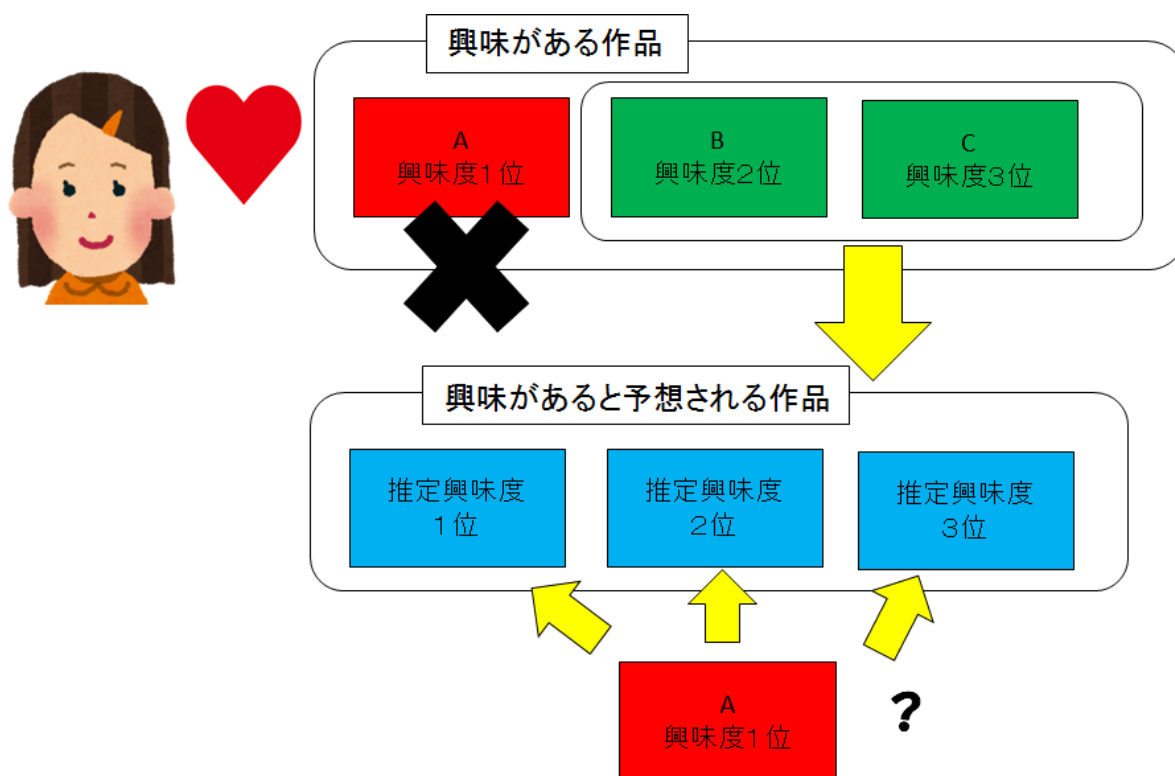


図 12 評価方法

## 4.2 評価結果

図 13 は、はてなキーワードを用いて推薦を行った結果、図 14 は、はてなキーワードを用いずに推薦を行った結果である。いずれも、隠した作品タグを最も正確に推薦できたのは推定順位が 1 位となった 7 回であった。20 位以内に推薦できたのは、ともに約 50% のユーザーに対してである。詳細な評価結果を付録 B の表 1 にまとめた。

比較のため、個人ごとの情報を用いずに、Pixiv のイラスト総数の順に候補を決めたのが図 15、2 位の作品を隠して、1 位と 3 位以下の作品から推薦を行ったのが図 16 である。図 15 では 1 位に推薦できたのは 9 人、20 位以内に推薦できたのは 31 人。図 16 では 1 位に推薦できたのは 8 人、20 位以内に推薦できたのは 35 人である。

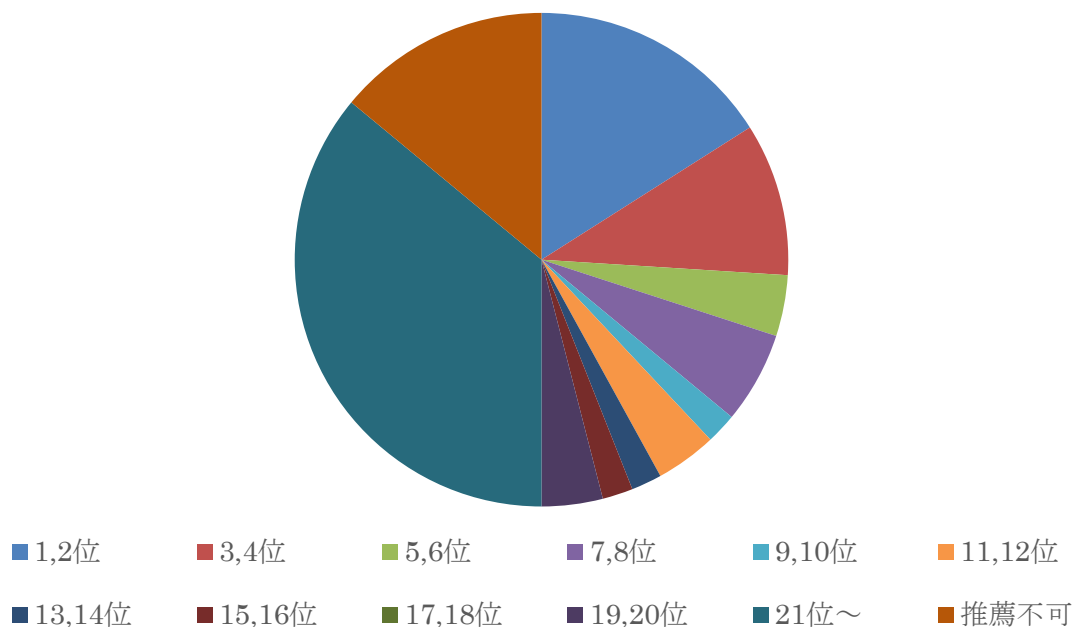


図 13 はてなキーワードを使った結果

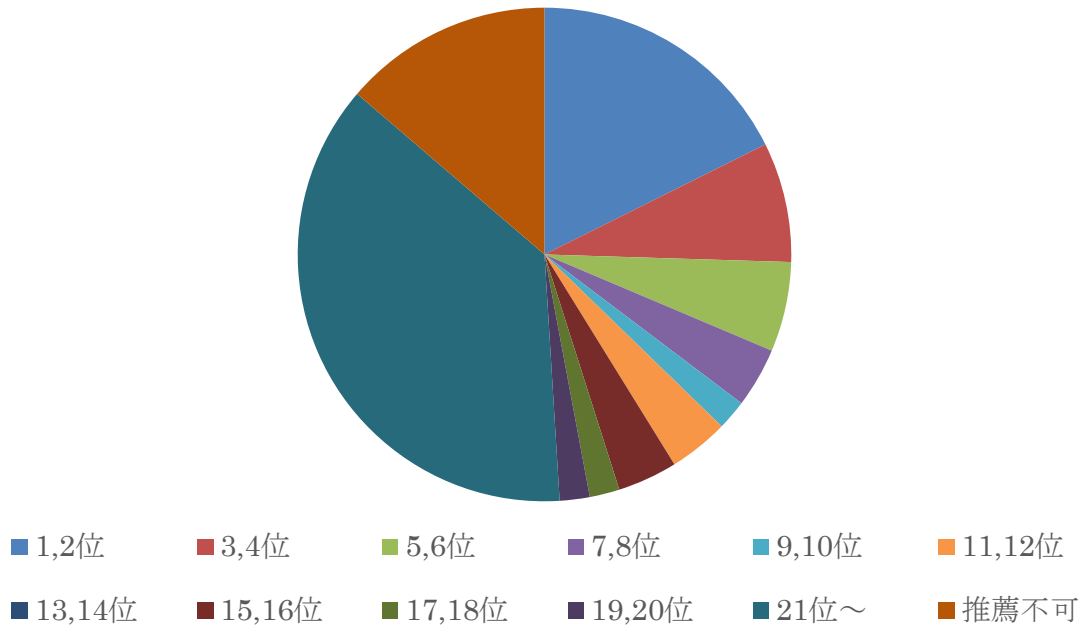


図 14 はてなキーワードを使わない結果

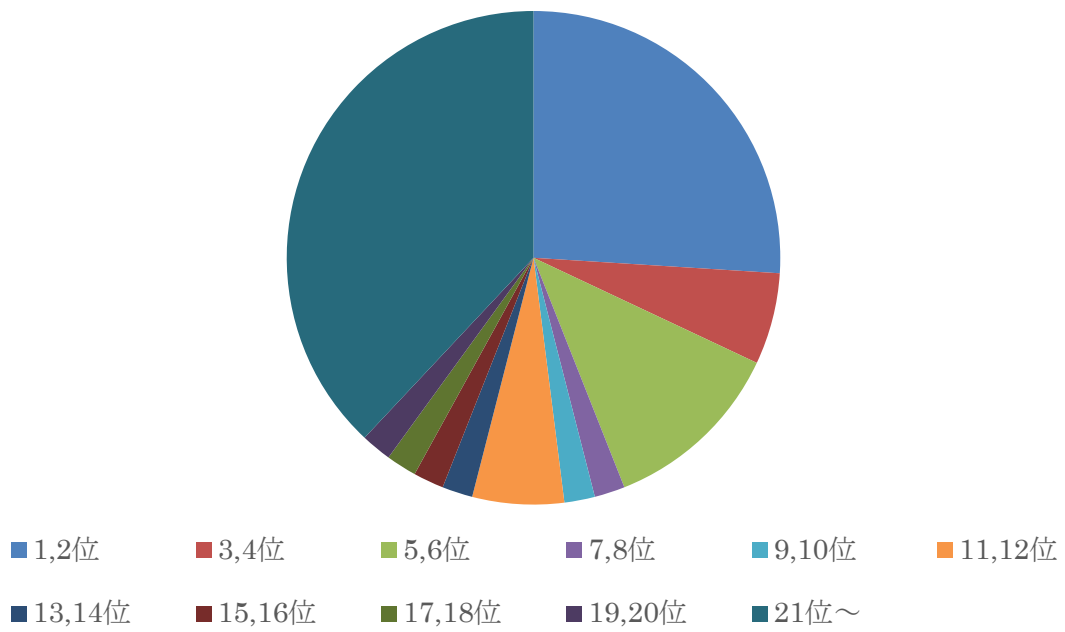


図 15 イラスト総数順に推薦した結果

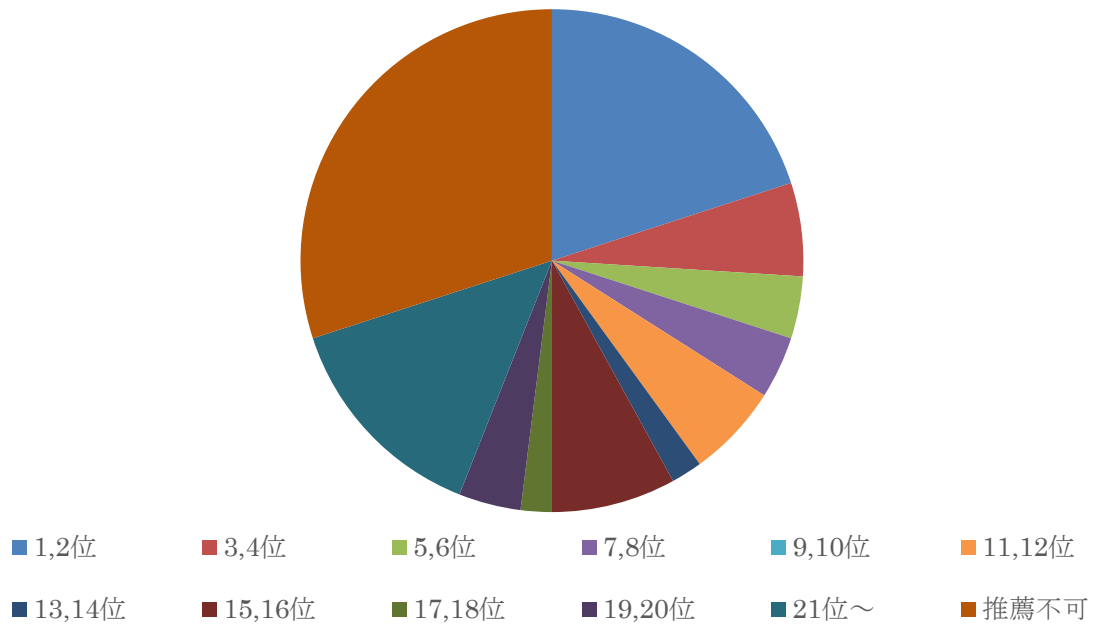


図 16 2位を隠して推薦した結果

# 第5章 考察




## 5.1 全体の結果の考察

推定順位が1位から20位になった25人、つまり約50%のユーザーに対して20位以内に推薦できており、作品タグにカテゴリ分けされた作品が約400種類あることを考えると高い精度で推薦することができた。しかし、推定順位が1位となった回数は合計7回と全順位の中で最も多いものの、作品の内訳をみると艦これ3回、アイドルマスターシンデレラガールズ2回、東方1回、おそ松さん1回と作品に偏りがあり、いずれも知名度の高い作品ばかりである。知名度順(イラスト総数順)に推薦した図15を見ると1位が9回と本研究より良い結果となってしまった。これは、1位に知名度が高い作品が固まりやすいことや、影響が大きい1位を隠すことで正しい推薦を行えなかったからだと思う。そこで、2位を隠して推薦をした図16を見ると、推定順位1位が8人、1位から20位が35人と、1位を隠した推薦より精度が良くなった。

## 5.2 はてなキーワードを用いることによる結果の差の考察

はてなキーワードを用いても用いなくても評価はほとんど変わらなかった。これは類似度の計算方法はノイズの影響を受けないからである。はてなキーワードを用いないと図17のように作品タグの中にノイズが多く混入するが、類似度を計算すると結果は9ページの図6と同じである。

**興味作品**

ユーザーA		<u>ポケットモンスター</u> <u>ノイズ1</u> <u>ドラゴンボール</u>
ユーザーB		<u>ドラえもん</u> <u>ポケットモンスター</u> <u>ノイズ2</u> <u>サザエさん</u>
ユーザーC		<u>ワンピース</u> <u>ノイズ3</u> <u>ドラゴンボール</u> <u>ノイズ4</u>

**ポケットモンスターとドラゴンボールの類似度は $1/3=0.333333...$**

図17 ノイズが混入した場合の類似度の計算例

## 5.3 今後の課題

### ■ マイナーな作品の推薦

推定順位が 1 位となった 4 作品のイラスト総数の平均は 194478 件と 3.3 節の作品と比べても膨大で、有名な作品ばかりであると言える。これらは既に多くの人に好まれているため他の人に推薦しても好まれる確率が高いことは明らかで、推薦に意外性がない。あまり有名ではないマイナーな作品を推薦する個人向けの推薦アルゴリズムを考える必要がある。

### ■ 評価方法の変更

今回の実験は機械的に評価を行った。評価方法もユーザーが既に興味を持っているものを当てるという方法で、将来的に興味を持つ物を薦めるという本来の意味の推薦では評価が変わると思われる。

### ■ 誰でも使えるツール化

今回の推薦では所々に手作業が入っており、瞬時に推薦できるわけではない。今後は CGI を用いて Pixiv のイラスト ID を入力するだけで、その人に合わせたアニメ・漫画作品が表示できるようにする。そのためには作品タグ、キャラクタータグに含まれるノイズが問題となる。はてなキーワードを使うことで、作品タグのノイズはだいぶ減らすことができたが、まだ完全にはなくすことはできていない上にキャラクタータグのノイズを減らすためには使えない。

# 謝辞

本研究を進めるにあたって、数々のご指導、ご助言を下さった河合敦夫准教授に心から感謝致します。また、有益なご助言、ご意見、ご協力をいただきました井須教授をはじめ、本学人間情報学研究室の皆様や甲南大学の永田教授には深く感謝致します。



# 付録

## A 各種学会発表時の質疑応答

◎改善すべきもので、実際に修正したもの

○改善すべきであったが、時間などの問題で修正しなかったもの

△改善する必要はないと判断したもの(福田)

×改善する必要はないと判断したもの(河合、福田)

2015年2月 大野先生

◎はてなブックマークというタグを使った Web サイトがある。

×イラスト画像から絵師を割り出すシステムがあり、画像認識を使ったシステムであるが、興味を持つ層は自分と似ている。

△新しい高度な手法を売りにするのではなく、特定の分野に応用したシステムであるので、多くの人に聞いてもらえる全国大会で発表するとよい。

△発表手法は興味を持ってもらえる人と長時間討論できるポスター発表がふさわしい。

○Web で公開してアンケートに答えてもらう。

2015年5月 合宿

△単純なタグの出現回数を興味度とするのはあまり良くないのではないか。(井須)

◎機械的に評価する方が良い。(永田)

2015年8月 永田さん

○最後のグラフ 2 種類を見ても一見あまり良い結果には見えない。評価のグラフを他の研究と比較しないと良いのかわからない

×Pixiv に特化した推薦ではないためパワーポイントで Pixiv の詳しくは説明いらず、省略した方が良い。タイトルからも外した方が良い。

○Pixiv 以外にもアニメのデータベースもあり、そちらはカテゴリ分けなどの手順を省略してアニメタイトルや登場キャラクターといった正確な情報を持っている。

×厳密にいうと協調フィルタリングではないので発表パワーポイントから外さないで詳しい人から突っ込みが入る可能性がある。

○難しいことをすると精度が下がるので、カテゴリ分けのような作業はしない方が良い。不要タグを削除するのではなく、全てのタグを同じように扱ってタグ同士の関係を見ると良い。

○Pixiv 百科事典、閲覧数、関連する商品が存在することを表すラクダのアイコンなどの Pixiv 特有の情報を使って推薦を行うべき。ユーザーの性別といった利用者の情報は他のデータベースも持っていると思われる。

## 2015年10月 国際シンポジウム

◎何か新しいポイントがほしい

→はてなキーワードを組み合わせでカテゴリ分けの精度を高めた

○分類の共起率の0.7を自動的に調整すると良い(佐々木)

◎このグラフで良いのかどうかわからない

→4.2節

×学術的にどうかアニメで意味があるのか(成瀬)

○既にユーザーの興味があるものが推薦されるかで評価をしているが、本来なら未知のものを推薦することが目的なので評価が変わるのではないか(大野)

○アマゾンに比べてどうか(鈴木)

→アニメの商品のページでのおすすめした他アニメ商品をアマゾンのアニメ推薦として比較する

## 2015年10月 東海支部

◎実際は知名度が低いアニメが推薦されるのか、その視点から見ると評価はどうか

○他の研究との共通点と比較

◎順位が高い作品は知名度が高い作品に集中しているのではないか

→順位が上位の作品は知名度が高い作品に集中しているのか調べるため、作品タグの総数の平均を順位ごとに図18にまとめた。以下の作品は上位5位までの作品である。昨年の実験のため今年とは結果が異なる。

1位 艦これ、銀魂、艦これ、東方、東方、VOCALOID

2位 艦これ

3位 アイドルマスターシンデレラガールズ、鬼灯の冷徹

4位 東方、テイルズオブゼスティリア

5位 刀剣乱舞

これだけでは順位が上位の作品はタグ総数が多い、つまり知名度が高いとは言えないように見える。しかし、作品ごとにタグ数は大きく異なる。特に、最もタグが多かった東方のタグ総数が1726644と2番目に多かったVOCALOIDの424936に比べても圧倒的に多いので、その順位に東方が含まれているかどうかでも平均が大きく変わってしまう。

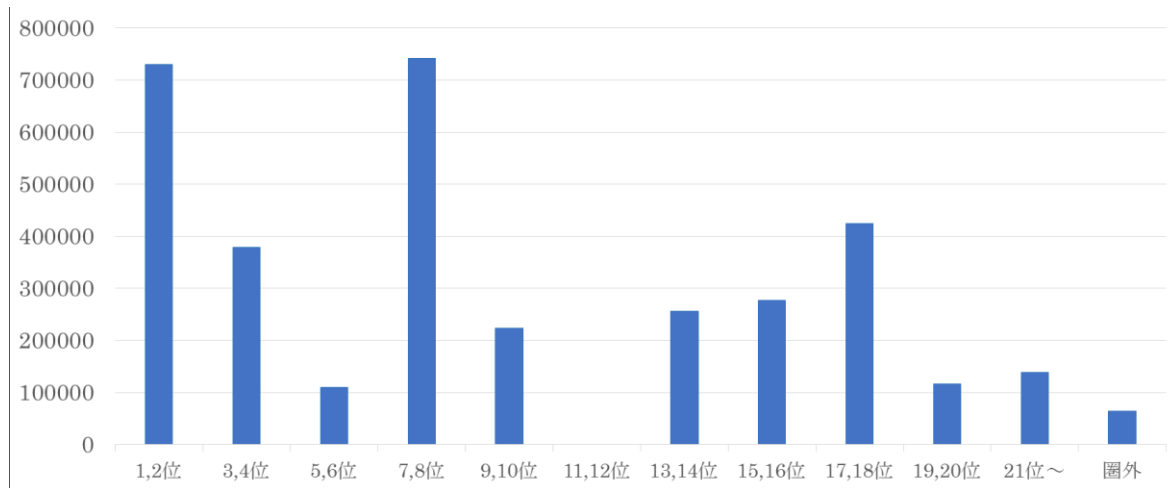


図 18 順位ごとの作品タグ総数の平均

修士論文発表質問

土井君

△類似度だけでなく、アニメの放送した年代のような要素も嗜好に影響するのではないかと？

◎マイナーな作品を推薦することが難しいと言っていたが、有名な作品を推薦すればもっと評価が良くなるのではないかと？

大山先生

○協調フィルタリングの研究が多くある中、独自性は何か？

→Pixiv のタグ構造を利用している

Amazon のように階層構造を持っているサイトは多くあり、他にも似たような研究がされているのではないかと？

→他にもはてなキーワード API を使ったカテゴリ分けのような様々な手法を組み合わせることでシステムを作っている

1つ1つの手法に新しさはないのではないかと？

○他の研究の推薦システムに比べて高い精度で推薦できているのか？

→本研究ではアニメ作品を推薦するシステムであるため直接的な比較はできない  
探せばあるのでは？

西野先生

×過去の英語の研究との共通性はあるのか？

## B 評価結果の順位まとめ

No	作品名	順位
1	東方	1
2	弱虫ペダル	11
3	星のカービィ	25
4	Free!	16
5	コードギアス	30
6	おそ松さん	19
7	テニスの王子様	4
8	艦これ	22
9	艦これ	8
10	おそ松さん	13
11	刀剣乱舞	4
12	ポケモン	4
13	アイドルマスターシンデレラガールズ	2
14	アイカツ!	29
15	東方	44
16	おそ松さん	7
17	おそ松さん	1
18	おそ松さん	42
19	艦これ	25
20	艦これ	1
21	白猫プロジェクト	79
22	ぷよぷよ	圏外
23	聖闘士聖也	7
24	刀剣乱舞	6
25	カードファイト!ヴァンガード	39
26	艦これ	1
27	東方	圏外
28	薄桜鬼	26
29	東方	圏外
30	魔法少女☆まどかマギカ	24
31	白猫プロジェクト	81
32	ヘタリア	4

	作品名	順位
33	薄桜鬼	76
34	おそ松さん	圏外
35	SD ガンダム	圏外
36	ハイキュー!	57
37	ワンピース	圏外
38	鉄血のオルフェンズ	圏外
39	Psycho-pass	48
40	アイドルマスターシンデレラガールズ	1
41	艦これ	50
42	魔法少女☆まどかマギカ	10
43	ドラゴンボール	49
44	フェアリーテイル	39
45	アイドルマスターシンデレラガールズ	1
46	ワールドトリガー	20
47	艦これ	1
48	ワンパンマン	圏外
49	ジョジョの奇妙な冒険	27
50	艦これ	1

表 1 評価結果まとめ

## C 実験作業の手順と注意点

### ① 辞書作成のためのタグ収集

作品タグとキャラクタータグの辞書を作成するために、まず `newtag.rb` を使って Pixiv の新着投稿イラストのタグを収集する。新着何件までのイラストからタグを収集するかは `def new_arrival_works` の `per_page` の値を変更することによって変えることができるが、一定の値を超えると動かなくなってしまうため、複数回の収集が必要である。

### ② 辞書作成

`push.pl` で作品タグとキャラクタータグの辞書を作成する。作品タグ->キャラクタータグという形式になっている。閾値は現在 0.7 である。

### ③ はてなキーワード自動リンク API によるノイズの削除

`Hatena.java` で②の辞書の作品タグのノイズを削除する。はてなキーワードにおいてカテゴリがアニメ、漫画、ゲームに所属しない作品タグは削除される。ファイル出力がうまくいかなかったので、今回の実験では標準出力した結果をファイルにコピーした。

### ④ 類似度計算のためのタグ収集

作品タグ同士の類似度を計算するために、`collecttag.rb` を使って Pixiv の新着投稿をしたユーザーの全イラストのタグを収集する。新着何人までのユーザーからタグを収集するかは `newtag.rb` と同じく `def new_arrival_works` の `per_page` の値を変更することによって変えることができるが、こちらも一定の値を超えると動かなくなってしまうため、複数回の収集が必要である。

### ⑤ 推薦対象者のタグ収集

推薦を行うユーザーの全イラストのタグを収集してくる。このときも `collecttag.rb` を実行して新着投稿をしたユーザーを対象にすればよいが、④のユーザーと重複しないように注意が必要である。

### ⑥ 類似度計算のための作品タグに置き換え

②の辞書と④のタグを入力にして `kanren.pl` を実行することでタグを全て作品タグに置き換える。重複は削除される。

※`kanren.pl` のプログラム上の注釈を入れた二行は削除する

※タグに [ や ( の文字が出てくるとプログラムが止まってしまうので、今回はその都度手作業で削除した。

### ⑦ 推薦対象者の興味度の測定

②の辞書と⑤のタグを入力にして `kanren.pl` を実行することで推薦対象者の興味度を計算する。⑥と違って重複は削除されないので作品タグの出現回数を手作業で数える。このとき、最も興味度の大きい作品が推定対象となるので2位以下の作品が⑧の入力となる。

※`kanren.pl` のプログラム上の注釈を入れた二行は削除しない

※タグに [ や ( の文字が出てくるとプログラムが止まってしまうので、今回はその都度手作業で削除した。

### ⑧ 推薦対象者の推定興味度の測定

⑦で求めた作品タグとその興味度を入力にして `ruijido.pl` を実行することで推定興味度を計算し、値が大きいものから順に出力する。⑦でもっとも興味度が高かった作品が何番目に出現しているかが評価結果となる。

## D 各種発表時に使用したパワーポイント

別紙で付録



# 参考文献

[1] 竹渕暎一 他, “Pixiv の二次創作イラストに含まれるジャンルタグの自動分類” 情報処理学会研究報告(GN), No.24,pp.1-5,2013

[2] 竹渕暎一 他, “Pixiv におけるキャラクタータグを検出する手法の検討” マルチメディア、分散協調とモバイルシンポジウム 2013 論文集,pp.84-88,2013

[3] 「情報推薦システムの基本 第5回 協調フィルタリング」  
<<http://gihyo.jp/dev/serial/01/information-recommendation-system/0005>>(2016/2/2 アクセス)

[4] 「Analyze It. 協調フィルタリングについてまとめてみた。」  
<<http://d.hatena.ne.jp/EulerDijkstra/20130407/1365349866>>(2016/2/2 アクセス)

[5] 「simanman のブログ Pixiv の API について調べてみた」  
<<http://simanman.hatenablog.com/entry/2013/02/11/222300>>(2016/2/3 アクセス)

[6] 「Hatena Developer Center」  
<<http://developer.hatena.ne.jp/>>(2016/2/3 アクセス)