

Twitter データを活用したラジオ番組推薦システムの検討

丸山 司[†] 岡本 一志[†] 柴田 淳司^{††}

[†] 電気通信大学情報理工学研究所 〒182-8585 東京都調布市調布ヶ丘 1-5-1

^{††} 東京都立産業技術大学院大学 〒140-0011 東京都品川区東大井 1-10-40

E-mail: [†]{tsukasa.maruyama,kazushi}@uec.ac.jp, ^{††}shibata-atsushi@aait.ac.jp

あらまし ラジオ番組では、リスナーに対し放送の感想ツイートを促す番組も多く、推薦システムへの活用に期待できるデータが豊富であるが、ラジオ番組推薦にソーシャルメディアを活用した研究は十分になされていない。本研究では、既存研究で提案されているソーシャルメディアを用いた推薦手法を Twitter データを用いたラジオ番組推薦に適用し、有効な推薦手法を明らかにする。ユーザのフォロー関係やツイートのテキスト、ツイートから抽出したユーザの関心を活用した手法の推薦精度を評価し比較する。実験では、ニッポン放送と TBS ラジオの 113 番組を対象に、2020 年 4 月から 2022 年 3 月のツイートを活用して評価実験を行う。

キーワード 推薦システム, ソーシャルメディア, ラジオ番組, ソーシャル推薦, 協調フィルタリング

1 はじめに

新型コロナウイルスの流行に伴う在宅勤務をはじめとした生活環境の変化から、ラジオ番組の聴取人数が増加している [1,2]. また、パソコンやスマートフォンでラジオ視聴ができる radiko¹ が登場した時期から、ラジオを聴く時間や機会が増えた人が増加している [3]. 従来のリアルタイム放送の番組から選択する以外に、radiko では 1 週間以内に放送された番組から選択し、好きなときに視聴できるタイムフリー機能により、番組選択の自由度が向上している。以上のような聴取者の増加と自由な番組選択を背景に、パーソナライズされたラジオ番組推薦のニーズが高まっているといえる。

10 代・20 代を中心にソーシャルメディアにラジオ番組の感想を投稿するユーザが増えている [3]. また、多くの番組が番組ウェブページや公式 Twitter アカウントで番組公式ハッシュタグを指定し、感想ツイートの投稿を促している。我々の調査では、2021 年の TBS ラジオとニッポン放送で放送されている番組のうち、120 番組が公式 Twitter アカウントや公式 HP, radiko の再生ページで公式ハッシュタグを指定している。

番組公式ハッシュタグを含むツイートをユーザの番組視聴と見なすことで評価ラベルを含めたデータセットを構築し、ユーザのツイートやフォロー関係から推薦モデルを構築して、ラジオ番組推薦の評価実験を行えると考えられる。一方で、ラジオ番組で放送される音楽などの情報からラジオ番組を推薦する研究 [4] があるが、ソーシャルメディアを活用したラジオ番組推薦に関する研究は十分になされていない。

本研究では、ラジオ番組に関するツイートやその投稿ユーザの過去のツイート、フォロー関係で構成されるデータセットを作成し、Twitter ユーザに対して番組を推薦するパーソナライズド推薦を想定する。既存研究で提案されているソーシャルメディア上のデータを活用した推薦手法をラジオ番組推薦に適用

し、各手法を評価することで、ラジオ番組推薦に適した手法を明らかにする。また、推薦結果やデータセットの分析から、ラジオ番組と Twitter データの関係性について考察する。評価実験では、ニッポン放送と TBS ラジオで 2021 年度に放送された 113 番組についてデータセットを作成し、各推薦手法についてランキング評価や多様性評価の指標を用いて推薦結果を評価する。

2 関連研究

本節では、ソーシャルメディアを活用した推薦手法について説明した後、ラジオ番組と類似メディアの推薦システムに関する研究について述べる。

2.1 ユーザ関係を活用した推薦

ソーシャルネットワークのユーザ間の信頼は親密度や興味・関心の類似により生み出されるとされている [5]. ユーザ間の親密度や信頼が高い場合は興味や関心の類似度も高いという仮説に基づき、ソーシャルメディア上のユーザ間の友人関係や信頼関係のユーザネットワークを活用した推薦研究が多く行われている [6-8]. このようなアプローチは、評価履歴からユーザの類似度計算が困難なコールドスタート問題に対して有効とされている。

Jeong らは、FaceBook²の親密な友人のアイテム評価値を集計するソーシャル推薦手法を提案し、従来の協調フィルタリング手法におけるコールドスタート問題を改善している [6]. 一般にユーザベース協調フィルタリングでは、推薦対象ユーザと他ユーザの類似度を計算し、興味・関心が近い他ユーザの評価値から推薦アイテムを決定する。Zarei らはソーシャルメディア上で推薦対象ユーザが信頼しているユーザと他ユーザの類似度計算を行うことで、特に評価アイテムが少ないユーザ（コールドスタートユーザ）に対する推薦精度の向上に成功してい

1 : <https://radiko.jp>

2 : <https://www.facebook.com/>

る [7]. Seo らは、ユーザ間の友好度をインタラクションによる親密度、ソーシャルネットワーク上の友人の類似度、投稿内容から取得する興味・関心の類似度により定義し、推薦に活用する協調フィルタリングモデルを提案している [9].

2.2 ユーザ生成コンテンツを活用した推薦

ユーザのコメントやツイートのユーザ投稿を用いた推薦手法の研究がある [10–13]. Lee らは、ユーザのツイート集合とニュースのテキストをそれぞれ文書とし、Term Frequency-Inverse Document Frequency (TF-IDF) を用いて文書ベクトルを作成、ユーザとニュース類似度計算を行うコンテンツベースフィルタリング (CBF) 手法を提案している [10]. Abel らは、ユーザのツイート中にあるハッシュタグやトピック、固有名詞の情報をを用いてユーザモデリングを行い、ニュース推薦に最も適している情報源を明らかにしている [12]. 林らは、レビューテキスト中のキーワードを肯定語・否定語に分割して抽出し、TF-IDF を用いて類似レビューを提示する映画推薦システムを構築している [13].

ユーザが生成するタグを用いた推薦が提案されている [14–16]. Konstas らは、Last.fm³の音楽推薦タスクについて、楽曲とタグ、ユーザをノードとする無向グラフを用いたグラフベース推薦を提案している [16]. ユーザが付与したタグや楽曲に付与されたタグ、ユーザが視聴した楽曲などに無向エッジを張り、Random Walk with Restart (RWR) を用いて、ユーザと楽曲のリンク予測を行うことで楽曲を推薦する。

2.3 ラジオ番組推薦とテレビ番組推薦に関する研究

テレビ番組推薦にソーシャルメディアを用いる研究がされている [17–20]. 澤井らは、NHK の番組について Twitter データを活用したテレビ番組推薦システムを構築している [17]. ハッシュタグ #NHK がついたツイートから、その投稿時間と同時刻の放送番組を視聴番組と想定し、投稿ユーザのフォロー情報と併せてデータセットを構築している。このデータセットを用いてカリスマユーザを中心に興味の伝播を考慮した協調フィルタリングによる番組推薦手法を提案している。

ラジオ番組推薦では、主に放送された音楽をもとにした推薦手法が提案されている [4, 21]. Ignatov らは、ラジオ番組で放送される音楽とユーザの音楽の好みからラジオ番組を推薦する手法を提案している [4]. しかし、日本のラジオ番組では番組の音楽に占める割合は小さく、適用は厳しいと考える。Jesus らは、放送される音楽に加え、出演者の話し方や番組のトークと音楽の割合、ユーザの好みのトピックや番組のトピックなどの複合的な補助情報を活用したラジオ番組推薦システムを構築している [21]. この研究では、ラジオ番組のトピックを手でラベル付けしていたり、事前のアンケートにユーザが好みを入力していたりとデータ取得のコストとユーザへの負担が大きいシステムとなっている。

太田らは [22], ラジオ番組の感想ツイートが多い時間を盛り上げられている箇所とし番組の放送音声を圧縮する手法を提案し

表 1: 検証手法の一覧

手法名	使用データ	モデル
User-based CF	視聴履歴のみ	CF
Friend-based CF	フォロワー	コサイン類似度
Trust-friend-based CF	フォロワー	CF
Friend-based Graph	フォロワー	RWR
Friend-group based CF	フォロワー	CF
Tweet-based CBF	ツイートテキスト	TF-IDF
Tweet-based CF	ツイートテキスト	TF-IDF, CF
Tweet-topic based CF	ハッシュタグ 人物名	CF, Jaccard 係数
Tweet-topic based Graph	ハッシュタグ 人物名	RWR

ており、ラジオ番組分野でも Twitter データの活用が目ざされ始めている。しかし、推薦システムに Twitter データを活用した研究は十分になされていない。

Twitter データでは、ユーザの投稿から番組視聴や関心のあつトピック、テキスト全体、ユーザネットワークから関心の近いユーザの情報をユーザへの負担やラベル付コストなしで取得できる。本研究ではこれらのデータに対して、ソーシャルメディアを使った既存の推薦手法をラジオ番組推薦に応用する。

3 検証する推薦手法

ユーザ u が視聴した番組の集合を P_u , フォロワー集合を F_u , ツイートの集合を T_u , 関心語を I_u とする。番組 $p \in P$ が公式にしているハッシュタグを h_p , 番組に関する口コミツイートの集合を T_p とする。本研究のタスクは、ユーザ u に対して、ユーザと番組に関する情報から、未来に視聴する番組集合 $P_{u'}$ を予測することである。予測対象の $P_{u'}$ は既に視聴済みの番組を除いた番組集合であり、 $P_u \cap P_{u'} = \phi, P_{u'} \subset P \setminus P_u$ である。推薦では、推薦時点で未視聴の番組 $p \in P \setminus P_u$ に対し、推薦スコア $s_{u,p}$ を計算し、推薦スコア順に並び替えた推薦番組リスト (順序集合) $Z_{u'}$ を作成する。評価では、 $P_{u'}$ と $Z_{u'}$ を比較し、ランキング評価を行うことで推薦精度を評価する。検証する推薦手法の一覧を表 1 にまとめる。

3.1 ユーザベース協調フィルタリング

ユーザ番組評価値行列 V を

$$V_{i,j} = \begin{cases} 1 & \text{ユーザ } i \text{ が番組 } j \text{ を視聴} \\ 0 & \text{ユーザ } i \text{ が番組 } j \text{ を未視聴,} \end{cases} \quad (1)$$

とする。ユーザ番組評価値行列 V に対して、アイテムベース CF を適用し、推薦スコアを計算する。ユーザの類似度計算には、コサイン類似度

$$\text{sim}(u_a, u_b) = \text{cossim}(V_a, V_b) \quad (2)$$

を使用する。本研究では、この手法を User-based CF と呼ぶ。

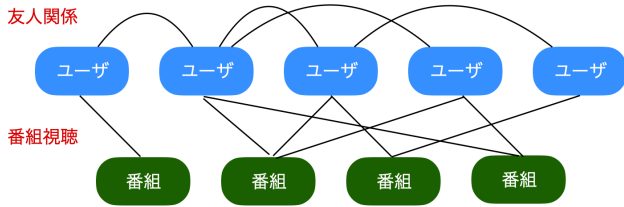


図 1: Friend-based Graph の構造の概形

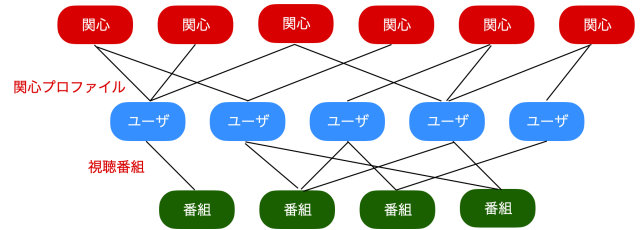


図 2: Tweet-topic based Graph の構造の概形

3.2 友人の評価値を利用した協調フィルタリング推薦

Jeong らの手法 [6] と Zarei らの手法 [7] を検証し、本研究では、それぞれ Friend-based CF, Trust-friend based CF として説明する。Friend-based CF では、友人 $f \in F_u$ の評価値の平均を計算し、番組の推薦スコア

$$s_{u,p_j} = \frac{\sum_{f \in F_u} V_{f,j}}{|F_u|} \quad (3)$$

とする。Trust-friend based CF では、推薦対象ユーザ u の信頼しているユーザ（本研究では友人） F_u との類似度計算

$$sim_t(u, u_t) = \frac{\sum_{f \in F_u} sim(f, u_t)}{|F_u|} \quad (4)$$

により、ユーザ類似度を定義し、ユーザベース協調フィルタリングを行う。

3.3 友人ネットワークを利用したグラフ推薦

本手法を本研究で Friend-based Graph と呼ぶ。この手法は、Konstas らのタグとユーザ、アイテムのグラフを用いる推薦手法 [16] からタグを除いた簡易的なグラフベース手法である。ユーザと番組をノードとし、ユーザが視聴した番組との間に重み 1 のエッジ、ユーザとフォロワーに重み 1 のエッジを持つ無向グラフ G を構築する。グラフ G の構造を図 1 に示す。本来 Twitter のフォロー関係は一方方向の関係であり、有向エッジにするべきだが、データ収集の都合からフォロワーの収集が困難なため無向エッジとしている。

グラフ G で各ノードからのエッジの重みの合計が 1 になるように正規化を行い、隣接行列 A を作成する。リスタート率 c としてユーザ u を始点、 r_u をノードの定常確率ベクトル、 q_u を u に対応するノードを 1、それ以外を 0 とする one-hot ベクトルとする。RWR によりノードの定常確率

$$r_u = (1 - c)Ar_u + cq_u \quad (5)$$

を計算する。番組 p に該当するノードの定常確率を推薦スコア $s_{u,p}$ とする。

3.4 友好度を計算し利用する協調フィルタリング推薦

Seo ら [9] の検証する手法のうち、ユーザコミュニティの類似度計算を用いた協調フィルタリングを Friend-group based CF とする。ユーザの友人集合を Jaccard 係数によりユーザ類似度

$$sim(u_a, u_b) = \frac{|F_{u_a} \cap F_{u_b}|}{|F_{u_a} \cup F_{u_b}|} \quad (6)$$

を計算し、ユーザベース協調フィルタリングにより推薦スコア $s_{u,p}$ を計算する。

3.5 ツイートテキストを利用したコンテンツベース推薦

Lee らの手法 [10] を本研究では、Tweet-based CBF と呼ぶ。ユーザ u のツイート集合 T_u を 1 つの文書と見なし、番組 p についてのツイート集合 T_p も同様に 1 つの文書と見なす。TF-IDF を用いて、文書ベクトル v_u, v_p を作成する。文書ベクトルのコサイン類似度 $cosim(v_u, v_p)$ を推薦スコア $s_{u,p}$ とする。

3.6 ツイートテキストを利用した協調フィルタリング推薦

Tweet-based CBF と同様に、ユーザ u のツイート集合 T_u に対し、TF-IDF により文書ベクトル v_u を作成する。ユーザ間の類似度計算を文書ベクトルの類似度 $sim(u_a, u_b) = cosim(v_{u_a}, v_{u_b})$ により定義し、ユーザベース協調フィルタリングにより推薦スコア $s_{u,p}$ を計算する。本研究では、Tweet-based CF と呼ぶ。

3.7 ツイートの関心語を利用した協調フィルタリング推薦

Seo ら [9] が検証している手法のうち、ユーザ投稿内のトピックを用いた協調フィルタリングを Tweet-topic based CF として適用する。トピックの抽出には、Abel ら [9, 12] の検証に従い、ツイートの固有名詞、ハッシュタグの 2 種類について比較する。ユーザ u のツイート集合 T_u から、固有名詞とハッシュタグをユーザの関心語 I_u として抽出する。固有名詞には人物名を使用する。関心語 I_u の出現回数による Bag of Words (BoW) から特徴ベクトル v_{I_u} を作成する。特徴ベクトルのコサイン類似度によりユーザ類似度計算を定義し、ユーザベース協調フィルタリングにより、推薦スコア $s_{u,p}$ を計算する。

3.8 ツイートの関心語を利用したグラフ推薦

Konstas らの手法 [16] を、本研究では、Tweet-topic based Graph と呼ぶ。Tweet-topic based CF と同様に固有名詞とハッシュタグを用いて関心語 I_u を考える。ユーザと番組、関心語をノードとし、ユーザが視聴した番組との間に重み 1 のエッジ、ユーザと関心語に重み 1 のエッジを持つ無向グラフ G を構築する。グラフ G の構造を図 2 に示す。

グラフ G から、各ノードからのエッジの重みの合計が 1 になるように正規化を行い、隣接行列を作成する。各ユーザ u を始点とした RWR を行い、番組 $p \in P$ の定常確率を計算し、推薦スコア $s_{u,p}$ とする。

表 2: テストデータセットの概要

番組数	113	
ユーザ	実験ユーザ 8331 人 評価ユーザ 6486 人	
実験ユーザ内の友人数	平均 1.48 人	標準偏差 3.43 人
実験ユーザ内の友人有無	有 4332 人	無 3999 人

4 実験データセットの構築

ニッポン放送と TBS ラジオの 113 番組の視聴ユーザを収集し、各ユーザのユーザプロフィールを収集する。表 2 に収集した番組や実験ユーザ、友人に関する概要を示す。

4.1 視聴者の収集

各番組は公式のハッシュタグを指定しており、そのハッシュタグをつけてツイートした場合そのユーザがそのツイートの投稿日時で該当番組を視聴したと見なす。また、番組公式アカウントをフォローしている場合もユーザ作成時に番組視聴し始めたとする。2021 年 4 月 1 日から 2022 年 3 月 31 日に対象のいずれかの番組を視聴したユーザを収集する。このユーザのうち、2020 年 4 月 1 日から 2022 年 3 月 31 日のツイート件数が 2000 件以下のユーザの中からランダムに選択したユーザを、実験で使用する実験ユーザとして使用する。

4.2 ユーザプロフィールの収集・構築

各実験ユーザについて、(1) 2020 年 4 月 1 日から 2022 年 3 月 31 日に投稿したツイート、(2) 収集時点でのフォロワー、を収集する。TwitterAPI では、ツイートテキスト中の人物名やハッシュタグを抽出して取得できる。この情報を Tweet-topic based CF や Tweet-topic based Graph で用いる関心語 I_u に使用する。

4.3 番組プロフィールの収集・構築

各番組について、番組指定のハッシュタグを含む 2020 年 4 月 1 日から 2022 年 3 月 31 日に投稿したツイートを収集し、番組に関するコメントツイートとする。

5 評価実験

5.1 実験設定

本実験では番組が指定するハッシュタグを含むツイートがある場合、その番組を視聴したと見なす。多くのユーザが継続してラジオ番組を視聴しているため、過去に視聴した番組を再度推薦することはほとんどユーザにとって有益ではない。そのため、3 節で説明したとおり、本実験ではユーザが過去に視聴していない番組を並び替えて推薦リストを作成する。

データ収集期間の 2020 年 4 月 1 日から 2022 年 3 月 31 日で、次のように分割してデータセットを構築する。まず、2020 年 4 月 1 日から 2021 年 3 月 31 日を学習期間、2021 年 4 月 1 日から同年 4 月 30 日を評価期間とする。各実験ユーザについて、学習期間に視聴していない番組のうち、評価期間に初めて

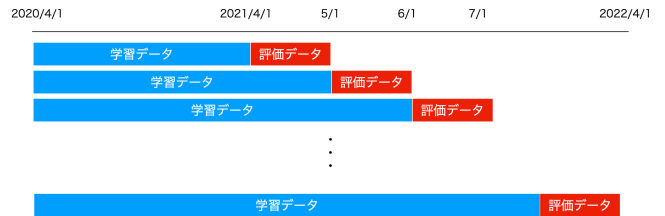


図 3: データセット分割の略図

表 3: 推薦精度と多様性

	Precision@5	MRR@10	Recall@10	Diversity@10
User-based CF	0.0529	0.151	0.322	0.959
Tweet-based CF	0.0712	0.247	0.412	0.954
Tweet-topic based CF (person)	0.0874	0.284	0.519	0.954
Tweet-topic based CF (hashtag)	0.0896	0.290	0.539	0.955
Friend-based CF	0.0492	0.150	0.303	0.964
Trust-friend based CF	0.0661	0.195	0.402	0.958
Friend-group based CF	0.104	0.334	0.610	0.953
Friend-based Graph	0.103	0.339	0.621	0.954
Tweet-topic based Graph (person)	0.0854	0.281	0.529	0.953
Tweet-topic based Graph (hashtag)	0.0920	0.298	0.557	0.954
Tweet-based CBF	0.0317	0.118	0.201	0.977

視聴した番組がある場合、そのユーザを評価ユーザとし、推薦精度の評価に用いる。学習期間、評価期間を 1ヶ月ずつづらしながら、図 3 のように 12 個のデータセットを構築する。

5.2 評価指標

推薦精度の計算には、Precision@ k 、Recall@ k 、MRR (Mean Reciprocal Rank) @ k の 3 つのランキング評価指標を用いる。 k は推薦リストの要素数を意味する。多様性評価 $Diversity@k$ には、推薦リスト上位 k 個の類似度の逆数により

$$Diversity(P_{u,z}, k) = \frac{\sum_{i \in \text{top}(P_{u,z}, k)} \sum_{j \in \text{top}(P_{u,z}, k) \setminus i} 1/\text{sim}(i, j)}{k(k-1)} \quad (7)$$

のとおり計算する。アイテムの類似度計算には評価値ベクトルのコサイン類似度を用いる。

5.3 推薦精度と多様性の評価結果

推薦リストサイズ毎の推薦精度および多様性の評価結果を図 4, 5 に示す。推薦リストのサイズ k を 3 から 10 にの範囲で評価し、図の k を横軸、各指標の評価値を縦軸としている。全手法における推薦リストのサイズを 10 としたときの各指標の評価値を表 3 に示す。Precision, Recall, MRR のいずれでも Friend-group based CF と Friend-based Graph で高い推薦精度となっている。関心語情報を使った手法である Tweet-topic based CF (person, hashtag), Tweet-topic based Graph (person, hashtag) が一定程度の推薦精度となっている。

また、Friend-based CF, Trust-friend based CF, Tweet-based CBF が推薦精度と同様に低い多様性の評価結果となっている。Tweet-based CF が推薦精度に反して、高い多様性の評価結果となっている。

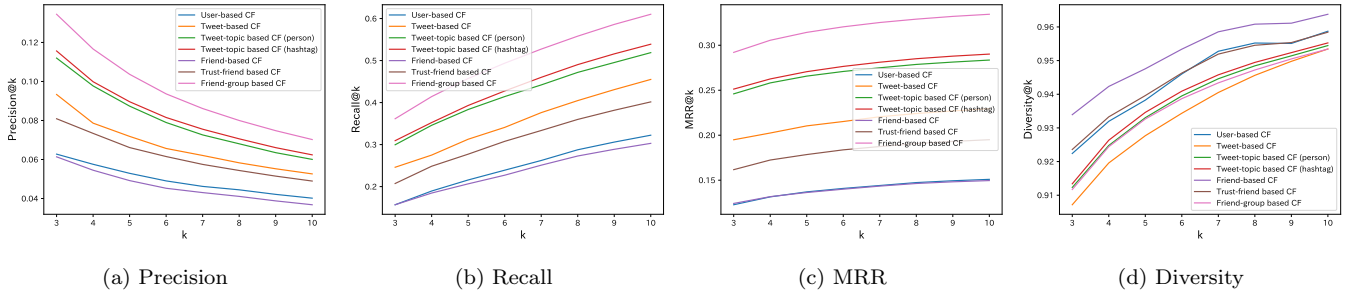


図 4: 推薦精度と多様性の評価結果 (CF)

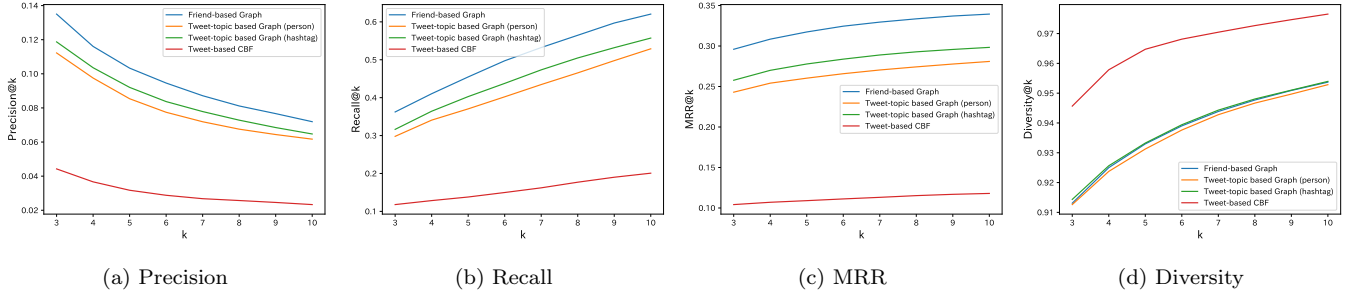


図 5: 推薦精度と多様性の評価結果 (Graph と CBF)

6 分析・考察

6.1 補助情報と番組視聴の関連

Friend-group based CF と Tweet-topic based CF では、Jaccard 係数を用いるか BoW ベクトルのコサイン類似度を用いるかの違いはあるが、いずれも補助情報の類似度からユーザ類似度を求める協調フィルタリングである。そこで、フォロワー (Friend-group based CF) と人物名、ハッシュタグ (Tweet-topic based CF) の補助情報が、番組視聴とどの程度関連しているかを分析する。

各補助情報から番組の TF-IDF 特徴ベクトルを作成し、番組類似度 sim_v を計算する。評価値ベクトルによる番組類似度 sim_r との相関を求め、ユーザ u 毎に補助情報を数え上げ、BoW ベクトル \mathbf{b}_u を作成する。全ユーザの BoW ベクトルにより、TF-IDF モデルを構築する。各ユーザの TF-IDF ベクトルを \mathbf{v}_u とし、番組 p の視聴者 $u \in U_p$ について集計することで、番組の TF-IDF 特徴ベクトル

$$\mathbf{v}_p = \frac{\sum_{u \in U_p} \mathbf{v}_u}{|U_p|} \quad (8)$$

を計算する。番組 p の TF-IDF ベクトル \mathbf{v}_p 、評価値ベクトル \mathbf{r}_p について、いずれもコサイン類似度を用いて

$$sim_v(p_i, p_j) = \text{cossim}(\mathbf{v}_{p_i}, \mathbf{v}_{p_j}) \quad (9)$$

$$sim_r(p_i, p_j) = \text{cossim}(\mathbf{r}_{p_i}, \mathbf{r}_{p_j}) \quad (10)$$

を計算する。

図 6 に、各補助情報による番組類似度を横軸、評価値ベクトル

を用いた類似度を縦軸にプロットしている。用いる番組は、実験ユーザで視聴者が 50 人以上いる 66 番組に限定し、それらの番組同士の類似度を計算している (同一番組の類似度はプロットから除外している)。補助情報を用いた類似度と評価値ベクトルを用いた類似度の相関係数は、フォロワーで 0.840、ハッシュタグで 0.530、人物名で 0.403 となっている。

以上の結果から、補助情報のうち、フォロワーが最もラジオ番組視聴と関連が強く、視聴履歴が不十分な場合に、視聴履歴の代わりにユーザプロフィールとして有効であるといえる。

6.2 友人データを用いた推薦手法の比較

Friend-based CF と Trust-friend based CF では、友人の評価値を集計したり、友人の評価値と実験ユーザの評価値を比較したりと、全ての友人アカウントを活用できるわけではない。評価値を持つユーザ、つまり実験ユーザ内の友人に限り活用できる。実験ユーザ 8331 人のうち 3999 人と、半数近くの実験ユーザが実験ユーザ内に友人を持たない。この 2 つの手法は、実験ユーザ内の友人がいない場合、User-based CF と同様の推薦となるため、多くのコールドスタートユーザに補助情報が提供できず、推薦精度が他の手法と比べ低くなったと考える。実験ユーザ数を拡大すれば解決できるとも考えられるが、データの収集コストは高くなる。ソーシャルメディアを活用した推薦では、データ収集コストを考慮した効率の良い推薦モデルが求められる。

一方、推薦精度が優れる Friend-group based CF と Friend-based Graph では、共通のフォロワーを通じて実験ユーザと接続できるモデルである。視聴履歴を持たなくてもフォロワーアカウント単体で推薦モデルに活用でき、実験ユーザとのつながり

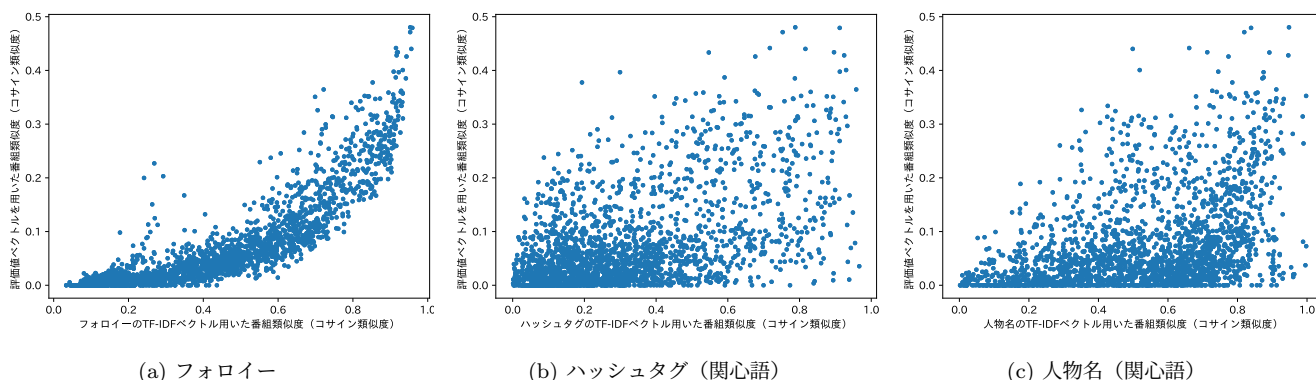


図 6: 補助情報を用いた番組類似度と評価値ベクトルを用いた番組類似度の相関

りの強さを計算するノードとして役割を果たすことができる。

友人データを用いた推薦手法では、直接評価値を用いる手法 (Friend-based CF や Trust-friend based CF) より、フォロワーを介して実験ユーザとのつながりを計算し、その評価値を用いることができる手法 (Friend-based Graph と Friend-group based CF) が友人データを効率よく活用でき、推薦精度の優れる手法であると考えられる。

6.3 実験ユーザサイズと推薦精度の関係

前節で Friend-group based CF と Friend-based Graph は視聴履歴がない友人を活用できる効率の良いモデルであると述べた。本節では実験ユーザが少ないデータセットで Friend-group based CF の推薦精度を計測し、効率の良いモデルであるという仮説を検証する。Friend-group based CF について、実験ユーザ数 8331 (実験ユーザ全体)、5000、3000、1000、500 人をランダムにサンプリングしてデータセットを作成し、推薦精度を比較する。推薦リストのアイテム数 k を 3 から 10 について推薦精度を計測し、実験ユーザサイズ毎に評価結果を図 7 に示す。

推薦精度を $MRR@10$ で比較すると、実験ユーザが 500 人の小規模なデータセットを用いた Friend-group based CF の推薦精度は 0.311 であり、表 3 で示した全ての実験ユーザを用いた Tweet-topic based CF や Friend-based CF の推薦精度より優れている。以上の結果から、Friend-group based CF では少ない実験ユーザでも有効な推薦ができるモデルであるといえる。小規模なデータセットでも推薦が可能であることは、Friend-group based CF の元となった Seo らの文献 [9] で主張されていない内容である。

6.4 推薦に有効な友人層の分析・考察

Friend-group based CF では、コミュニティの類似度を用いてユーザの類似度を計算し推薦している。本節では、コミュニティの類似度計算でどのようなアカウントが推薦に有効であるかを考察する。Friend-group based CF では、有効な友人として 2 つの仮説がある。

仮説 (1) は、フォローしている有名人がそのユーザの関心データとなっており、ラジオ番組と直接結びついている仮説である。ラジオ番組は、番組ジャンルは少なく、パーソナリティ

が誰であるかが番組を大きく特徴付けられる。トレンド総研 [3] による調査でも、アーティスト・タレントなどと密接に関わるメディアとして考える視聴者も多い。

仮説 (2) は、共通のフォロワーを持つ実験ユーザは、友人の友人のように、親密度の高い潜在的な友人に近い場合である。ユーザ間の親密度や信頼が高い場合、興味や関心の類似度も高いという仮説 [5] と同様に、コミュニティの類似度で計算したユーザ類似度は潜在的な友人となる確率と近く、友人の評価値を利用して推薦ができているという仮説である。これは、Friend-group based CF のベースとなる Seo らの手法 [9] で、前提とされている考え方である。

6.4.1 フォロワー数による友人セグメントの精度比較

フォロワー数による友人データの 7 つのセグメント (1~10 人、10~100 人、100~1K 人、1K~10K 人、10K~100K 人、100K~1M 人、1M 人~) を作成し、セグメントの友人データセットによる推薦精度の差を比較する。フォロワー数の多いセグメントの精度が優れば、仮説 (1) の有名人の関心として利用しており、少ないセグメントの精度が優れば、仮説 (2) の友人関係を利用している可能性が高いと考える。

ここで実験ユーザのフォロワーが属するセグメントの分布を確認する。実験ユーザについて、フォロワーが 7 つの区分にどのように分布しているかを図 8 に示す。100~1K、1K~10K、10K~100K、100K~1M のセグメントでは、フォロワーの割合が同程度になっており、これらのセグメントでは、ボリューム差による影響を考慮せず比較する。各セグメントの友人データのみを用いて推薦を行い、推薦精度と多様性を図 9 に示す。

$MRR@k$ により推薦精度を比較すると、10K~100K 人のセグメントが最も優れ、次点で 1K~10K、100K~1M のセグメントで優れた推薦精度となっており、 $Recall@k$ や $Precision@k$ でもおおそ同様の推薦精度の順位になっている。次節で各セグメントに属するフォロワーについて定性的に分析する。

6.4.2 フォロワーの定性的な特徴分析

佐久間宣行の ANN0 について、6.1 項と同様に TF-IDF を作成し、フォロワー数のセグメント毎に重要なアカウントを計算する。表 4-6 に、各セグメントの TF-IDF 値の上位 20 件についてリストアップしている。表では、番組公式アカウントや

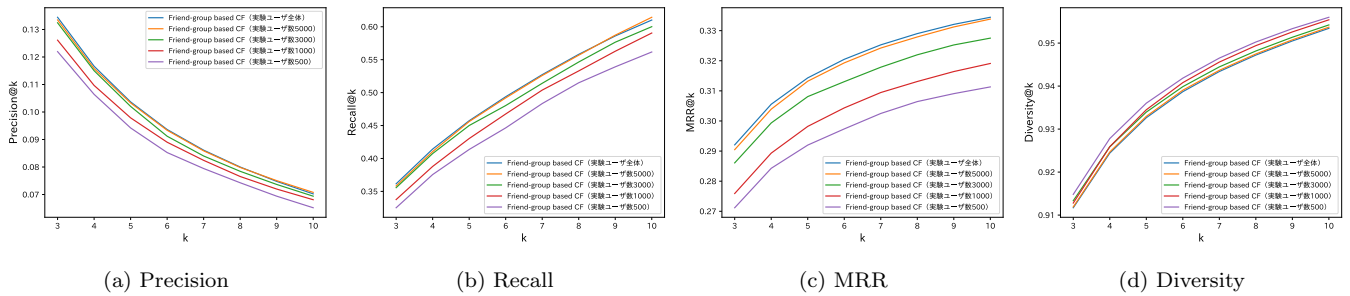


図 7: 実験ユーザサイズを変化させたときの推薦精度と多様性の推移

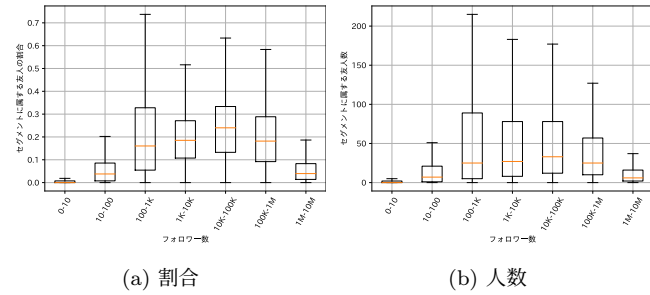


図 8: 実験ユーザの友人のフォロワー数の分布

表 4: 「佐久間宣行の ANN0」視聴者のフォロワー (1K-10K)

TF-IDF 値	表示名	フォロワー数	番組の視聴回数 からのフォロワー	パーソナリティ からのフォロワー
0.121	ジャイアントあつひこ	7599	2166	-
0.0970	天草大王	4969	19	-
0.0956	金子 司	9572	73	-
0.0951	論より証拠の大ライス	9685	0	-
0.0932	ポテチ食べたい、あとモテたい	6568	284	-
0.0927	サクマ&ピース【公式】	6551	0	○
0.0892	サハラサザンカ	3551	437	-
0.0838	福田 眞太	9991	0	-
0.0828	徳泉	3183	0	-
0.0819	げすば	5948	0	-
0.0817	よもぎもち	5983	165	-
0.0816	岳山健	8620	0	-
0.0812	ハートマト	3908	164	-
0.0807	さらば青春の光が Ta ダ、Ba カ、Sa ワギ...	8515	0	-
0.0804	それはもう白石 TENGA	2762	234	-
0.0789	(´ ˘ ˘) 3は二十歳になってから	3651	780	-
0.0776	一蘭食ってまた一蘭	4516	179	-
0.0765	アルミ缶	3150	1648	-
0.0763	駄作	5734	50	-
0.0761	マイナビニュース・ラジオ【公式】	5526	92	-

パーソナリティのアカウントからフォローされているかと、該当番組ハッシュタグの含むツイートをした回数を視聴回数として、併せて記載している。100-1K, 1K-10K のセグメントでは、番組の視聴回数が多い（公式ハッシュタグを含めた番組のコメントを頻繁に投稿している）アカウント（以下コアラリスナーと呼ぶ）が多い。10K-100K, 100K-1M のセグメントでは、番組パーソナリティと関連する有名人と思われるアカウントが多い。

5 節の結果と本節の分析をふまえると、有名人のフォロワー (10K~のセグメント) を関心として用いた推薦手法がラジオ番組推薦に最も有効であり、コアラリスナーのフォロワー (1K-10K セグメント) を用いた推薦も一定程度有効なことがわかる。

7 おわりに

本研究は、ラジオ番組推薦に Twitter データを活用した推薦手法を適用し、推薦精度の比較とデータセットの分析を行なっ

表 5: 「佐久間宣行の ANN0」視聴者のフォロワー (10K-100K)

TF-IDF 値	表示名	フォロワー数	番組の視聴回数	公式アカウント からのフォロワー	パーソナリティ からのフォロワー
0.224	佐久間宣行のオールナイトニッポン0【公式】	80307	1068	-	○
0.157	福田卓也	51974	37	○	○
0.143	石井 玄 (ひかる)	62862	441	○	○
0.131	三四郎・相田周二	92788	4	○	○
0.126	三四郎のオールナイトニッポン0【毎週金曜深夜 27 時〜】	57054	24	○	-
0.125	あちこちオードリー【水曜よる 11:06 テレビ東京系列】	83008	0	-	○
0.119	宮崎 守史	77312	0	-	○
0.109	デレクジカー	47393	10	-	○
0.102	佐久間宣行の NOBROCKTV	18662	1	-	○
0.101	飯塚	43581	14	-	-
0.099	マチカルラブリーのオールナイトニッポン0【公式】	54820	2	-	-
0.0986	霜降り明星のオールナイトニッポン 金曜 25 時	82971	12	○	-
0.0985	JUNK (TBS ラジオ)	46243	0	-	-
0.0905	ハイチのターン! (TBS ラジオ)	50660	0	-	-
0.0905	たりないふたり-山里亮太と若林正恭-	69053	0	-	○
0.0874	アルコ&ピース D.C.GARAGE (TBS ラジオ)	26929	0	-	-
0.0841	沈黙の金曜日	63066	0	-	-
0.0834	フワちゃんのオールナイトニッポン0	38567	1	-	-
0.0822	『ラジオの時間』編集部	25328	68	-	-
0.0815	人気ラジオ番組完全ガイド&サマ編集部公式	12636	75	-	-

表 6: 「佐久間宣行の ANN0」視聴者のフォロワー (100K-1M)

TF-IDF 値	表示名	フォロワー数	番組の視聴回数	公式アカウント からのフォロワー	パーソナリティ からのフォロワー
0.221	佐久間宣行	330405	725	○	-
0.162	オードリー-のオールナイトニッポン	228634	16	○	○
0.153	オールナイトニッポン	187481	426	○	○
0.152	DJ 松永 (Creepy Nuts)	300883	23	○	○
0.145	Creepy Nuts のオールナイトニッポン【毎週月曜 25 時〜】	123008	41	○	-
0.138	Creepy Nuts	299922	12	-	-
0.132	アルコ&ピース平子	275090	0	-	○
0.132	岩井勇気 ハイイチ	551833	0	-	○
0.124	さらば青春の光 森田哲矢	417750	0	-	○
0.116	佐藤潤春	102348	4	-	-
0.111	マチカルラブリー 野田クリスタル	303122	0	-	○
0.111	藤井健太郎	185045	5	○	○
0.109	M-1 グランプリ	532392	0	-	-
0.108	粗品 (霜降り明星)	888880	0	-	○
0.107	東京 03 飯塚	292389	6	○	○
0.0992	霜降り明星 せいや	502372	0	-	○
0.0984	日向坂 46	825800	0	-	○
0.0941	お笑いナタリー	324778	25	-	○
0.0865	乃木坂 46 のオールナイトニッポン【公式】	193378	16	○	-
0.0865	日向坂で会いましょう	400151	0	-	-

ている。推薦精度の比較から、Twitter 上のユーザコミュニティを活用した協調フィルタリングやグラフベース手法が最も推薦精度が優れたことを明らかにしている。分析結果から、ラジオ番組推薦ではツイートから抽出する関心情報より、有名人の関心を表現するフォロワーデータが番組視聴と類似する補助情報であるといえる。これは、ラジオ番組が、他のドメイン（テレビや映画、音楽など）に比べ、出演者やその関連する有名人が、視聴行動に密接に関わっている性質があるためと考える。先行研究のソーシャルメディアを用いた推薦で、フォロワーを有名人の関心として活用する提案はされておらず、これは本研究の貢献の 1 つである。

今後は、ラジオ番組の Twitter 上のユーザコミュニティに關する更なる分析や、有名人のネットワークを用いた推薦手法の提案を考えている。

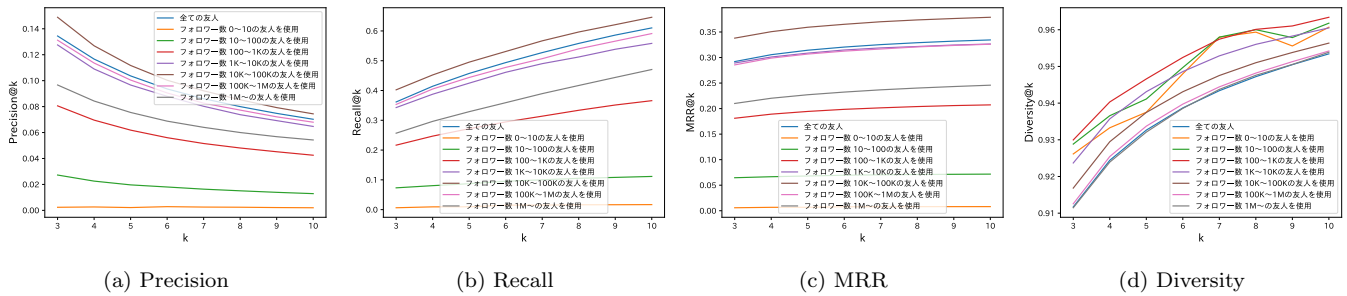


図 9: 友人のフォロワーの数による推薦精度と多様性の推移

謝 辞

本研究は JSPS 科研費 JP21H03553 の助成を受けたものです。

文 献

- [1] TimesDigital Shift. “コロナ禍でラジオが復権!? 民放ラジオ業界 70 年の歴史を塗り替えた radiko (ラジコ) の「共存共栄型 dx」とは”. https://digital-shift.jp/dx_strategy/210708, 7 2021. (Accessed on 2022/12/1).
- [2] ビデオリサーチ. “新たな生活環境下でラジオリスナーが増加傾向”. <https://www.videor.co.jp/press/2020/200625.html>, 6 2020. (Accessed on 2022/12/1).
- [3] トレンド総研. “10~60 代の「ラジオ聴取」実態を調査”. https://kyodonewsprwire.jp/prwfile/release/M102314/201506301561/_prw_0R1f1_h6281jD0.pdf, 6 2015. (Accessed on 2022/12/1).
- [4] D. I. Ignatov, S. I. Nikolenko, T. Abaev, and J. Poelmans. “Online recommender system for radio station hosting based on information fusion and adaptive tag-aware profiling”. *Expert Systems with Applications*, Vol. 55, pp. 546–558, 2016.
- [5] M. Kaptein, C. Nass, and P. Markopoulos. “The effects of familiarity and similarity on compliance in social networks”. *International Journal of Internet Marketing and Advertising*, Vol. 8, p. 222, 01 2014.
- [6] O.-R. Jeong. “Sns-based recommendation mechanisms for social media”. *Multimedia Tools Appl.*, Vol. 74, No. 7, p. 2433–2447, apr 2015.
- [7] M. R. Zarei and M. R. Moosavi. “A memory-based collaborative filtering recommender system using social ties”. In *2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA)*, pp. 263–267, 2019.
- [8] L. Chao, Y. Jian, L. Xiang, and C. J. Hui. “A social network system oriented hybrid recommendation model”. In *Proceedings of 2012 2nd International Conference on Computer Science and Network Technology*, pp. 901–906, 2012.
- [9] Y.-D. Seo, Y.-G. Kim, E. Lee, and D.-K. Baik. “Personalized recommender system based on friendship strength in social network services”. *Expert Systems with Applications*, Vol. 69, pp. 135–148, 2017.
- [10] W.-J. Lee, K.-J. Oh, C.-G. Lim, and H.-J. Choi. “User profile extraction from twitter for personalized news recommendation”. In *16th International Conference on Advanced Communication Technology*, pp. 779–783, 2014.
- [11] O. Phelan, K. McCarthy, M. Bennett, and B. Smyth. “Terms of a feather: Content-based news recommendation and discovery using twitter”. In P. Clough, C. Foley, C. Gurrin, G. J. F. Jones, W. Kraaij, H. Lee, and V. Mudoch, editors, *Advances in Information Retrieval*, pp. 448–459, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- [12] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. “Analyzing user modeling on twitter for personalized news recommendations”. In J. A. Konstan, R. Conejo, J. L. Marzo, and N. Oliver, editors, *User Modeling, Adaption and Personalization*, pp. 1–12, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- [13] 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*, Vol. 30, No. 1, pp. 102–111, 2015.
- [14] D. Sánchez-Moreno, M. N. Moreno-García, N. Sonboli, B. Mobasher, and R. Burke. “Using social tag embedding in a collaborative filtering approach for recommender systems”. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pp. 502–507, 2020.
- [15] D. Yang, L. Chen, J. Liang, Y. Xiao, and W. Wang. “Social tag embedding for the recommendation with sparse user-item interactions”. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 127–134, 2018.
- [16] I. Konstantas, V. Stathopoulos, and J. M. Jose. “On social networks and collaborative recommendation”. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '09*, p. 195–202, New York, NY, USA, 2009. Association for Computing Machinery.
- [17] 澤井里枝, 有安香子, 藤沢寛. “Sns を利用した協調フィルタリングによる番組推薦手法”. *情報処理学会研究報告*, 2010 年度, No. 4, pp. 1–8, 12 2010.
- [18] 袴田裕紀. “マイクロブログを用いた番組推薦に関する研究”. 第 80 回全国大会講演論文集, Vol. 2018, No. 1, pp. 279–280, 03 2018.
- [19] Y. Zhang, W. Chen, and Z. Yin. “Collaborative filtering with social regularization for tv program recommendation”. *Knowledge-Based Systems*, Vol. 54, pp. 310–317, 2013.
- [20] S. Pyo, E. Kim, and M. Kim. “Lda-based unified topic modeling for similar tv user grouping and tv program recommendation”. *IEEE Transactions on Cybernetics*, Vol. 45, pp. 1476–1490, 2015.
- [21] A. J. Fernández-García, R. Rodríguez-Echeverría, J. C. Preciado, J. Perianez, and J. D. Gutiérrez. “A hybrid multidimensional recommender system for radio programs”. *Expert Systems with Applications*, Vol. 198, p. 116706, 2022.
- [22] 太田暢, 村山太一, 矢田峻太郎, 荒牧英治, 若宮翔子. “Twitter を用いたラジオ番組圧縮手法の提案”. *deim2022 最終論文集 DEIM*, 2022.