

ニコニコ生放送配信者の可視化と推薦 Visualization and Recommendation of Niconico Live Broadcasters

英 良治

Ryoji Hanabusa

法政大学情報科学部デジタルメディア学科

E-mail: ryoji.hanabusa.6w@stu.k.hosei.ac.jp

Abstract

This paper presents a system that supports users of a video streaming service called Niconico Live in searching for programs and also that visualizes relations among broadcasters. Although Niconico Live provides a search engine based on tags and keywords, many broadcasters label their programs with inconsistent tags and titles, which makes the search engine almost useless. To solve the problem, this paper proposes a method that combines visualization and recommendation. It visualizes similarities of broadcasters by using viewing histories of other users, which is done by multi-dimensional scaling. This method uses two types of recommendation that are item-based and user-based. The item-based recommendation uses the result of visualizing similar broadcasters. By contrast, the user-based recommendation computes users with similar viewing histories to recommend broadcasters. The combination of visualization and recommendation allows users to get to know, not only actual broadcasters for them to watch, but also their preferences for broadcasters.

1. はじめに

大手動画共有サイトであるニコニコ動画では、ユーザーがリアルタイムに配信を行い、視聴者がコメントを投稿できるニコニコ生放送 [1] というサービスを提供している。現在、ニコニコ生放送ではピーク時で同時刻に 5000 人もの配信者が生放送を行っているが、一般的にユーザーは生放送の中から数人の配信者のものだけを視聴している。ニコニコ生放送ではユーザーが配信 URL に接続した時点で画面上の来場者数にカウントされるため、配信の内容に関係なくとも、よりクリックされやすいタグを配信者が設定する風潮が多く見られる。そのため、キーワードやタグにより番組を検索するシステムが提供されているが、配信内容とタグが一致しているものが少なく、ユーザーの好みの配信を見つけるのが困難である。

ちくわちゃんランキング [2] では、ニコニコ生放送の配信の盛り上がり度を示すアクティブ数による番組推薦を行っている。このサービスは公式の番組検索システムよりも優れているが、リアルタイムで配信を行っている配信者のみのランキング表示であることと、配信の内容に関係の無い表示をしていることから、人気のある配信者

を見つけることはできても、好みの配信を見つけるのは困難である。

この問題を解決するために、本研究では配信者の配信内容やジャンルの集合を視覚的に把握できるよう、配信者間の繋がりの可視化を行う。さらに直感的な番組推薦を行うため、推薦結果を可視化結果上に組み込んで提示する。

2. 関連研究

ニコニコ動画を対象とした研究は、動画タグを用いた動画間の関係を可視化するものや、動画に付けられたコメントから動画内容を分析するものなど、いくつか行われている。しかし、ニコニコ生放送を対象とした研究は行われていない。

推薦に関する研究として、人の好みに合わせた音楽を表示する楽曲推薦がある。推薦方法としては、内容に基づくフィルタリングと協調フィルタリングが存在する。内容に基づくフィルタリングは、曲自体の分析を行って曲調の似た曲を推薦し、協調フィルタリングは、曲を購入したユーザの購入履歴を元に推薦する。しかし、協調フィルタリングでは利用しているユーザの多い曲ばかりが推薦されてしまい、新しい曲の発見といった面が薄れてしまう問題が生じる。

また、楽曲推薦の研究として、推薦と可視化を組み合わせた Songrium [3] がある。このシステムはインターネット上で流行している音楽の特徴と他の曲との関係性を可視化する機能と、さらにユーザーが付加した矢印タグによる楽曲推薦機能を提供し、ユーザーの好みの楽曲を視覚的に見つけられるようにしている。

3. 準備

3.1. 多次元尺度法 (MDS)

多次元尺度法とは多変量解析の一種であり、対象間の距離あるいは類似度を空間上の距離に置き換え、二次元平面上での相対的位置関係を求める手法である。可視化する行列が各ノード間の距離で構成されている場合

$$\text{距離 } d_{ij} = \|\text{個体}_i - \text{個体}_j\|$$

である。これを次のように変換し、 n 次正方行列 (Z_{ij}) を作る。

$$Z_{ij} = \frac{1}{2} \left(\sum_{i=1}^m \frac{d_{ij}^2}{m} + \sum_{j=1}^m \frac{d_{ij}^2}{m} - \sum_{i=1}^m \sum_{j=1}^m \frac{d_{ij}^2}{m^2} - d_{ij}^2 \right)$$

さらに Z を対角化して対角行列 $\sqrt{\Lambda}$ を作る

$$\sqrt{\Lambda} = \begin{bmatrix} \sqrt{\lambda_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{\lambda_n} \end{bmatrix}$$

また、 Z を対角化する直交行列 X と掛け合わせることで n 次元正方行列 $P = X\sqrt{\Lambda}$ を作る。

このとき行列 P の第 i 行目 $(P_{i1}, P_{i2}, \dots, P_{in})$ は個体 i の n 次元ユークリッド空間での座標となる。二次元の場合は (P_{i1}, P_{i2}) を座標とし、ノード間の距離を図示した知覚マップを作成できる。類似している個体同士は近くに配置される。

3.2. 推薦

一般的に推薦には、ユーザベース推薦とアイテムベース推薦がある [4]。

ユーザベースの推薦ではユーザ同士の類似度を算出するため、推薦対象のユーザを追加する度に、すでに登録されている全てのユーザとの類似度を算出する必要があり、計算量が多くなる問題が生じる。

アイテムベースの推薦ではアイテム同士の類似度を算出し、それを元に推薦を行う。具体的には、ユーザが追加された際に、そのユーザの持っているアイテム間に重みを持たせてアイテム同士の類似度行列を作成し、次に推薦対象となるユーザが持っているアイテムと事前に作成しておいた類似度行列を比較する。これにより、計算量が少なくなり、ユーザベースより遥かに早い時間で推薦を行うことが可能である。

4. 提案手法

4.1. 視聴データの抽出

本研究では配信者に紐づくデータではなく視聴者の視聴履歴を元に配信者の関係性を間接的に取得し、可視化を行う。

ニコニコ生放送では配信中に視聴者がリアルタイムでコメントを付ける機能がある。またユーザはニコニコ動画に登録した際に一意的に割り当てられるユーザ ID を所持しており、配信にコメントを付ける際に自分の ID を公開することができる。ID を公開することにより配信者に視聴者のユーザ名と ID を認識してもらえようになる。またニコニコ生放送では匿名コメントの機能も存在しており、誹謗中傷などを行う際は匿名でコメントをすることが多く見られる。

生米保管庫 [5]では匿名でないユーザ ID を入力することにより、視聴データをカレンダー形式で表示することができる。視聴データは視聴者がコメントを付けた配信者の ID や、コメント内容、またユーザ ID が配信者でもあった場合は配信履歴まで取得することができる。

本研究ではこの視聴データを分析し、各視聴者が視聴した配信者の ID と視聴回数だけを抽出して視聴履歴として新しく定義した。

4.2. 非類似度行列の作成

4.1 節で作成した視聴履歴を基に配信者間に重みを付け、配信者間の非類似度行列を作成することにした。例として以下の視聴履歴を可視化する場合を考える。

- 視聴履歴 1 [A: 5 回, B: 2 回, C: 1 回, D: 2 回]
- 視聴履歴 2 [B: 3 回, C: 2 回, E: 6 回]
- 視聴履歴 3 [A: 3 回, B: 1 回, C: 1 回, E: 7 回]
- 視聴履歴 4 [C: 3 回, E: 1 回]
- 視聴履歴 5 [F: 7 回]

このとき、視聴回数のカウントは表 1 のように、各視聴者がデータ取得期間内に視聴している配信者間に重みを持たせている。また、A と F のように関係性が全く得られなかったセルは Null とした。

表 1 配信者間の重み

	A	B	C	D	E	F
A	Null	2	2	1	1	Null
B	2	Null	3	1	2	Null
C	2	3	Null	1	3	Null
D	1	1	1	Null	Null	Null
E	1	2	3	Null	Null	Null
F	Null	Null	Null	Null	Null	Null

次に重みの最大値から各項の数値を引くことで非類似度行列を求めると、表 2 のようになる。Null のセルは非類似度を 4 とした。この行列は数値が低いものほど関係性が大きいことを意味し、C と B は関係性が強く、A と F は関係性が最も弱いことを意味する。

表 2 非類似度行列

	A	B	C	D	E	F
A	0	2	2	3	3	4
B	2	0	1	3	2	4
C	2	1	0	3	1	4
D	3	3	3	0	4	4
E	3	2	1	4	0	4
F	4	4	4	4	4	0

4.3. 可視化

4.2 節で作成した非類似度行列に多次元尺度法を適用し、座標を取得すると表 3 のようになる(ここでは小数点第 3 位で四捨五入している)。この座標を基に R 言語 [6]で可視化を行うと、図 1 のように各ノード間の関係性を見ることができる。非類似度が低いほど近くに表示され、配信者と関係を持たないノードは他ノードから離れた位置に配置されている。

表 3 多次元尺度法による座標

	X 座標	Y 座標
A	-0.51	-0.63
B	-0.92	0.02
C	-1.06	0.30
D	0.65	-2.17
E	-1.02	1.46
F	2.87	1.02

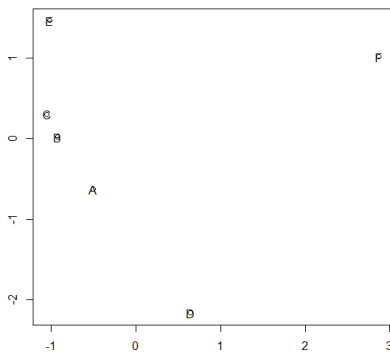


図1 R言語によるテストデータの可視化

4.4. 推薦手法

本研究では推薦結果を単に ID として表示するのではなく、推薦結果を可視化結果上に重ねて表示することで、ユーザの好みの配信を視覚的に捉えられるようにする。

アイテムベース推薦では、配信者の非類似度行列に多次元尺度法を適用することにより得られた座標を用いて、入力した ID 郡の各座標から距離の近い 2 点を候補 1 として定義した。

ユーザベース推薦では、入力した配信者の ID 郡を最も多く含む視聴者を視聴履歴から見つけ、類似ユーザとして定義し、類似ユーザの視聴している配信者全てを候補 2 として定義した。

さらに、候補 1 と候補 2 を照合し、両方の候補に含まれている配信者を推薦優先度「高」とし、片方のみに含まれる配信者を推薦優先度「低」として定義した。

5. 実装

5.1. 元データの取得

本研究ではニコニコ生放送からのユーザ ID の取得にニコ生コメントビューア [7]を使用した。ニコ生コメントビューアではニコニコ生放送での配信の URL を入力することにより、配信に付けられたコメントとユーザ ID を結び付けて表示することができる。本研究では様々な視聴傾向のデータが必要なため、100 人の配信者の配信から匿名でないユーザ ID を 3 人ずつ取得し、合計で 300 人の視聴者から ID を取得した。

次に視聴データを取得するため、生米保管庫を使用した。生米保管庫ではユーザを入力することで視聴者の視聴データをカレンダー形式で表示することができる。wget を使用して、300 人の視聴者が 2014 年 8 月 1 日～31 日までの 1 ヶ月間に視聴した視聴データを生米保管庫から取得した。

5.2. データの抽出

データの抽出のため、以下の処理を Processing で実装した。取得した視聴データには余分な情報が多く含まれているため、各視聴者の視聴した配信者の ID と視聴回数をカウントし、視聴履歴としてテキストファイルに書き出した。

視聴履歴から配信者の非類似度行列を作るためには、視聴履歴を 1 人ずつ読み込み配信者間に重みを付ける必要がある。このため、可変長の行列を用意する必要があり、本研究ではその実装に 2 重の HashMap を用いた。

次に視聴履歴を再度読み込み、各ユーザから配信者間の重みカウントしていき、行列を作成した。次に重みの最大値を求め、各セルとの差を求めることで非類似度行列に変換した。しかし、この状態だと配信者間に関係性を持たないセルが Null のままであるため、Null のセルは非類似度を重みの最大値よりも大きい値に設定した。次に、R で非類似度行列を読み込み、多次元尺度法の関数である cmdscale()を用いて各ノードの二次元平面上での座標を取得した。R には plot 機能も用意されており、R 上で可視化を行うこともできるが、本研究では可視化結果を推薦結果と組み合わせて使用するため、配信者の座標データのみをテキストファイルとして書き出した。

5.3. 可視化

可視化には Processing を使用した。5.2 節で求めた配信者の座標を読み込み、配信者のノードに座標と配信者 ID を持たせて可視化を行った。

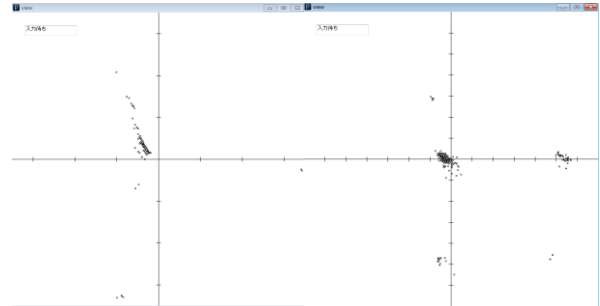


図2 Processingによる可視化

図 2 の左側は視聴者データが 100 人の場合の可視化結果である。メモリは距離 40 毎に表示している。視聴者 100 人に対する配信者ノードの数は 775 個であった。要素数が少ないため、重みの最大値は 3、非類似度の最大値は 100 であり、他の配信者との関係性が全く無いノード郡が中心付近に少し散らばっていることがわかる。

次に図 2 の右側の画像は視聴データが 300 人の場合の可視化結果である。視聴者 300 人に対する配信者ノードの数は 2008 個であった。重みの最大値は 7、非類似度の最大値は 100 であり、左側の画像に比べると関係性の深いノード郡が集中し始めていることがわかる。

5.4. 推薦

推薦機能の実装は Processing で行った。配信者の ID を推薦するものと推薦しないものの 2 種類に分類してしまうと可視化結果上に表示する際に単調になってしまうため、多次元尺度法を用いたアイテムベースによる推薦と、視聴データを基に類似ユーザを検索し、その視聴履歴を用いたユーザベース推薦の 2 種類を行った。

使用者が普段視聴している配信者の ID を複数入力すると、入力した各ノードと多次元尺度法の結果における

距離が近いノード 2 組を候補 1 として追加した。ただし、入力した ID 郡の中に距離が最短の組がある場合、推薦候補が入力ノードと同じになってしまうため、最短ではなく距離が近い 2 組を候補として追加した。次に、入力した ID 郡が最も多く含まれている他の視聴者を検索し、その視聴者の視聴履歴全てを候補 2 として追加した。

次に両方の推薦手法に適合するものを優先度「高」として赤色で大きく表示し、片方だけに適合するものは優先度「低」として橙色で表示し、他のノードは紺色で表示した。

6. 実験

作成したプログラムを用いて可視化と推薦の実験を行った。以下では視聴データ 100 人での実験の結果を与える。入力する視聴者 ID を 27042280 と 41237297 にした場合、優先度「高」として出力されたのが 21735538, 27042280, 41183171, 41237297 の 4 人、優先度「低」として表示されたのが 38 人となった。

実際に上記 4 組のユーザーページやニコニコ大百科を検索してみると以下のことがわかった。

表 4, 表 5 からわかる通り、入力ノードと関係の深いノード郡が上手く推薦されている。

41183171 は入力した配信者のミラー配信者であり、推薦優先度「高」として表示された。

21735538 は入力した配信者と似た配信内容であり推薦された。この結果は本研究の目的に即していると考えられる。また、27042280 と 41237297 は入力した ID が推薦として出力されているが、これは多次元尺度法上での距離が近かったためであり、入力自体に視聴傾向の偏りが生じていたためである。

表 4 入力した配信者

入力	配信ジャンル
27042280	萌え声・ゲーム
41237297	ゲーム・料理

表 5 推薦された配信者

出力	配信ジャンル	関係性
27042280	萌え声・ゲーム	41237297 と師弟関係
41237297	ゲーム・料理	27042280 と師弟関係
21735538	萌え声・ゲーム	27042280 と似た配信内容
41183171	配信ミラー	27042280 のミラー配信者

図 3 を見ると 2 点が近い位置に重なって表示されることがわかる。X 座標が -40~-80, Y 座標が 80~120 付近に推薦ノードが集中しており、視聴傾向と推薦傾向に偏りが出ていることがわかる。

7. 考察

5.3 節の表示結果からわかるようにサンプルとして取得した視聴者データが少なかったと考えられる。またこれにより類似ユーザを検索する際に、単に視聴している配信者の種類が多いユーザが候補として保持されてしまい、

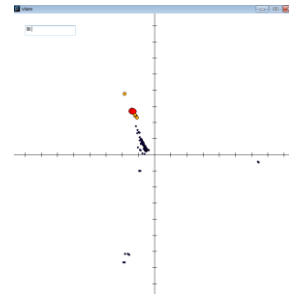


図 3 Processing による可視化推薦結果

推薦結果の優先度「低」のノード数が多くなってしまった。この問題を解決するためには、類似度を求める際に各視聴データの視聴回数を考慮して重み行列を作成する必要がある。

また、推薦結果が可視化と ID での表示のため、その配信者がどのような人物なのか ID をもとに別途検索する必要がある。ノードにコミュニティページへのリンクを表示できればより良い推薦システムになることが見込まれる。

8. まとめ

本研究では視聴者の視聴履歴をもとに配信者の関係を可視化し、配信者の推薦システムの実装を行った。実験では入力したデータに対して正しい推薦結果を得ることができた。また、可視化結果に推薦結果を組み込むことで、結果的に配信者の推薦に留まらず、使用したユーザの視聴傾向や好みまで把握することができた。

本論文では少ない視聴者から可視化や推薦を行ったが、ニコニコ生放送の全ての ID を基にこのシステムを用いることで、ニコニコ生放送の現状を知り、公式の検索エンジンより優れたシステムを実現することができると考えられる。

文 献

- [1] "ニコニコ生放送," [Online]. Available: <http://live.nicovideo.jp/>.
- [2] "ちくわちゃんランキング," [Online]. Available: <http://www.chikuwachan.com/live/>.
- [3] 濱崎雅弘, 後藤真孝, "Songrium: 多様な関係性に基づく音楽視聴支援サービス," 情報処理学会研究報告, vol. 2012-MUS-96, no. 1, pp. 1-8, 2012.
- [4] T. Segaran, 集合知プログラミング, 2008, pp. 7-30.
- [5] "生米保管庫," [Online]. Available: <http://namagome.com/>.
- [6] 中村永友, R で学ぶデータサイエンス 2 : 多次元データ解析法, 2009, pp. 177-188.
- [7] "ニコ生コメントビューア," [Online]. Available: <http://www.posite-c.com/application/ncv/>.