

ニュース記事における共起情報を利用した アーティスト推薦アルゴリズムの検討

松尾 繭 正田 備也

長崎大学工学部情報システム工学科

b609416@yahoo.co.jp, masada@nagasaki-u.ac.jp

概要 近年、携帯音楽プレーヤーやインターネット上の音楽サービスの普及により、ユーザが音楽を聴く機会は増加している。また、時代に沿って音楽のスタイルが増え続け、多種多様なアーティストが無数に存在している。この状況の中で、自分で自分の嗜好に合うアーティスト・楽曲を見つけ出すのは困難であり、自動的に推薦してくれるシステムの需要が高まっている。本研究では、ユーザの嗜好に合うアーティストを、ランキング形式で表示できるアルゴリズムについて検討する。その際、ニュース記事でアーティストが共起する頻度を用いる。この手法を、有名なアーティストと有名でないアーティストの2つの場合で検証する。

キーワード 情報推薦, 音楽情報, 共起

1 はじめに

近年、iPod 等の携帯音楽プレーヤーの普及により、ユーザが生活の中で聴く楽曲は増加傾向にある。また、youtube やニコニコ動画等の、インターネット上で自分の聴きたい楽曲を無料で聴くことの出来るサイトの普及も重なった。この事を背景に、現在、無数にいるアーティスト・楽曲から、自分の嗜好に合うものを探し出してくれる推薦システムの需要が高まっている。[3][4]

2 従来手法

情報推薦の従来手法は、協調フィルタリングと内容ベースフィルタリングが挙げられる。前者は、ユーザ間の類似度に基づく推薦である。また、後者は、アイテム間の類似度に基づく推薦であり、システムを使い始めたばかりのユーザに対しても情報提供が可能であるというメリットがある。本研究は後者の手法に分類される。[1]

3 提案手法

本研究では、株式会社ナターシャが運営する、最新音楽ニュース「ナタリー」のニュース記事を使う。各ニュース記事には、記事に出現するアーティストのリンクが関連アーティストとして記事上部に付けられており(図1参照)、これを共起情報として利用する。8月7日の時点でナタリーに登録されていたアーティストの記事を新しい順に最大30件取得した。扱った記事は10,364件、アーティストの出現回数は55,663回であり、1つの記事に平均約5.4アーティストが出現していることになる。



図1 「ナタリー」のニュース記事例

ユーザは推薦して欲しいアーティストの指針となるアーティスト名を入力する。そのアーティストと一緒に記事の出現回数が多い順に、アーティスト名を表示する。これにより、ユーザの嗜好にあったアーティストを推薦できると考えた。表1の6種類のアルゴリズムを提案する。

Table1 6種類のアルゴリズム

アルゴリズム名	特徴
N	出現回数が多い順にアーティスト名を表示
Nlog	N+アーティストの重要度をlogを使って反映
Nsqrt	N+アーティストの重要度を√を使って反映
I	N+記事の重要度を反映
Ilog	N+I+アーティストの重要度をlogを使って反映
Isqrt	N+I+アーティストの重要度を√を使って反映

基本となるのはNのアルゴリズムである。これにアーティストの重要度や記事の重要度を反映させたものが、他の5つのアルゴリズムである。

アーティストの重要度 有名なアーティストの方が、話題にされやすいので出現回数が多い。また、同じ話題に対して複数の記事が掲載されることがある。この場合の記事1つと、有名でないアーティストの記事1つを同じ重みと考えるのはおかしい。情報検索における IDF[2]と似た考え方を使い、出現回数の多いアーティストの重みは小さくなるようにする。今回は、 \log や $\sqrt{\quad}$ を用いて1つの記事の重みを $1/(\log(df+1))$, または $1/\sqrt{df}$ にする。

ここで df はアーティストの記事出現回数を表している。

記事の重要度 一緒の記事に出現するアーティストの数が少ない方が、アーティスト同士の関連性が強いと考えられる。アーティストの重要度と同様に、情報検索における IDF と似た考え方を使い、出現するアーティストの数が多記事の重みは小さくなるようにする。今回は、アーティストの数の逆数を用いて反映させた。

4 実験

有名でないアーティストの例として、嘘つきバービー（記事出現回数14回）、有名なアーティストの例として、チャットモンチー（記事出現回数175回）を使った。実験結果は、推薦結果の1位～3位のアーティストが自分の嗜好に合うかどうかを、○×で判定する。それを全てのアルゴリズムで行い、表にまとめる。

指針となるアーティストに嘘つきバービーを入力した場合の結果を表2に、チャットモンチーを入力した場合の結果を表3に示す。縦の数字は順位を、横の文字はアルゴリズム名を表している。図2のNのアルゴリズムの2位の「○×○○」のように記号が並んでいるものは、同率の順位のものを表している。

	N	Nlog	Nsqrt	I	Ilog	Isqrt
1	○	×	×	○	○	×
2	○×○○	○	○	○	×	○
3		○	×	×	×	×

図2 嘘つきバービーの推薦結果の評価

	N	Nlog	Nsqrt	I	Ilog	Isqrt
1	○	○	○	○	○	○
2	×	×	○	×	○	×
3	○	○	○	○	×	○

図3 チャットモンチーの推薦結果の評価

嘘つきバービーの結果では、アーティスト、記事の重要度を反映させなくても良い結果を得ることが出来たため、アルゴリズムNが一番有効だと思われる。チャットモンチーの結果では、結果にあまり違いが見られなかった。

5 考察

ナタリーの記事では、有名なアーティストは有名なアーティスト同士、有名でないアーティストは有名でないアーティスト同士で共起することが多く、アーティストの重要度を反映させても、効果は無かった。また、記事の重要度を反映させた場合も、有名でないアーティストの場合は、元々出現する記事が少なく、関連性の高いアーティストとしか記事に出現しないため、効果は無かった。

しかし、有名なアーティストの場合は、元々一緒の記事に出現するアーティストの数が多いため、記事の重要度を反映させるのとさせないのでは、結果に大きく違いが出た。チャットモンチーを入力した場合のNとIのアル

ゴリズムでの推薦結果を、表2に示す。人がどのようなアーティストを推薦して欲しいかと考えると、自分の好きなアーティストにテイストが似ているアーティスト・似ていないが意外と自分の嗜好に合っているアーティストの2種類があると考えられる。表2の結果の場合、推薦されるアーティストは違いがどちらも有効な結果であったと思われる、最初の種類が記事の重要度を反映させた場合に、後の種類が反映させなかった場合に相当すると考えられる。記事の重要度を反映させた場合は、出現するアーティストの数が少ない記事の重みが大きくなるので、そのような記事と一緒に出現するアーティストが上位に位置しやすくなる。そして、それらのアーティストは音楽性が類似していると考えられる。これとは対照的に、記事の重要度を反映させない場合は、出現するアーティストの数が多記事も1つとして数えられるので、頻繁に同じ記事には登場するが音楽性が類似していないアーティストに行き着く場合がある。この場合は、「意外な情報」[5]をユーザに与えていると思われる。

Table2 チャットモンチーの場合のNとIでの推薦結果

順位\記事重要度	N(反映させない)	I(反映させる)
1	9mm Parabellum Bullet	ASIAN KUNG-FU GENERATION
2	THE BACK HORN	ストレイテナー
3	THE BAWDIES	奥田民生

6 今後の展望

現段階では、基本の結果の算出方法+アーティストの重要度・記事の重要度を反映させた場合を研究してきた。今後、何か他に反映させて良い結果を生む要素があるかを引き続き考えていく。また、有名なアーティストと有名でないアーティストの違いをさらに考え、それぞれに適しているデータのある他のサイトを探し(例えば、Wikipedia は有名なアーティストの場合に有効なデータ源ではないかと考えている)、ナタリーから集めたデータと組み合わせることを検討する。さらに、被験者を使った評価実験も予定している。

参考文献

- [1] 田中克己, 角屋和俊: 情報推薦システム入門 理論と実践, 共立出版, 2012.
- [2] C.D. Manning, P. Raghavan, H. Schütze (岩野, 黒川, 濱田, 村上 訳): 情報検索の基礎, 共立出版, 2012.
- [3] 吉井和佳, 後藤真孝: 音楽推薦システム, 情報処理学会誌, Vol.50, No.8, pp.751-755, 2009.
- [4] 西尾翼, 河野浩之: 感性情報とランキング情報による楽曲推薦システム, 南山大学修士論文, 2011.
- [5] 後藤清豪, 高田秀志: ソーシャルメディア上での行動に基づく「意外な情報」の提供者になり得る人物の推薦手法, Vol2010-DBS-151, No.41, 2010.