

## 興味に基づく音楽情報推薦システムの提案

野澤 大介<sup>†</sup> 河合 由起子<sup>†</sup> 熊本 忠彦<sup>††</sup><sup>†</sup> 京都産業大学理学部コンピュータ科学科    <sup>††</sup> 千葉工業大学情報科学部情報ネットワーク学科

## 1 はじめに

膨大な音楽情報の中から曲名や歌詞、メロディーといった音楽に関する情報を用いて検索する研究が広く行われている [1][2][3]。しかしながら、音楽情報と直接的には関係のないキーワードを用いるような、ユーザの幅広い興味に合わせた検索は困難である。また、これまでの協調フィルタリングに基づく情報推薦では、多数のユーザ情報が必要であり、小規模な利用では高い精度が得られ難い。そこで、本研究では、それぞれのアーティストに関して書かれたブログ記事とその書き手（ブロガー）に関する個人情報（年齢や性別、出身地）からある年代・性別・出身地の人らにどのようなアーティストが話題となっているかを特定し、ユーザの年齢・性別・出身地に合った音楽情報（アーティストやブログ記事）を推薦する手法を提案する。具体的には、yahoo! ブログや goo ブログから各アーティストごとに約 1,000 件程度のブログ記事を収集して形態素解析し、 $tf \cdot idf$  値から特徴的な語彙を選別する。また、ブログを書いているブロガーのプロファイル約 12,000 人分を抽出した個人情報（年齢・性別・出身地）を各ブログ記事とアーティストプロファイルに対応付ける。これらの語彙と個人情報を用いることで、ユーザは、どういった年齢や性別、出身地のブロガーが、どのような話題をもとに、どのアーティストに興味を持っているか（話題度）が分かる。さらに、ユーザが閲覧した記事の話題と類似する話題を持つ記事や、関連のあるアーティストを推薦できる。

## 2 システム設計

図 1 に提案手法を実装したシステムの概要を示す。システムの流れは、ユーザの問合せ前にブログ情報を収集し分析する機構と、ユーザの問合せが発生した際に検索し選択する推薦機構とに分けられる。

ユーザの問合せ前では、まず、ブログサイトごとにアーティスト名をキーワードとして検索した結果の上位 1,000 件のブログ記事を収集する。収集したブログ

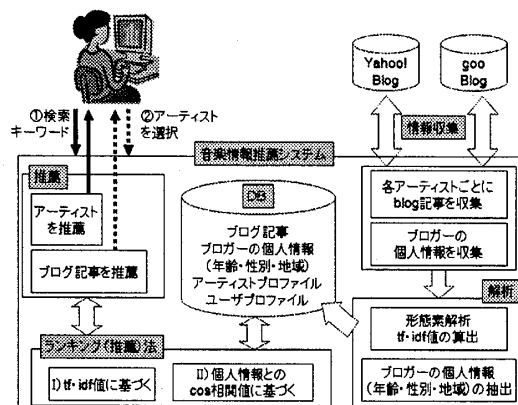


図 1: 音楽情報推薦システムの概要

記事を茶釜 [4] を用いて形態素解析し、単語（名詞）の  $tf \cdot idf$  値を算出する。さらに、取得したブログ記事の URL よりブロガーのプロファイルページを取得する。取得したプロファイルページから、年齢・性別・出身地となる箇所を特定し、各ブロガーの記事とアーティストに対応付ける。

次に、ユーザの問合せでは、ユーザがキーワード検索するとユーザ ID、検索キーワード、アクセス時間を記録したユーザプロファイルが作成される。検索では、キーワードを含む記事集合をアーティストごとに分類し、分類した記事集合からそのキーワードの  $tf \cdot idf$  値の合計値を算出し、合計値の高い順にアーティスト名をユーザへ推薦する。ユーザがアーティストを選択すると、そのアーティストの記事が  $tf \cdot idf$  値の高い順に推薦され、同時にユーザプロファイルにアーティスト名が追加される。さらにユーザが提示された記事を選択すると、記事の URL がプロファイルに追記される。また、アーティストと記事には事前にブロガーの年齢・性別・出身地が付与されているので、ユーザの年齢・性別・出身地に合わせて記事やアーティストを検索・推薦することもできる。

2.1  $tf \cdot idf$  に基づく推薦

提案手法には、I) 語彙に基づく推薦と II) ブロガーの情報に基づく推薦の 2 種類がある。語彙に基づく推薦では、キーワードから記事を特定するのではなく、アーティストを特定する。具体的には、 $tf \cdot idf$  値を算出する際に、 $df$  値はユーザが入力したキーワード

<sup>†</sup> Daisuke Nozawa(g438023@cc.kyoto-su.ac.jp)<sup>†</sup> Yukiko Kawai(kawai@cc.kyoto-su.ac.jp)<sup>††</sup> Tadahiko Kumamoto(kumamoto@net.it-chiba.ac.jp)  
Department of Computer Science, Kyoto Sangyo University  
(†)

Department of Information and Network Science, Chiba Institute of Technology (††)

を含む記事全てを対象とせず、アーティストごとに分類した  $df_{(artist)_i} = (\text{キーワードとアーティスト } i \text{ とを含む記事}) / (\text{アーティスト } i \text{ の記事総数})$  として算出する。この算出した  $tf \cdot idf$  値をアーティストごとに集計し、合計値を基にアーティストを推薦する。また、キーワードとアーティストを含む記事は、アーティストごとに  $tf \cdot idf$  値の高い順に推薦される。

## 2.2 年齢・性別・出身地に基づく推薦

ブロガーの年齢・性別・出身地（個人情報）に基づくアーティストの推薦は、アーティスト名を含む記事からブロガーの個人情報を取得し、それらを基にベクトル  $V = (v_{age}, v_{sex}, v_{local})$  を決定する。ここで、任意の記事  $p$  のベクトルを  $V_p$ 、任意のアーティスト  $a$  のベクトルを  $V_a$ 、任意のユーザ  $u$  のベクトルを  $V_u$  とする。ベクトルの要素  $v_{age}$  は、0~100歳の年齢値で（年齢/100）とし、 $v_{sex}$  は、性別で男性が1、女性を0の値とする。また、 $v_{local}$  は、出身地とし北海道から沖縄の都道府県を0~1の範囲で設定する。

記事ベクトル  $V_p$  は、収集された時点では、その記事を書いたブロガーの個人情報を要素値とする。任意のユーザ  $u$  が閲覧した場合、 $(V_p + V_u) / 2$  が新たな  $V_p$  となる。また次に他のユーザ  $u_2$  が閲覧した場合は、同様に更新されて  $V_p = (V_p + V_{u_2}) / 2$  となる。

任意のアーティスト  $a$  のベクトル  $V_a$  は、収集された時点では、アーティスト  $a$  について書いてある各記事のブロガーの個人情報の平均値とする。ユーザ  $u$  がアーティストを選択した場合、 $(V_a + V_u) / 2$  が新たな  $V_a$  となる。また、次に他のユーザ  $u_2$  が閲覧した場合も同様に更新され、 $V_a = (V_a + V_{u_2}) / 2$  となる。

任意のユーザ  $u$  のベクトル  $V_u$  は、初めて検索した時点では、そのユーザの個人情報が要素値となる。ユーザが記事を閲覧したりアーティストを選択した場合は、 $(V_u + V_p \text{ または } V_a) / 2$  が新たな  $V_u$  となる。

以上の  $V_p$ 、 $V_a$ 、 $V_u$  を用いて、アーティストの推薦は  $V_u$  と  $V_a$  との  $\cos$  相関を算出し、類似度の高いアーティストを推薦する。記事推薦では、 $V_u$  と  $V_p$  との  $\cos$  相関を算出し、類似度の高い記事を推薦する。

## 3 実験

アーティスト名とブログ記事を推薦する音楽情報推薦システムを構築し、システムの評価を行った。実験ではYahoo!ブログからアーティスト27人分のブログ記事24,925件（ブロガー11,738人分）を収集した。また、取得した記事から、Yahoo!のブロガープロフィールの形式に基づき、年齢を4,799人分、性別を9,047人分抽出した。尚、gooより6,326人分の性別と2,550人分の出身地を抽出できたが、本稿の実験では総数の

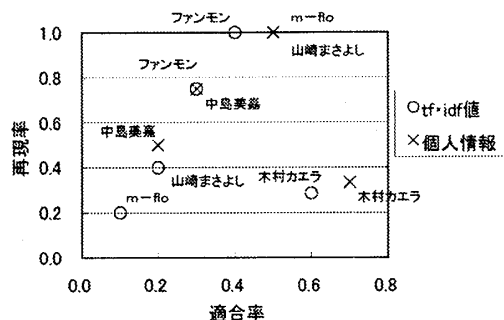


図2:  $tf \cdot idf$  値と個人情報とを用いた記事推薦の適合率と再現率（検索キーワード「バイク」）

多いYahoo!の年齢と性別のみを用いた。また、適合率と再現率は被験者1名（22歳男子学生）の主観的判断（適合/不適合）に基づいて算出した。

図2に、「バイク」で検索した結果の上位5名のアーティストの、 $tf \cdot idf$  値に基づいて推薦した上位10件の記事に対する適合率・再現率と、ブロガーの個人情報に基づいた推薦記事10件の適合率・再現率を示す。尚、適合率は、(10件中適合と判断された記事数) / 10とし、再現率は、(10件中適合と判断された記事数) / (キーワードとアーティスト名を含む記事のうち適合と判断された記事数)とする。また、各  $V_a$  と  $V_p$  は1回、 $V_u$  は3回更新された。実験結果より、上位5名のアーティストのうち3名に対しては、ブロガーの個人情報を用いた記事推薦の優位性が示された。

## 4 まとめと今後の課題

本論文では、各アーティストに関して書かれたブログ記事とその書き手に関する個人情報（年齢と性別）から、ある年代・性別の人らにどのようなアーティストが話題となっているかを特定し、ユーザの年齢・性別に合った音楽情報（アーティストや記事）を推薦する手法を提案し、評価実験を行った。今後、出身地による地域情報も利用した推薦の評価実験と、異表記問題の解決、さらにアーティストがグループの場合の対処が課題となる。

## 参考文献

- [1] 石先広海, 帆足啓一郎, 菅谷史昭, 複数検索クエリによる音楽情報検索結果統合手法, 第69回情処全大講演論文集, 4C-4, pp.2-117-2-118, 2007.
- [2] 橋口博樹, 西村拓一, 矢部博明, 赤坂貴志, 岡隆一, 鼻歌による音楽検索と歌詞音声検索の統合処理の検討, 情処研報 (音楽情報科学), Vol.2001, No.16, pp.57-62, 2001.
- [3] 園田智也, 後藤真孝, 村岡洋一, WWW上での歌声による曲検索システム, 信学論 (D-II), Vol.J82-D-II, No.4, pp.721-731, 1999.
- [4] 形態素解析システム茶釜  
http://chasen.naist.jp/hiki/ChaSen/