Technology Reports（特集）
レコメンド　深層学習　楽曲
生活を充実させるAI技術特集

歌唱履歴の時系列性を考慮した楽曲レコメンド技術の開発

サービスイノベーション部　田中（たなか） 茂樹（しげき）　深澤（ふかざわ） 佑介（ゆうすけ）†

カラオケにおいては検索機能が充実しており（歌手別，年代別など），ユーザはこれらを用いて楽曲を選択できる一方，システムが楽曲を能動的にレコメンドする機能を有することは稀である．これはカラオケにおけるレコメンドの困難さによるところで，特に対象がグループであること，過去に来店された時の歌唱履歴がないこと，場の雰囲気で選択される楽曲が変化することに起因する．そこで，ドコモでは深層学習を用いたレコメンドモデルを開発した．本モデルでは時系列を考慮した深層学習を用い，さまざまな楽曲情報と併せて学習させることで，未知のグループに対してシステムが適切なレコメンドを行うことが可能となった．

1. まえがき

機械学習*1の進歩により，レコメンド技術はオンラインショッピングサイトを始めとしたさまざまなサービスに導入され，ユーザ体験の向上やビジネスの活性化に貢献している．一方，カラオケにおいては下記のような性質から，適切なレコメンドの実現が非常に困難であった．

・ユーザは歌唱履歴のないグループである
・グループメンバはグループ構成や場の雰囲気によって歌唱曲を選択する

レコメンドの代表的な手法であり，さまざまな分野で活用されている協調フィルタレコメンド手法は，一般に個人の商品購入履歴やWeb閲覧履歴などから趣味・思考を理解し，似たユーザが興味をもつ商品を関連商品として推薦する．また，個々のプロファイルを組み合わせてグループのプロファイルへ拡張する手法も提案されている［1］．一方，カラオケではカラオケ端末へログインせず，過去に訪れた際の歌唱履歴をもたないユーザも多いため，プロファイルが困難である上にグループの人数すら不明

©2020 NTT DOCOMO, INC.
本誌掲載記事の無断転載を禁じます．
本誌に掲載されている社名，製品およびソフトウエア，サービスなどの名称は，各社の商標または登録商標．

† 現在，総務部

＊1 機械学習：事例をもとにした統計処理により，計算機に入力と出力の関係を学習させる枠組み．

である．

そのため，例えば第一興商では一番最後に歌われた楽曲に類似した楽曲をレコメンドするレコメンド機能を提供していた．しかし，カラオケをグループで利用する場合，メンバが順番に歌唱することが多く，直近に歌唱したメンバに対して類似楽曲をレコメンドできても，次に歌唱するのが別のメンバである場合にレコメンドの効果が得られない問題があった．

また，メンバがグループの構成や雰囲気によって歌唱曲を選択するという点も，レコメンドを困難にする要因である．例えば，グループが会社の同僚で構成される場合と友人・家族の場合では，選択されやすい楽曲が異なる可能性が高い．また，カラオケでは入室から退室までの間に場の雰囲気は変化し，盛り上がっている時の楽曲や締めの1曲をもっているメンバも多い．このように，その場の空気が選曲に影響する場合は多く，レコメンドにおいてもその変化を読み取ることが肝要であるが，前述の協調フィルタに代表されるレコメンド手法は一般に長期的な趣味・嗜好を捉えるものであり，場や雰囲気によって常に変化する領域では期待される効果が得られないことが多い［2］．

そこで，ドコモでは深層学習*2を用いたレコメンドモデルを開発した．

このモデルでは時系列を考慮し，楽曲が選択されるたびにその楽曲のさまざまなメタ情報を基に，次に選択される可能性の高い楽曲および歌手を予想する．これにより，過去の履歴がないグループにも効果的なレコメンドが可能となり，加えて歌唱の流れを捉えることでその場の雰囲気も考慮したレコメンドが可能となった．

本稿では，提案手法の詳細と，本手法を用いたリアルタイム楽曲レコメンドエンジンについて解説する．

2. 時系列考慮・多変量深層学習レコメンドモデル

ドコモが提案するレコメンドモデルは，直前に予約された楽曲のさまざまなメタ情報を入力することで次に選択する楽曲，もしくは歌手のレコメンドを行う．

2.1 モデル概要

提案モデルの概要を**図1**に示す．

本モデルは直前に歌唱された楽曲のIDに加え，歌手名，作曲者名，楽曲ジャンルといった情報を入力し，それらをモデル内で特徴量に変換する．例えば歌手名の特徴量は歌手同士の類似度などを表し，モデルは学習を通して各歌手をより良く表す特徴量を獲得する．これらの特徴量は再帰型ニューラルネットワーク（RNN：Recurrent Neural Network）へ入力される．RNNは隠れ状態と呼ばれる情報を保持する機能を有し，入力された特徴量と隠れ状態から出力演算を行い，同時に隠れ状態を更新する．この機構によって，過去に入力された情報もある程度加味することができる．モデルはRNNの出力を用いて次の楽曲と歌手のランキングを生成する．

本モデルの特長は，①RNNを用いていること，②多様な楽曲情報を入力していること，③複数のタスク（次の楽曲ランキング出力と歌手ランキング出力）を同時に行うことである．

①RNNを用いることで，今まで入力された楽曲情報をある程度記憶することが可能となる．これにより，入力された直近の楽曲情報だけでなく，グループが入室後に予約した1曲目からレコメンド直前に予約した曲までの履歴を時系列として捉え，順序などを考慮したレコメンドが可能となる．

②入力として楽曲のIDのみを扱う手法も一般的であるが，それに対し本モデルの学習では，多

*2 深層学習：多層構造のニューラルネットワークを用いた機械学習．

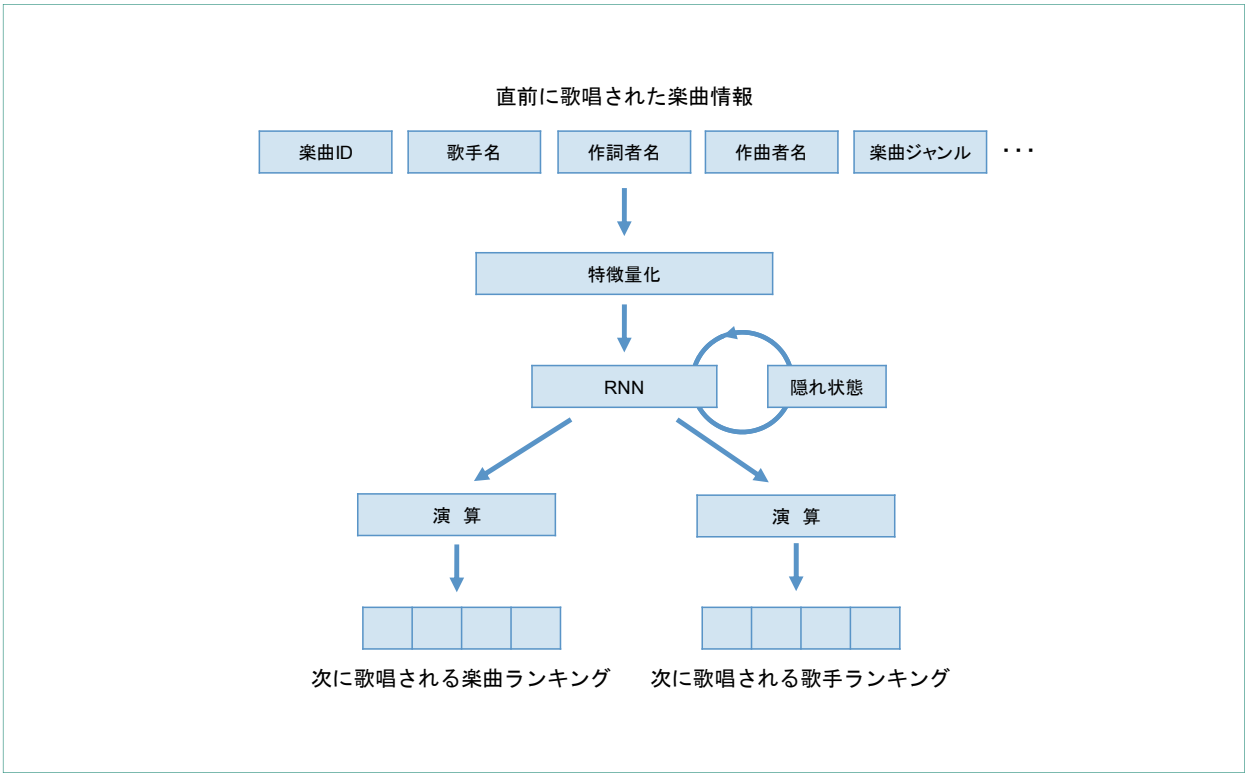

図1　提案レコメンドモデルの概要

様な楽曲情報を扱うことで，モデルが楽曲をより深く理解でき，レコメンドされた結果にメンバが歌おうとした楽曲が含まれる割合であるレコメンド精度の向上が期待できる．

③複数のタスクを同時に行うことでリソース活用の効率化が可能である．一般的な機械学習モデルは1つのタスクに特化して学習され，複数タスクがある場合は個々にモデルを生成するため，タスク数に比例してモデルの学習時間やモデル運用のためのリソース消費が増大する．複数のタスクを同時に解く設計はマルチタスク学習と呼ばれ，モデルは与えられた関連する複数のタスクを解くために，より広い視野での学習を行い，各タスクにおける精度の向上に寄与する．また，今回1つのモデルで2つの機能を実装できたことで，機能ごとにモデルを用意する場合に対しリソース消費を半分に抑えることができた．

2.2　精度検証

提案手法の精度検証のため，第一興商の協力のもと1カ月間のカラオケ店舗のデータを用いて既存手法との比較を行った．前半3週間のデータを用いて各モデルを学習し，後半1週間の推定精度を比較した．

既存手法は以下の2手法とした．

・ランキング：学習期間で人気の高かった楽曲を常にレコメンドする．

・Item-kNN（k-Nearest Neighbor Algorithm）：アイテムベースの協調フィルタレコメンド手法で，学習期間での楽曲同士の類似度テーブルを事前に作成する．このテーブルはそれぞれの楽曲同士がどの程度同一グループに歌唱されやす

いかを示しており，レコメンド時には直近に歌唱された楽曲と類似した楽曲をレコメンドする．

MAP@N*3での精度比較は**図2**の結果となり，楽曲と歌手のレコメンドいずれにおいても提案手法が既存手法を10%以上上回る精度を達成した．

また，**図3**では顧客の入室から退室までの履歴を3分割し，それぞれの精度の平均を評価することで時間経過による精度の推移を示している．ランキングによる精度が入室後，時間の経過とともに減少していることから，一般にグループ内ではメジャーな楽曲を始めに歌唱し，徐々に多様な楽曲を選択する

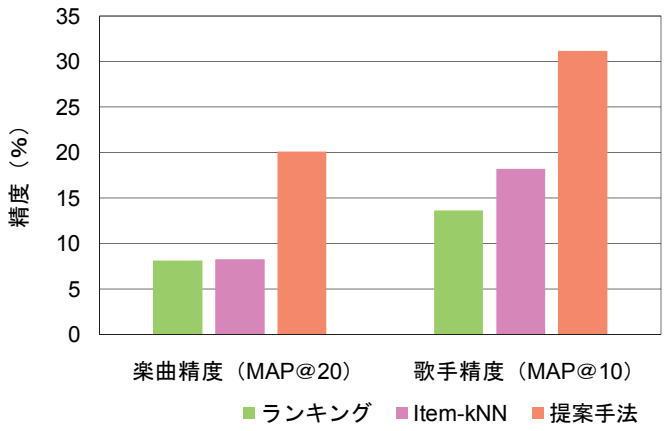

図2　各モデルの精度比較

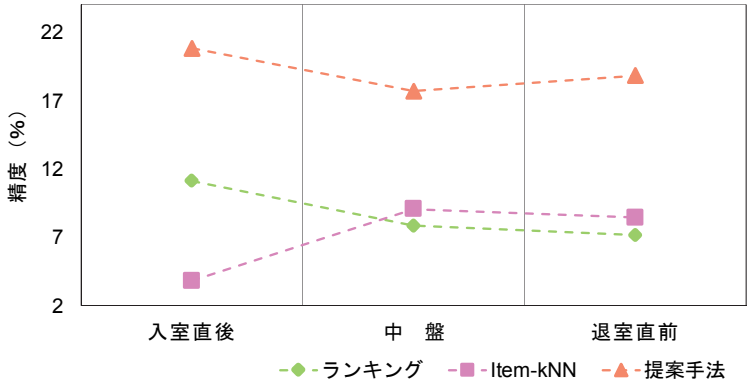

図3　各モデルの時間経過による楽曲精度推移（MAP@20）

*3　MAP@N：N個アイテムをレコメンドした際，次にユーザが選択したアイテムがその中に含まれる確率．

ことが分かる．また，Item-kNNによる精度が入室直後は非常に低く，中盤上昇してから再度低下していることから，メンバは始めに好きな楽曲を個々に選択し，中盤は他メンバの選曲に影響されるが，最後には自らの歌いたい楽曲で締める傾向があると推察される．これらの両モデルに対し，提案手法はすべてのタイミングで大きく上回る精度を達成しており，グループ内の雰囲気の変化も捉えられていることが分かる．

3. リアルタイム楽曲レコメンドエンジン

リアルタイム楽曲レコメンドエンジンは，提案モデルを用いたシステムとしてAWS（Amazon Web Services）[*4]上に構築され，第一興商へAPI（Application Programming Interface）[*5]経由でレコメンド機能提供を行っている．

システム概要図を**図4**へ示す．

本システムはバッチサーバとAPIサーバで構築される．バッチサーバはデータベースに蓄積された歌唱履歴や楽曲情報からレコメンドモデルの再学習を定期的に行い，APIサーバへ学習済みのモデルを送信し，APIサーバは学習済みモデルを最新に更新する．バッチサーバは，定期実行時以外は停止するが，APIサーバはレコメンドリクエスト処理のため，常に起動を続ける．レコメンドリクエストを受信したAPIサーバは，パラメータとして送信された端末情報や歌唱履歴を基に学習済みモデルで演算し，次にレコメンドする楽曲・歌手をレスポンスする．データベースの歌唱履歴は必ずしもリアルタイムで更新される訳ではないが，APIサーバは端末での歌唱履歴をレコメンドリクエストごとに受領するため，楽曲を予約するたびにレコメンドのリアルタイムな更新が可能となる．

レコメンド機能は**図5**のようにDAM（Daiichikosho Amusement Multimedia）端末上の予約前画面で提供され，対応するタブをタップすることで今までの履歴に基づきレコメンドが表示される．

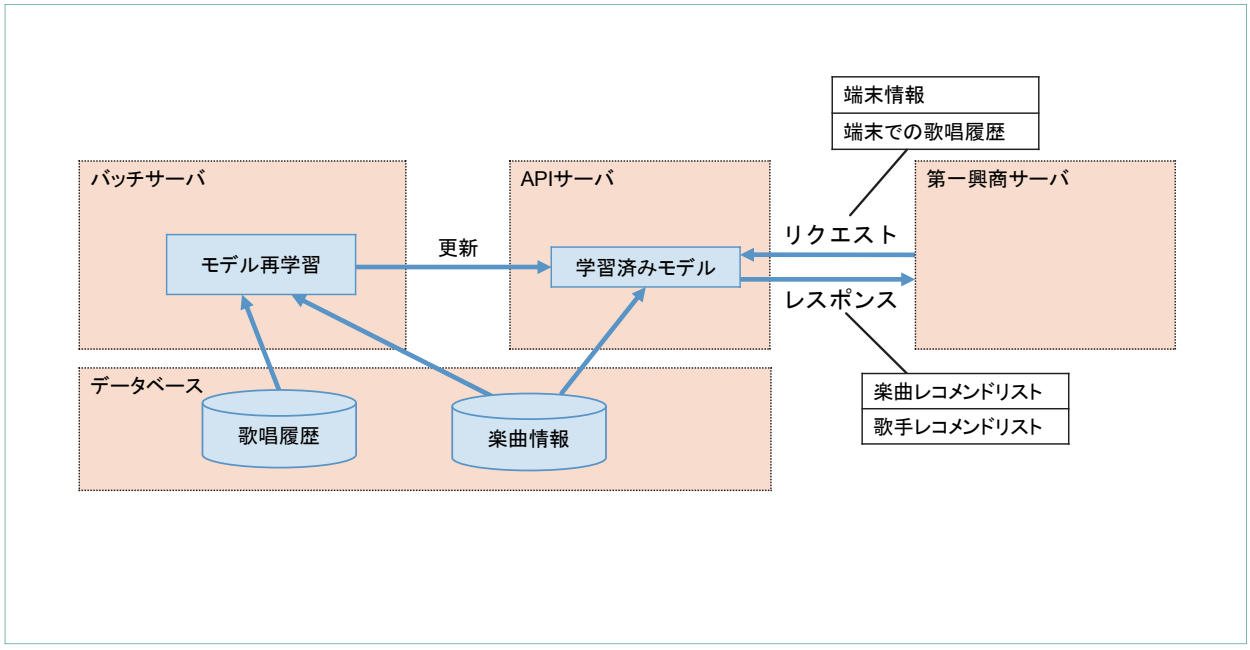

図4　レコメンドシステム概要

＊4　AWS：Amazon.comが提供するクラウドコンピューティングサービス．

＊5　API：ソフトウェアの機能を他のプログラムから利用できるように切り出したインターフェース．

図5　カラオケ端末上でのレコメンド表示画面

4. あとがき

本稿では，カラオケシステムにおいてレコメンドを行う際の課題と，それらを解決する深層学習を用いたレコメンド手法とその性能評価，および当手法を用いたレコメンドシステムについて解説した．

本稿執筆時点ではトライアルとしてレコメンド機能をカラオケ店舗に提供しており，フィードバックをいただきながら改善を行っている．今後正式な商用サービスとし，広く使用してもらうことでユーザ体験の向上に貢献したい．

文　献

[1] A. Felfernig, L. Boratto, M. Stettinger and M. Tkalčič: “Group recommender systems: An introduction,” Springer, 2018.

[2] M. Pazzani and D. Billsus: “Learning and revising user pro les: The identi cation of interesting web sites,” Machine learning, Vol.27, No.3, pp.313–331, 1997.