

状況依存型ユーザ嗜好モデリングに基づく **Context-Aware** 情報推薦システム

清水 豪士

富山県立大学 情報基盤工学講座
t715038@st.pu-toyama.ac.jp

July 17, 2021

背景

- Web の発達にともない、ユーザがアクセスできる情報が膨大になり、この膨大な情報の中から適切な情報を獲得することを目的とした情報推薦手法に関する研究が行われてきた。
- ユーザの各コンテクスト（ユーザの状況）に応じた細やかな情報推薦を行うためには、想定するコンテクストを細かく設定する必要があるが、各コンテクストに対応する十分な学習データを確保することは容易でない。

目的

- コンテクスト依存型情報推薦システムの開発
- 提案手法を適用した情報推薦システムのプロトタイプの実装。

コンテクスト

コンテクストとは、一般にユーザを取り巻くさまざまな状況を意味するが、本研究では「人の情報選択の判断に対する意思決定に影響を及ぼす状況」と定義する。

以下のコンテクストを考慮した情報推薦システムについて検討する。

コンテクスト

- 時間情報：季節、曜日、時刻、etc.
- 気象情報：天気、気温、湿度、etc.
- ユーザ情報：予算、目的、気分、体調、etc.
- 同伴者情報：関係（友人、恋人など）、人数、etc.

SVM を用いた場合のユーザ嗜好モデリング

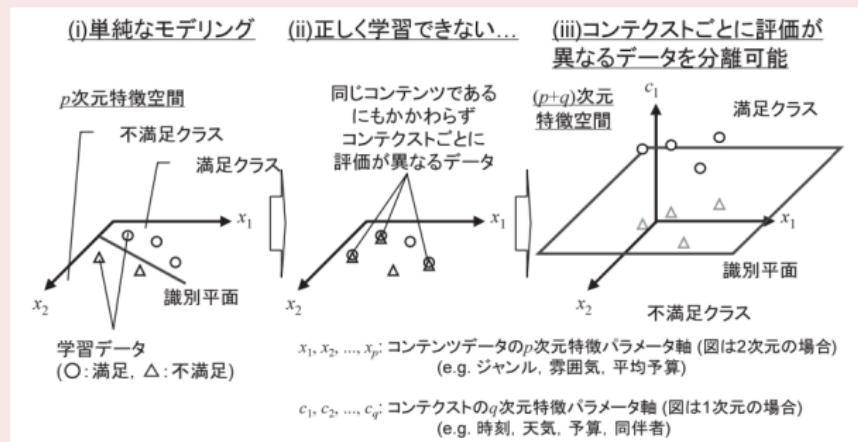


図 1: コンテクスト依存型ユーザ嗜好モデリング

- 人の嗜好は同じコンテンツであるにもかかわらず、コンテクストごとに価値が異なる。 (ii)
- この手順 (i) では、その時々のコンテクストに応じて多様に変化するような人の嗜好をモデル化することはできない。

SVM を用いた場合のユーザ嗜好モデリング

コンテンツの特徴を表すパラメータに加えてコンテキストの特徴を表すパラメータも含めて扱う。

具体的なモデリング手順は (iii)

- ① コンテンツデータを p 次元の特徴パラメータ, コンテキストを q 次元の特徴パラメータの形式で表現する.
- ② ユーザはコンテンツデータに対して 1 つ 1 つそのときのコンテキストにおける好みの評価（満足・不満足）を与える.
- ③ 評価付けされたコンテンツデータの特徴パラメータとその評価時のコンテキストの特徴パラメータとを統合したデータ集合を学習データ集合として $(p + q)$ 次元特徴空間上に配置する.
- ④ 学習データ集合を満足クラスと不満足クラスとに分類する識別平面を構成する.

このモデリング手法をコンテキスト依存型ユーザ嗜好モデリング手法（CP モデリング）とよぶ。

CP モデリングにより, コンテキストに応じて変化するユーザ嗜好が適切に学習されたユーザ嗜好モデル（CP モデル）を構築することができる。

ユーザ嗜好型モデリングに基づくユーザ類似度の算出手順

6/26

ユーザの評価時のコンテキストを考慮しないと、ユーザの嗜好の類似度を適切に算出することはできない問題に対して、互いのユーザの CP モデルの類似度に基づいてユーザ類似度を算出する方法をとる。

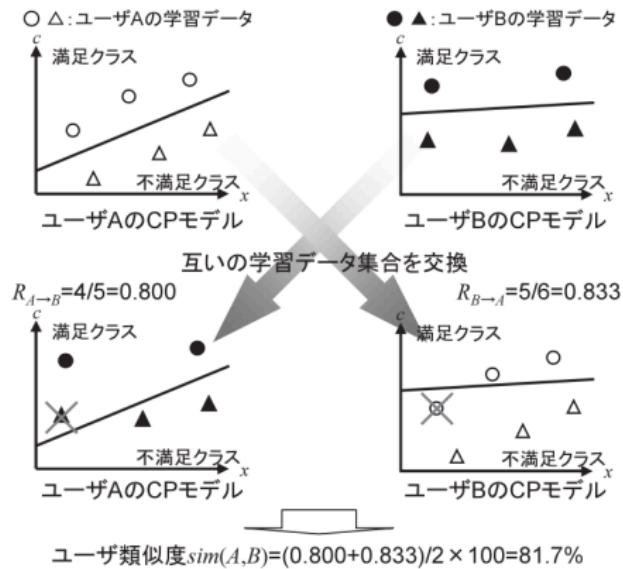


図 2: CP モデルに基づくユーザ類似度の算出

ユーザ類似度の算出手順

- ① 各ユーザの学習データを互いの CP モデル上に交換する
- ② ユーザ A の CP モデルにより、ユーザ B の学習データ 1つ1つに対して属するクラス（満足・不満足）を判定する。
- ③ そのユーザ A のモデルにより判定されたクラスとユーザ B の評価値とが一致したデータ数 ($M_{A \rightarrow B}$) を数える。
- ④ ユーザ B の全学習データ数 (N_B) に対する一致データ数の割合 $R_{A \rightarrow B}$ を求める。
- ⑤ ユーザ A の学習データ集合に対しても同様に行い $R_{B \rightarrow A}$ を求める。
- ⑥ $R_{A \rightarrow B}$ と $R_{B \rightarrow A}$ との平均値をユーザ類似度 $sim(A, B)$ とする。

以上の手順を一般化すると、ユーザ u とユーザ v とのユーザ類似度は以下の式で求まる。

$$sim(u, v) = \frac{1}{2} \left(\frac{M_{v \rightarrow u}}{N_u} + \frac{M_{u \rightarrow v}}{N_v} \right) \times 100(%) \quad (1)$$

このユーザ類似度が高いユーザ同士を類似ユーザとして扱う。

情報推薦方式

- コンテクスト依存型情報フィルタリング (C-IF)
- コンテクスト依存型協調フィルタリング (C-CF)

コンテクスト依存型情報フィルタリング (C-IF)

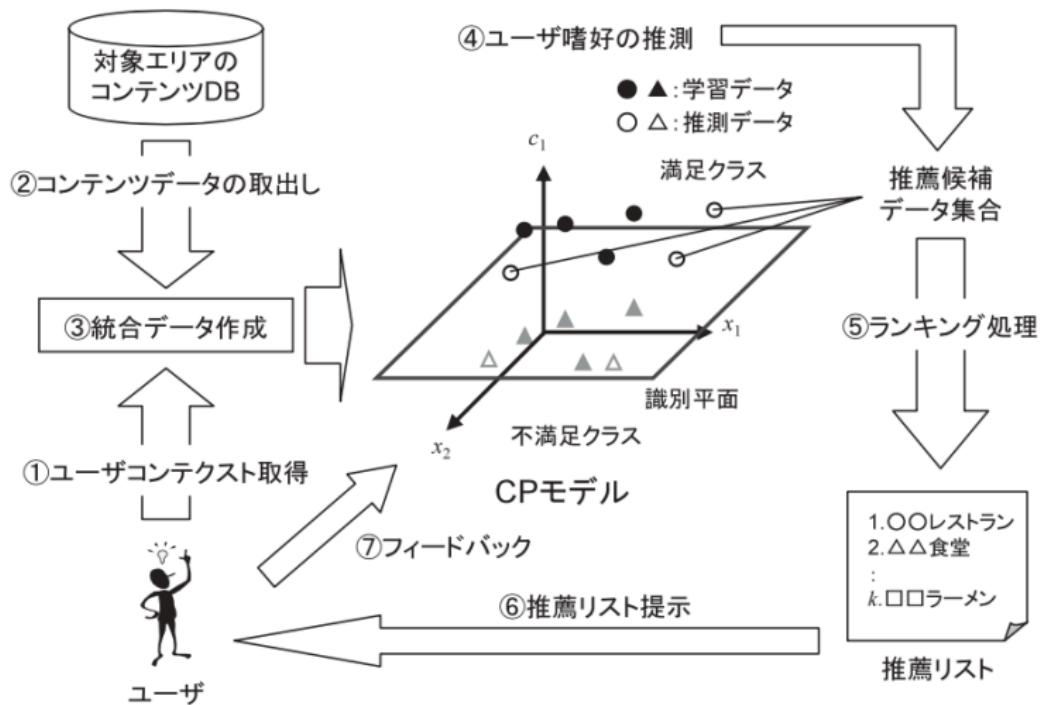
- 対象ユーザの CP モデルに基づいて、データベースから、ユーザのおかれているコンテクストにおけるユーザの嗜好に合致するコンテンツを検索し、そのコンテンツをユーザに推薦する。

コンテクスト依存型協調フィルタリング (C-CF)

- ユーザ同士の CP モデルの類似度に基づいて、類似ユーザを探し出し、対象ユーザのコンテクスト近傍において、その類似度ユーザが満足と評価しているコンテンツを対象ユーザに推薦する。

コンテクスト依存型情報フィルタリング

9/26



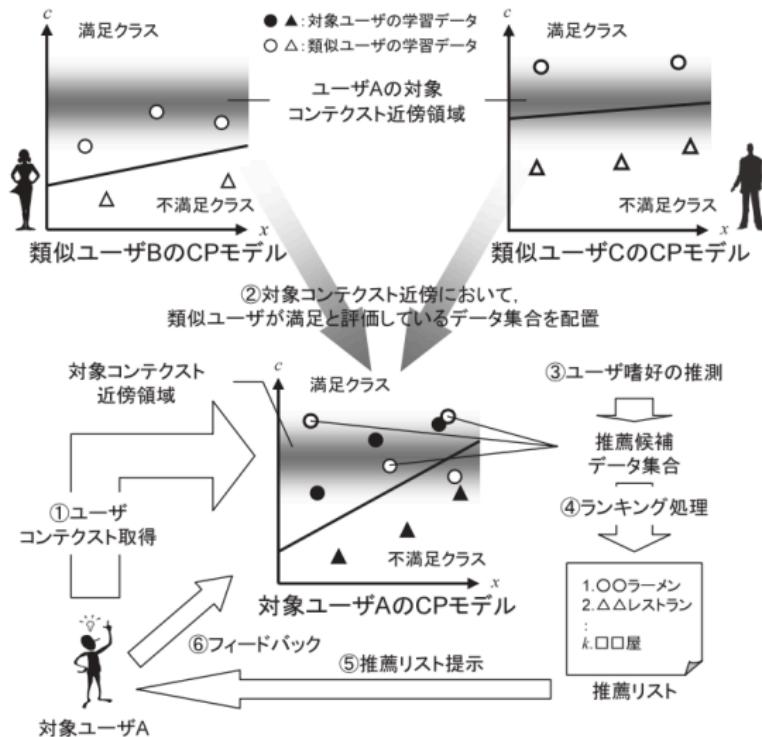


図 4: コンテクスト依存型協調フィルタリング (C-CF) によるシステムの動作手順

ユーザ嗜好モデリング手法として SVM を適用することの妥当性の検証

11/26

- 情報推薦システムは、ユーザが把握しきれていないコンテンツを適切に推薦することに価値がある。
- 未知データに対して高い汎化性能を示すことが望ましい。

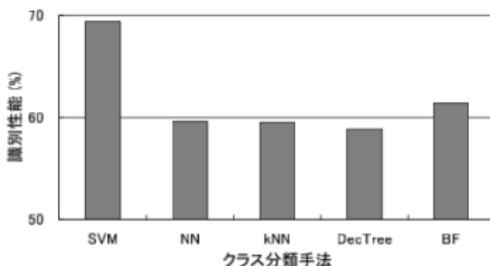


図 5: 実データセットを対象とした時の未知データに対する識別性能の比較

- クラス分類手法として、SVM, NN, kNN, DecTree, BF の 5 つに対して定量評価実験を行った結果、Support Vector Machine (SVM) が比較的高い汎化性能を示した。

被験者満足度によるシステムの評価

12/26

データセット

- 飲食店を対象としたデータセットを用いる.
- 「Yahoo グルメ」を参考に図 6 に示すエリアを対象とする.
- 登録されているデータ件数のうち、有効なデータだけを用いる.
- 図 7 に示す 28 次元の飲食店特徴パラメータを設定.
- コンテクスト特徴パラメータも図 8 に示す 24 次元の特徴パラメータを設定.

エリア	登録件数	有効データ数
I 心斎橋, 堀江	2,069	650
II 天王寺, 阿倍野	416	121
III 梅田, 中津	1,508	445

図 6: 飲食店データセット

パラメータ項目	パラメータ値
ジャンル (8)	[和食] / [洋食] / [中華] / [居酒屋] / [アジア] / [ラーメン] / [カフェ, スイーツ] / [その他]
予算 (昼夜) (2)	0~10,000 円
店の特徴 (11)	[個室] / [駐車場] / [禁煙席] / [カラオケ] / [深夜営業] / [食べ放題] / [ランチ] / [カード可] / [ライブ] / [テイクアウト] / [貸し切り可]
雰囲気 (7)	[デート向き] / [接待向き] / [宴会・パーティー向き] / [合コン向き] / [ファミリー向き] / [1人でも可] / [少人数向き]

図 7: 飲食店特徴パラメータ

パラメータ項目	パラメータ値
時間情報	月 (2) 1~12 月
	曜日 (2) 日~土
	時刻 (2) 00:00~23:59
気象情報	天気 (4) [晴れ] / [曇り] / [雨] / [雪]
	気温 (1) 0~40°C
ユーザ情報	休日 (2) [休日] / [休日前]
	予算 (1) 0~10000 円
	余裕時間 (1) 0~4 時間
同伴者情報	[1 人] / [家族] / [恋人] / [友人・先輩・後輩] /
	グループ (7) [上司・部下] / [ビジネス] / [該当なし]
	人数 (男女) (2) 0~10 人

図 8: コンテクスト特徴パラメータ

被験者と評価指標

- 飲食店データに対する評価（満足・不満足）を被験者に行ってもらう。
- 被験者は、研究室の学生及び教官で9人である（被験者 a,b,...,i）。
- いずれの被験者も図6に示したエリアには何度か訪れたことがあり、これらのエリアの雰囲気などについてはある程度把握している。
- 被験者が把握しているエリアを実験対象として選択したのは、できるだけ被験者にユーザコンテクストを想像しやすくするため。
- 被験者満足度を用いてシステムの評価を行う。
- 満足度とは被験者が情報推薦に対してどの程度満足したかを表す指標であり、以下の式で表される。

$$\text{満足度} = \frac{\text{被験者が満足した回数}}{\text{推薦回数}} \times 100(\%)$$

ランダム提示およびコンテクスト非依存の推薦方式との比較

C-IF および C-CF による情報推薦の有効性をランダム提示およびコンテクストに依存しない推薦方式との比較を行うことで評価する。

実験方法

- 本実験で用いるデータセットの対象エリアは、図 6 に示す内、「心斎橋、堀江」エリアとする。

各被験者のユーザ嗜好モデルを構築する。

ユーザ嗜好モデルの構築

- ① 被験者には、各自「心斎橋、堀江」エリアにおいて、現実に起こりうるコンテクストを想像してもらい、あらかじめ 10 パターンのコンテクストを設定してもらう。
- ② 被験者には、1 つのコンテクストあたり、データセットの中からランダムに選択された 10 件の飲食店データに対して、満足・不満足の評価を与えてもらう。
- ③ この評価に基づいて、被験者 1 人あたり合計 100 個の学習データを作成する。
- ④ 作成された学習データの基づいてコンテクスト依存/非依存型のユーザ嗜好モデルを構築する。
- ⑤ コンテクスト依存型ユーザ嗜好モデルに基づいて各被験者間のユーザ類似度を算出する。
ここで、各被験者に対してユーザ類似度が最も高い 3 人の被験者を C-CF における類似ユーザとして扱う。

以下の各方式で被験者に対して飲食店を推薦する。

飲食店の推薦方式

- ランダム提示：被験者の嗜好及びコンテクストに関係なく、ランダムの提示する。
- コンテクスト非依存型推薦：被験者のコンテクストは考慮しないが、ある程度の嗜好に基づいて推薦する。
- C-IF による推薦：コンテクストごとに変化する被験者の嗜好を考慮して推薦する。
- C-CF による推薦：コンテクストも考慮した類似ユーザの嗜好に基づいて推薦する。

飲食店の推薦方式

- 上記の 4 つの方式により推薦された飲食店に対して、満足・不満足の評価を与える。
- その評価に基づいて被験者満足度を算出する。

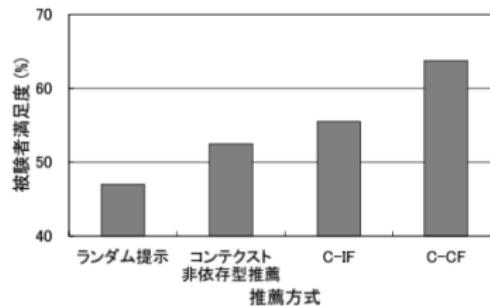


図 9: 各推薦方式に対する被験者満足度の比較

実験結果

- 図 9 に各方式に対する被験者満足度を示す.
- 9人の被験者の満足度の平均を示す.
- 図 9 に示すように、C-CF に対する満足度が最も高く、次に C-IF に対する満足度が高い.

考察

- ランダム提示は提示する飲食店に根拠がないため、最も満足度が低い。
- コンテクスト非依存型推薦は、コンテクストで変化するユーザの嗜好を正しくモデル化できていないことから、ときにコンテクストにふさわしくない飲食店が推薦されたことなど満足度が低くなった原因であると考える。
- C-IF は、コンテクストごとに変化するユーザの嗜好を正しくモデル化しているため、コンテクストに合った飲食店が適切に推薦されたため、比較的高い満足度が得られた。
- C-CF は、コンテクストに適合し、かつ 1 度他のユーザが満足と評価した飲食店を推薦しているため、推薦情報に対する信頼度が高く、結果、最も高い満足度が得られた。

実験方法

- ① 9人の被験者を3人ずつ3つの被験者グループ (A : (a,b,c), B : (d,e,f), C : (g,h,i)) に分ける。
 - ② 各被験者グループはそれぞれ対応するエリア (A : エリア I, B : エリア II, C : エリア III) で学習データを作成する (図6参照)。
 - ③ 3つのエリアにおいてC-IF および C-CF により各被験者に対して飲食店を推薦する。
 - ④ 被験者は推薦された飲食店に対して評価 (満足・不満足) を与える。
-
- 被験者 a がエリア II を推薦対象エリアとしていた場合、そのエリアにおいて学習データを作成した被験者 d,e,f が、被験者 a の類似ユーザとなる。
 - 被験者が学習データを作成したエリアを、その被験者にとって「既知エリア」とし、既知エリア以外を「未知エリア」とよぶ。

実験結果

21/26

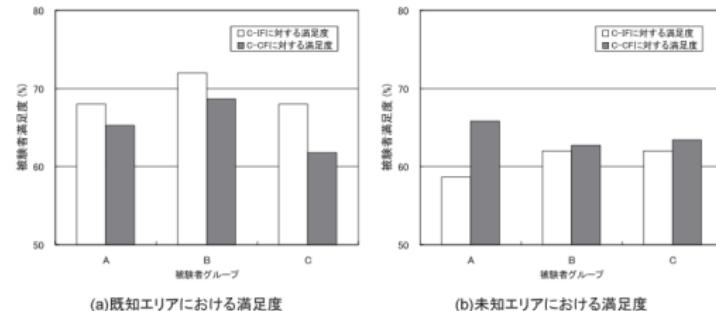


図 10: 既知/未知エリアにおける C-IF および C-CF に対する満足度の比較

- (a) は既知エリアを推薦対象としたときの, (b) は未知エリアを推薦対象としたときの被験者満足度

得られた結果

- 既知エリアを対象とした場合には、C-IF に対する満足度の方が高い
- 未知エリアを推薦対象とした場合には、C-CF に対する満足度の方が高い

特徴パラメータの最適化に関する検討

22/26

- 特徴パラメータの中には、ユーザの情報選択の判断に対する意思決定に影響をいい意味で与えるものと、ノイズとして与えるものがある。
- 各特徴パラメータがモデルの識別性能に寄与している度合いを分析することで最適化に近づく。
- 本実験の被験者に対して「情報選択の判断において、参考にした項目の順位」に関してヒアリング調査を行った。

飲食店情報

- ① ジャンル
- ② 予算
- ③ 雰囲気

コンテキスト項目

- ① 同伴者情報
- ② 時刻
- ③ ユーザ情報

実験方法

- ① あらかじめ構築した CP モデルの識別性能を得ておき、これを標準時の識別性能とする。
- ② 図 7 に示す飲食店特徴パラメータを「ジャンル」、「予算」、「店の特徴」、「雰囲気」の 4 組に、図 8 に示すコンテクスト特徴パラメータを「月」、「曜日」、「時刻」、「気象情報」、「ユーザ情報」、「同伴者情報」の 6 組にそれぞれ分ける。
- ③ 学習データを構成する特徴パラメータから、2 で挙げた特徴パラメータを 1 組ずる排除する。
- ④ 1 組の特徴パラメータが排除された学習データ集合から各被験者の CP モデルを新たに構築し、そのモデルの識別性能を得る。
- ⑤ 特徴パラメータ排除後の識別性能と標準時の識別性能を比較する。
- ⑥ これをすべての組について行う。

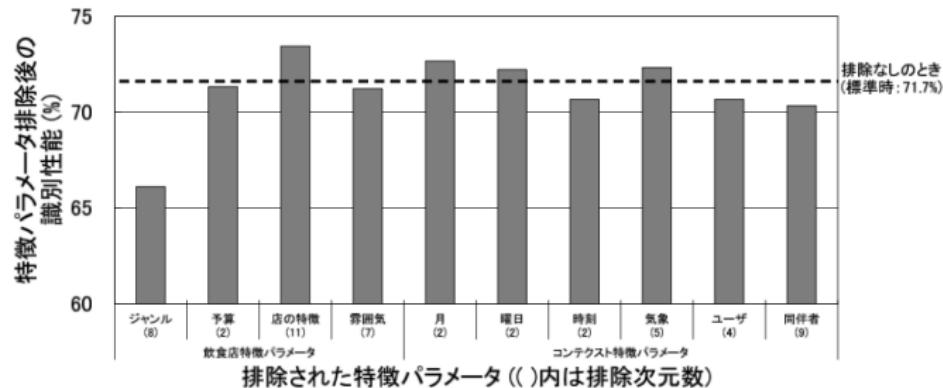


図 11: 特徴パラメータがモデルの識別性能に与える影響の分析

実験結果

- 「ジャンル」パラメータは、本実験においてモデルの識別性能の向上に大きく寄与している。
- 「店の特徴」、「月」、「気象情報」、「曜日」の4組はノイズとなった特徴パラメータである。

考察

- ヒアリングの結果と本実験結果から得られた順位とがほぼ一致したことから、ユーザの情報選択の判断基準となる特徴パラメータを抽出することが可能である。
- 図 12 に示すように、ノイズとなった特徴パラメータを排除していくことによって識別性能を高めていくことが可能であるという結果が得られたことから、本実験で行った分析方法は、特徴パラメータ最適化手法として、適用しうるものであるといえる。

特徴パラメータの組 (× : 排除パラメータ)					識別性能 (%)
店の特徴	月	気象情報	曜日	予算	
-	-	-	-	-	71.7
×	-	-	-	-	73.4
×	×	-	-	-	73.8
×	×	×	-	-	74.1
×	×	×	×	-	74.7
×	×	×	×	×	73.6

図 12: 特徴パラメータを累積的に排除したときの識別性能の変化

まとめ

C-IF および C-CF に対して以下の特長を明らかにした。

- 既知エリアを推薦対象とした場合には、C-IF による情報推薦が有効に働く。
- 未知エリアを推薦対象とした場合には、C-CF による情報推薦が有効に働く。

各特徴パラメータが識別性能に与える影響を分析することによって、特徴パラメータの最適化が可能であることを示した。

課題

- ユーザの意思決定に影響を及ぼすコンテキストとして、周辺環境や時系列的要因も考慮に入れた情報推薦システムについての検討
- 飲食店だけでなく音楽情報や映画情報など、さまざまなコンテンツを推薦対象とした情報推薦システムの開発を行う。