

卒業論文

飲食店推薦のための機械学習アルゴリズムの比較

関西学院大学 理工学部 情報科学科

27020729 須崎 良祐

2023年3月

指導教員：北村 泰彦 教授

内容梗概

本研究は、検索と推薦の違いに焦点を当て、特に飲食店の推薦において、どの機械学習アルゴリズムが最適かを明確化することを目的としている。検索はユーザーがアクティブに情報を求める一方で、推薦はユーザーの好みに基づいて自動的にアイテムを提示する。

本研究では、飲食店に関するデータを入力変数に変換し、飲食店に対する評価を目的変数に変換し、決定木、ロジスティック回帰、k-NN の 3 つのモデルで 10 フォールド交差検証を行った。

実験は、十分な飲食店情報と評価データが利用可能な場合と、それらの情報が限られている場合の両方で実施された。その結果、情報が豊富な場合には決定木が最良の結果を示した。一方、情報が限られている場合にはロジスティック回帰が最良の結果を示した。

この研究は、機械学習を用いた推薦アルゴリズムの開発において、使用するアルゴリズムの選択がデータの量や質に依存することを示している。それにより、各シナリオに最適なアルゴリズムの選定に関する洞察を提供している。

目次

第1章	はじめに	1
第2章	推薦システム	3
2.1	推薦システムに用いる用語の定義	3
2.1.1	推薦システムとは	3
2.1.2	既存の推薦手法	3
2.2	推薦システムに関する研究	4
2.2.1	決定木を用いたユーザー群に対しての飲食店推薦システム	4
2.2.2	個人に対しての飲食店推薦システム	6
2.2.3	決定木を用いた個人に対しての音楽推薦システム	6
2.3	本研究の目的	8
第3章	飲食店推薦のための機械学習アルゴリズムの比較	9
第4章	評価	10
第5章	まとめ	11
付録A	Latexの書き方	12
	謝辞	13
	参考文献	13

目 次

2.1	顧客情報と飲食店情報の詳細	4
2.2	システムへの入力	5
2.3	システムへの出力	5
2.4	変更する店舗と推薦される店舗	6
2.5	システムの構成	7
2.6	含まれている属性	7
A.1	サンプル	12

表 目 次

A.1 クラスタ数と選択肢の割合	12
----------------------------	----

第1章 はじめに

推薦システムは、ユーザーの過去の行動や好みに基づき、自動的に関連アイテムを提案する技術である。特に、本研究では飲食店推薦に焦点を当てている。このようなシステムは、ユーザーがまだ発見していない新しい飲食店を提案することで、個人化された食体験を提供する。

推薦システムの一般的な応用例としては、Netflix や Amazon のようなサービスがあり、これらはユーザーの視聴履歴や購入履歴に基づいて映画や商品を推薦する。しかし、飲食店の推薦の文脈では、この概念を飲食店の選択に応用している。

これは、ぐるなびや食べログなどの検索システムとは明確に異なる。これらの検索システムでは、ユーザーは特定のルールを入力し、システムはそれに最も適した飲食店情報を提供する。一方で、推薦システムでは、ユーザーの飲食店に対する過去の評価を分析し、ユーザー自身が未知である可能性のある新しい飲食店を提案する。つまり、推薦は自動的にユーザーに対して飲食店情報を提供し、検索はユーザー自身が飲食店情報を見つけ出さなければならないのである。

ユーザーの手間がかからない点や新しい未知の飲食店を発見できる可能性が高い点は、飲食店推薦システムの大きな価値となっている。飲食店の推薦は、ユーザーにとって便利であり、新鮮な食体験を提供する機会を創出する。

近年では推薦システムの研究は盛んに行われているが、本研究では3つの研究を関連研究として取り上げる。1つ目は、決定木を用いたユーザー群に対しての飲食店推薦システムの研究 [1] であるが、ユーザー群に対しての推薦システムになっているため、個人の好みに着目した推薦をすることができない。2つ目は、ユーザー個人の取捨選択によって各属性に対する重み変動し、そのユーザーの好みに沿った飲食店を提示できる推薦システムの研究 [2] であるが、推薦に用いる飲食店の属性として、価格帯とジャンルしか扱っていない。3つ目に、決定木を用いたユーザー個人で利用できる音楽推薦システムの研究 [3] であり、決定木と他の機械学習アルゴリズムを比較して、決定木の有効性を示している。しかし、飲食店推薦の文脈で、どの機械学習アルゴリズムが有効であるか明確でない。

そこで、本研究では、内容ベースフィルタリングを用いた飲食店推薦に、代表的な機械学習アルゴリズムのうち、どのアルゴリズムを使えば良いかを明確化する。

評価では、どのアルゴリズムを使えば良いかを明確化するために、独自に集めた飲食店情報とそれに対するユーザーの評価を用いて、各機械学習アルゴリズムで10フォールド交差検証を行い、その性能を示す。そして、最も性能が良かったアルゴリズムの精度と他のアルゴリズムの精度に統計的な有意差が存在するかど

うかも確認する．また，学習用データを増やした際の精度の推移を示す．

本論文の構成は以下の通りである．第2章では，専門用語の定義と関連研究，及び本研究の目的について述べる．第3章では，本研究で比較する機械学習アルゴリズムの詳細について述べる．第4章では，結果その考察を述べる．第5章では，本研究のまとめと今後の課題について述べる．

第2章 推薦システム

2.1 推薦システムに用いる用語の定義

2.1.1 推薦システムとは

推薦システムは、ユーザーの行動パターンに基づいて、商品、サービス、情報を自動的に提案するシステムである。このシステムの重要性は、インターネットの発展とともに急速に増しており、背景として情報過多の問題がある。現代のデジタル社会で、ユーザーが直面する情報の量が爆発的に増加し、関連性の高い、または興味を引くコンテンツを効率的に見つけることが困難になっているためである。

推薦システムは、この情報の海から有用な情報を引き出し、個々のユーザーにカスタマイズされた推薦を行うことで、情報選択のプロセスを簡素化し、ユーザー体験を向上させる。

推薦システムの開発は、ユーザーのニーズを理解し、パーソナライズされたサービスを提供することに重点を置いている。これにより、ユーザー満足度を高め、ビジネスにおいては顧客の関与を深め、売上向上に寄与することが期待される。

2.1.2 既存の推薦手法

推薦システムにおいて広く使用される二つの主要な手法は、以下の通りである。

内容ベースフィルタリング

内容ベースフィルタリングは、ユーザーのアイテムに対する過去の評価を分析し、アイテムの各属性とユーザーの評価の関係をモデル化し、そのモデルを用いて新たなアイテムを推薦する手法である。つまり、アイテムの内容に着目している。

協調フィルタリング

協調フィルタリングは、多くのユーザーのアイテムに対する評価を分析し、似た評価パターンを持つユーザー、あるいはユーザーが高評価を与えたアイテムに似た評価パターンを持つアイテムを見つけ出し、新たなアイテムを推薦する手法である。

つまり、内容ベースフィルタリングはユーザーの評価とアイテムの内容の関係を用いる推薦手法であり、協調フィルタリングは他ユーザー、あるいは他アイテムの評価パターンを用いる推薦手法である。

本研究では、内容ベースフィルタリングを用いた推薦に焦点を当てている。

2.2 推薦システムに関する研究

2.2.1 決定木を用いたユーザー群に対しての飲食店推薦システム

この研究では、決定木を用いて、飲食店を低価格帯、中価格帯、高価格帯カテゴリに分類し、ユーザーを低予算、中予算、高予算カテゴリに分類し、各カテゴリがマッチする飲食店をユーザーに推薦するシステムの開発を行っている。この研究で用いたデータセットには表 2.1 のような 95 件の飲食店情報と、135 人の顧客情報と顧客のいくつかの飲食店に対する評価情報が含まれている。これらの情報を使って、ユーザーと飲食店を各予算カテゴリに決定木を構築している。具体的なシステムへの入力、出力はそれぞれ表 2.2、表 2.3 に示す通りで、個人情報や個人の飲食店に対する好みを入力し、それに基づいて、いずれかの価格帯の飲食店情報が出力される。

しかし、この研究で開発しているシステムは、ユーザー群に対しての推薦システムになっており、個人の好みに着目した推薦を行うことができないという課題を抱えている。

Customer Master	Restaurant Master
userID	placeID
Latitude	Rcuisine
Longitude	Rpayment
Smoker	Parking
drink_level	Weekday time
dress_preference	Sat
Ambience	Sun
Transport	Latitude
marital_status	Longitude
Hijos	Name
birth_year	Alcohol
Interest	smoking_area
Personality	dress_code
Religion	Accessibility
Activity	Price
Color	Rambience
Weight	Franchise
Budget	Area
Height	other_services
Cuisine	
Payment	

図 2.1: 顧客情報と飲食店情報の詳細

Customer specifications Input				
Smoking	FALSE	TRUE		
	0	1		
Interest	variety	retro	eco-friendly	technology
	0	1	1	1
drink level	abstemious	casual drinker	sodal drinker	
	0	0	1	
dress preference	formal	informal	no-preference	elegant
	0	1	0	0
ambience	family	friends	solitary	
	1	1	0	
personality	conformist	hard-worker	hunter-ostentatious	thrifty-protector
	0	1	0	0
Activity	student	professional	unemployed	working class
	0	1		0

図 2.2: システムへの入力

Restaurant Recommendation system							
Restaurant Recommendation system	Distance	Cuisine	Weekday	Saturday	Sunday	Parking	Area
Name	Distance	Cuisine	Weekday	Saturday	Sunday	Parking	Area
Restaurant Bar Coty y Pablo	1,941.04	Bar	16-24	16-24	16-24	none	closed
El cotorreo	1,943.37	Family	0-8	0-8	0-8	none	open
Cafeteria cenidet	1,944.12	Cafeteria	0-8	0-8	0-8	public	closed
Log Yin	1,946.28	Mexican	16-24	16-24	16-24	yes	closed
Subway	1,946.80	Fast Food	16-24	16-24	16-24	public	closed
McDonalds Centro	1,946.90	American	16-24	16-24	16-24	none	closed
TACOS CORRECAMINOS	2,169.71	Mexican	16-24	16-24	16-24	none	closed
TACOS EL GUERO	2,170.56	Mexican	16-24	16-24	16-24	none	closed
tacos de la estacion	2,170.57	Mexican	0-8	0-8	0-8	none	open
little pizza Emilio Portes Gil	2,171.78	Armenian	16-24	16-24	16-24	none	closed
palomo tec	2,171.84	Mexican	16-24	16-24	16-24	none	closed
tacos de barbacoa enfrente del Tec	2,171.86	Mexican	8-16	8-16	8-16	public	open
Carreton de Flautas y Migadas	2,171.90	Mexican	0-8	0-8	0-8	none	open
puesto de gorditas	2,171.92	Regional	0-8	0-8	0-8	public	open
tacos abi	2,171.93	Mexican	16-24	16-24	16-24	none	closed
Hamburguesas La perica	2,172.16	Mexican	16-24	16-24	16-24	public	open

図 2.3: システムへの出力

2.2.2 個人に対しての飲食店推薦システム

この研究では、飲食店情報（ジャンルと価格帯）から、飲食店の取捨選択を行い、ジャンルと価格帯それぞれの重み変動し、徐々にユーザーの好みに合った飲食店を推薦するシステムの開発を行っている。具体的なシステムへの入力、出力は表 2.4 に示す通りで、ユーザーは適当に飲食店を選択し、その飲食店に対して評価（良い、悪い）を入力し、その飲食店が、新しい飲食店に変更される。

しかし、この研究で開発しているシステムは、推薦に用いる飲食店の属性として、価格帯とジャンルしか扱っていないという課題を抱えている。



図 2.4: 変更する店舗と推薦される店舗

2.2.3 決定木を用いた個人に対しての音楽推薦システム

この研究では、推薦に決定木、k-means、凝集法を用いた場合の評価を行い、決定木の有効性を示し、決定木を用いて音楽推薦システムの開発を行っている。システムの構成は表 2.5 に示す通りで、ユーザはいくつかの音楽データに対して評価（好き、嫌い、どちらでもない）を行い、その音楽データからシステムは曲の特徴量を抽出する。そして、ユーザの好みを表すユーザープロファイル（決定木）を作成し、それを元に推薦するかどうかを判定する。この研究で用いたデータセットには、表 2.6 のような属性が含まれている 200 曲分の音楽データ（入力変数）と、それらに対する 10 人分の評価情報（目的変数）が含まれている。各機械学習アルゴリズムの評価では、これらのデータセットを用いて、ユーザーの評価情報の予測に、それぞれの機械学習アルゴリズムを用いた場合の性能の比較を行っている。

しかし、飲食店推薦の文脈で、どの機械学習アルゴリズムが有効であるか明確でない。

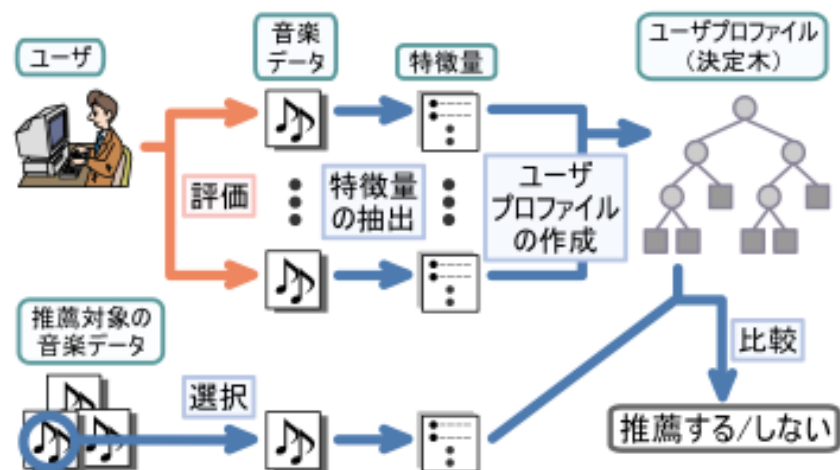


図 2.5: システムの構成

拍子 (曲全体)
調性 (曲全体)
平均テンポ (曲全体)
リズム (曲全体)
メジャーコードの割合 (曲全体)
マイナーコードの割合 (曲全体)
sus4 コードの割合 (曲全体)
キー (曲全体)
音色 (メロディCH)
平均音高差 (M-CH, B-CH, C-CH)
平均音長 (M-CH, B-CH, C-CH, D-CH)
平均音長差 (M-CH, B-CH, C-CH, D-CH)
M-CH : メロディCH, B-CH : ベースCH
C-CH : コードCH, D-CH : ドラムCH

図 2.6: 含まれている属性

2.3 本研究の目的

これらの課題を解決するために，本研究では，内容ベースフィルタリングを用いた飲食店推薦で，どの機械学習アルゴリズムを使えばいいかを明確化する．

第3章 飲食店推薦のための機械学習 アルゴリズムの比較

設計方針（目的達成に必要な機能の一覧），全体構成と各モジュールの役割や関係，各モジュールの詳細，システムの利用例，について書く．モジュールの説明においては，入力と出力を明確にすること．

第4章 評価

評価の目的，方法，結果，考察（研究の目的がどこまで達成できたのかについて），今後の課題について書く．実験結果はグラフや表にまとめる．実験結果に関しては必要に応じて，統計的検定を行う．考察に関してはその根拠となる実験結果を示すこと．

第5章 まとめ

研究の目的と結果を簡潔に書く．今後の課題に関して簡潔に書く．

付 録 A Latexの書き方

図 A.1 のように，図を挿入するときは必ず figure 環境を用い，label，caption をつけること．図は必ず本文中で引用し，説明を加えること．図は eps 形式にすること．パワポで図を書く場合は，InkScape に図をコピーし，eps 形式で保存すること．

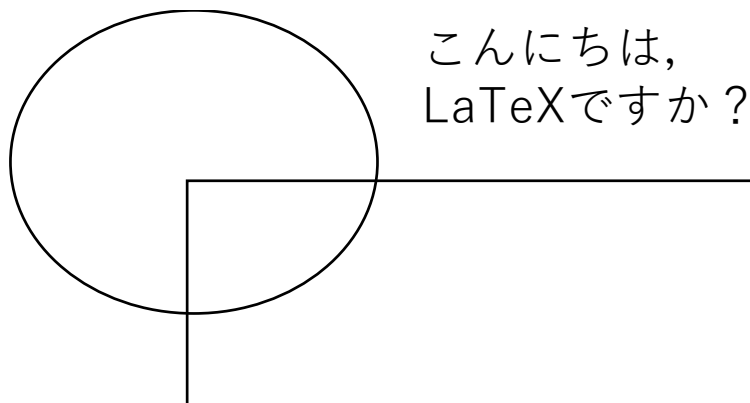


図 A.1: サンプル

表 A.1 のように，表を挿入するときは必ず table 環境を用い，label，caption をつけること．表は必ず本文中で引用し，説明を加えること．

表 A.1: クラスタ数と選択肢の割合

クラスタ数	当然	おもしろい	無意味	無回答	計
1 (10 例)	189	233	46	2	470
2 (8 例)	110	219	43	4	376
3 (2 例)	22	60	12	0	94
計	321	512	101	6	940

謝辞

（本当はそう思っていないけど）本研究を進めるにあたり，終始ご指導下さった北村泰彦教授に深く感謝いたします。

また，音声認識に関して協力していただいた人間システム工学科川端豪教授に厚くお礼申し上げます。ほかにお世話になった人もついでに書いておく。

北村研究室の同輩諸氏には，日頃から多くの（つまらん）ご助言や（いらない）ご支援をいただき，大変感謝しています。

最後に，4年間大学に通わせてくれた両親とペットのひなのに，心から感謝しています。

参考文献

- [1] Surajit Ghosh Dastidar. Restaurant recommendation system. *International Journal of Business Analytics and Intelligence*, 2017.
- [2] 児玉礼, 北山大輔. ユーザーの取捨選択行動に基づく嗜好情報による飲食店推薦システム. 第8回データ工学と情報マネジメントに関するフォーラム, F2-2, 2016.
- [3] 土方嘉徳, 岩濱数宏, 西田正吾. 決定木を用いた内容に基づく音楽情報フィルタリングとその有効性の検証. 情報処理学会, 研究報告, 2004.