

基礎から学ぶ推薦システム

— 情報技術で嗜好を予測する —

博士(工学) 奥 健太 著

コロナ社

ま え が き

「この商品を買った人はこんな商品も買っています」。いまや、Web上のいたるところでみかけます。オンラインショッピングサイトで、ある商品を購入しようとする、その商品に関連するおすすめの商品が提示されます。動画共有サイトにアクセスすると、あなたにおすすめの動画が提示されます。観光情報サイトで興味のある地域の情報を閲覧すると、その地域のおすすめの観光スポットやレストラン、ホテルなどの情報が提示されます。もしかしたら、あなたもおすすめの本として提示されて、本書を手にとられたのかもしれない。このような機能を支えている技術が推薦システムです。

さて、この推薦システムの中身はどのようになっており、どのようにわたしたちの好みを言い当てているのでしょうか。本書では、その推薦システムの中身に迫ります。推薦システムにおいては、その人が何を好むのか、人の嗜好をいかに予測するのが課題となります。その予測の鍵を握るのがデータ、すなわち、その人がこれまでに利用してきたコンテンツに対する評価履歴です。その評価履歴からの嗜好予測には、機械学習をはじめとしたさまざまな情報技術を駆使します。人の嗜好は多種多様で複雑です。そのような人の嗜好を情報技術でどこまで予測できるのか、これが推薦システムにおける究極的な問いであるといえます。この問いに対して挑みつづけることが、推薦システム研究の醍醐味ともいえるでしょう。

情報技術は、数学によって支えられています。そのため、本書では数式がふんだんに出てきます。数式が苦手な方にとっては少々難しく感じるかもしれませんが、しかし、安心してください。本書で出てくる数式については、一つひとつ紐解きながら丁寧に解説していきます。具体例を交えながら、計算過程もなるべく省略せずに書いていますので、一つひとつ記号の意味を理解しながら式を追っていくことができます。本書では、問題に親しみをもってもらいやすいように、具体的かつ身近な事例として、食をテーマにしたシナリオを設定しています。食に対する好き嫌いは個人差が現れやすいので、嗜好予測の問題設定

としてはうってつけです。また、数式の意味を理解するためには、実際に手で計算しながら式を追っていくことが重要です。問題の大きさも小規模なものにしていますので、一部の計算を除いて、プログラムに頼らなくても、(関数電卓ぐらいは必要ですが)手計算で追っていける内容となっています。本書を通して、数式の読み方にも慣れていただければ幸いです。

本書の読者対象は、これから推薦システムについて学びたいという大学生、および大学院生を想定しています。学部1~2年生レベルの微分、線形代数、確率・統計に関する知識を前提としていますが、先述したように、数式についてはその都度解説していきますので、授業で習ったことがある程度で十分です。必要に応じて数学の教科書などを参照し、知識を補いながら読み進めていただければと思います。本書をきっかけに、推薦システムの研究分野に興味をもつていただければ望外の喜びです。

また、本書は、大学の授業や研究室などで、学生に対して推薦システムや関連技術について指導される機会のある教員の方も読者対象として想定しています。推薦システムは、機械学習やデータ分析などの応用例の一つとして紹介されることが多いと思いますが、本書の目次を参照いただいてもわかるとおり、推薦システムはじつに多くの技術に関連します。そのため、推薦システム一つとっても、学生に伝えられる技術は多くあります。本書が推薦システムや関連技術の指導の一助になれば幸いです。なお、コロナ社の書籍紹介ページ^{†1}では、教科書採用者向け資料を提供しています。

一方で、本書は推薦システムの基礎に焦点をあてているため、先端的なトピックについてはカバーしていません。先端的なトピックについて興味のある方は、推薦システムの国際会議である ACM Recommender Systems¹⁾^{†2}やそれに併設されているワークショップで最新の研究動向を追うとよいでしょう。また、本書は実践向けではなく、一部の擬似コードを除いて、コードも出てきません。推薦システムの実装や実践に関しては、文献 40) や文献 178), 181) などが参考になります。もちろん、これら最新の研究動向の調査や、推薦システムの実装や実践を進めるうえで、本書で学ぶような推薦システムの基礎をしっかり押さえておくことは大事です。

^{†1} <https://www.coronasha.co.jp/np/isbn/9784339029284/>

^{†2} 肩付き数字は、巻末の引用・参考文献を示しています。

推薦システムは、大きく、内容ベース推薦システム、協調ベース推薦システム、知識ベース推薦システム、そして、それらを組み合わせたハイブリッド型推薦システムに分類されます。この分類を踏まえ、本書はつぎのような構成になっています。1章では、なぜ推薦システムが必要かという話からはじめ、推薦問題の定義について学びます。内容ベース推薦システムについては2章と3章の2章構成で、協調ベース推薦システムについては4章から6章の3章構成で、それぞれ学びます。知識ベース推薦システムは7章で学び、ハイブリッド型推薦システムは8章で学びます。そして、最後に9章で推薦システムの評価について学びます。

本書は、1章から順を追って読むことを想定しています。前の章で学んだ知識を後の章にも活用するように、知識を積み重ねていく構成となっています。例えば、3章では、ルールベース分類器や単純ベイズ分類器、決定木などの各分類器について学びます。そして、ここで学んだ各分類器は、5章で学ぶモデルベース協調フィルタリングにおいても活用します。また、6章では潜在因子モデルについて学びますが、これは欠損値を含む評価値行列の行と列の次元を同時に縮約するものです。その前の段階で、2章で評価履歴の次元削減、4章で欠損値を含む評価値行列の次元削減について学びます。このように、次元削減について簡単な問題設定から段階的に学んでいけるように構成を工夫しています。

また、コロナ社の書籍紹介ページには本書のサポートサイトを掲載しています。サポートサイトには、GitHub上に、本書の内容に対応した演習課題や計算例に沿ったPythonコード、問題集、解説スライドや動画などのコンテンツを公開していきます。

本書の内容については、記述が著者の主観にならないように、主要な文献を引用するように心がけました。特に、各トピックにおける主要なサーヴェイ論文や初期に提案された論文などをなるべく引用するようにしました。本書で紹介しきれなかった詳細部分は、これらの論文などを参照し、適宜補っていただければ幸いです。本書は、これら先人たちの知の積み重ねがあってのものです。この場をお借りして謝意を表します。

それでは、一緒に推薦システムの中身についてみていきましょう！

2022年5月

奥 健太

目 次

1. 推薦システム

1.1 推薦システムとは	1
1.2 なぜ推薦システムが必要か	2
1.3 ユーザ, アイテム, 評価値	5
1.3.1 ユーザ	5
1.3.2 アイテム	6
1.3.3 評価履歴	9
1.3.4 評価値行列	12
1.4 推薦問題の定義	13
1.5 推薦サイクル	15

2. 内容ベース推薦システム ～近傍ベース方式～

2.1 内容ベース推薦システム (近傍ベース方式) の基本	18
2.1.1 アイテムの特徴ベクトル	18
2.1.2 特徴空間	19
2.1.3 ユーザプロフィール	19
2.2 類似度に基づく推薦	21
2.2.1 類似度	21
2.2.2 推薦	24
2.3 適合性フィードバック	25
2.4 k 近傍法	28
2.4.1 距離	29

2.4.2	近傍アイテム	31
2.4.3	推 薦	32
2.5	次元削減	34
2.5.1	分散共分散行列	36
2.5.2	固有値・固有ベクトル	40
2.5.3	主 成 分	41
2.5.4	主成分得点	42
2.5.5	寄 与 率	45
2.5.6	推 薦	46

3. 内容ベース推薦システム ～モデルベース方式～

3.1	内容ベース推薦システム（モデルベース方式）の基本	48
3.1.1	アイテムの特徴ベクトル	48
3.1.2	訓練データ	49
3.1.3	予測対象データ	50
3.1.4	学習と予測	52
3.1.5	クラス分類問題	53
3.2	ルールベース分類器	54
3.2.1	ル ー ル	56
3.2.2	候補ルール	56
3.2.3	頻出ルール	58
3.2.4	相関ルール	59
3.2.5	ユーザプロファイル	61
3.2.6	嗜好予測	62
3.2.7	推 薦	63
3.3	単純ベイズ分類器	64
3.3.1	条件付き確率	66
3.3.2	問題設定	68
3.3.3	ベイズの定理	69
3.3.4	単純ベイズ仮定	70
3.3.5	ユーザプロファイル	74

3.3.6	嗜好予測	75
3.3.7	ラプラススムージング	76
3.3.8	推薦	78
3.4	決定木	79
3.4.1	ジニ係数	81
3.4.2	分割のよさ	83
3.4.3	決定木の学習	85
3.4.4	ユーザプロファイル	88
3.4.5	嗜好予測	88
3.4.6	推薦	89

4. 協調ベース推薦システム

～近傍ベース協調フィルタリング～

4.1	協調フィルタリングの基本	92
4.2	ユーザベース協調フィルタリング	94
4.2.1	ユーザ類似度	95
4.2.2	類似ユーザの選定	101
4.2.3	嗜好予測	102
4.2.4	評価値行列の補完	104
4.3	アイテムベース協調フィルタリング	105
4.3.1	アイテム類似度	107
4.3.2	アイテム-アイテム類似度行列	111
4.3.3	類似アイテムの選定	112
4.3.4	嗜好予測	114
4.3.5	評価値行列の補完	115
4.4	評価値行列の次元削減	116
4.4.1	分散共分散行列	119
4.4.2	固有値・固有ベクトル	122
4.4.3	潜在因子	123
4.4.4	嗜好予測	125
4.5	ユーザベース協調フィルタリング vs. アイテムベース協調フィルタリング	126

4.5.1 正確性とセレンディピティ	127
4.5.2 説明性	127
4.5.3 計算効率性	128

5. 協調ベース推薦システム

～モデルベース協調フィルタリング～

5.1 モデルベース協調フィルタリングの基本	129
5.2 ルールベース協調フィルタリング	133
5.2.1 ルール	135
5.2.2 候補ルール	135
5.2.3 相関ルール	135
5.2.4 学習モデル	138
5.2.5 嗜好予測	139
5.2.6 評価値行列の補完	139
5.3 単純ベイズ協調フィルタリング	140
5.3.1 問題設定	141
5.3.2 ベイズの定理	143
5.3.3 単純ベイズ仮定	144
5.3.4 事前確率	145
5.3.5 各アイテムに関する条件付き確率	146
5.3.6 ラプラススムージング	148
5.3.7 学習モデル	149
5.3.8 嗜好予測	150
5.3.9 評価値行列の補完	152
5.4 決定木に基づく協調フィルタリング	153
5.4.1 決定木の学習	156
5.4.2 学習モデル	158
5.4.3 嗜好予測	158
5.4.4 評価値行列の補完	158

6. 協調ベース推薦システム ～潜在因子モデル～

6.1 潜在因子モデルの基本	160
6.1.1 問題設定	161
6.1.2 定式化	164
6.2 勾配降下法	167
6.3 確率的勾配降下法	173
6.4 近傍ベース協調フィルタリング vs. モデルベース協調フィルタリング	179
6.4.1 正確性	179
6.4.2 シンプルさ	180
6.4.3 説明性	180
6.4.4 計算効率性	180
6.4.5 モデル表現	180
6.5 内容ベース推薦システム vs. 協調ベース推薦システム	181
6.5.1 コールドスタート問題	181
6.5.2 アイテム特徴量の抽出	182
6.5.3 正確性とセレンディピティ	183
6.5.4 説明性	183

7. 知識ベース推薦システム

7.1 知識ベース推薦システムの基本	185
7.2 制約ベース推薦システム	186
7.2.1 ユーザ要求	187
7.2.2 アイテムデータベース	188
7.2.3 適合性制約	190
7.2.4 フィルタ制約	191
7.2.5 推薦タスク	192

7.2.6 推 薦	194
7.2.7 要 求 の 緩 和	196
7.2.8 要 求 の 追 加	199
7.3 事 例 ベース 推 薦 システム	201
7.3.1 初 期 事 例	202
7.3.2 類 似 性 尺 度	202
7.3.3 批 評	210
7.4 知 識 ベース 推 薦 システム vs. 内 容 ・ 協 調 ベース 推 薦 システム	218
7.4.1 コ ー ル ド ス タ ー ト 問 題	218
7.4.2 個 人 化	218
7.4.3 制 御 性	218
7.4.4 アイテム特徴量の抽出・知識獲得	219

8. ハイブリッド型推薦システム

8.1 ハイブリッド型推薦システムの基本	220
8.2 アンサンブル型ハイブリッド	221
8.2.1 並列型ハイブリッド	222
8.2.2 直列型ハイブリッド	225
8.3 モノリシック型ハイブリッド	233
8.4 混合型ハイブリッド	236

9. 推薦システムの評価

9.1 推薦システムの評価の基本	237
9.1.1 評 価 目 的	238
9.1.2 評 価 指 標	239
9.1.3 ベースライン	239
9.1.4 ハイパーパラメタ	239
9.2 評 価 方 法	240
9.2.1 オフライン評価	240

9.2.2 ユーザ評価	246
9.2.3 オンライン評価	251
9.3 正確性に関する評価指標	253
9.3.1 嗜好予測の正確性	254
9.3.2 上位 K 推薦の正確性	259
9.4 発見性に関する評価指標	273
9.4.1 被覆率	274
9.4.2 多様性	276
9.4.3 新規性	279
9.4.4 意外性	283
9.4.5 セレンディピティ	286
引用・参考文献	289
あとがき	305
索引	308

推薦システム

Alice は 20 歳の大学生です。美味しいものを好みます。今日も、とあるクレープ店の Web サイトから美味しそうなクレープを探しています。いろいろあって悩みますね。すると、「このクレープを好む人は、こんなクレープも好みます」と、美味しそうなクレープを薦められました。

1.1 推薦システムとは

「この商品を買った人はこんな商品も買っています」。いまや、Web 上のいたるところでこのような表示をみかけます。例えば、Amazon⁹⁾^{†1,2}で、ある本の情報を閲覧していると「この商品を買った人はこんな商品も買っています」というフレーズとともに、他のおすすめの本が提示されます。また、マイストアを開くと、これまでの購入履歴や閲覧履歴を基に「おすすめ商品」が提示されます。映画を観ようと Netflix¹⁰⁸⁾^{†3}にアクセスすると、人気の映画や自身の興味に合った映画が推薦されます。音楽を聴こうと Spotify¹⁴⁰⁾^{†4}にアクセスすると、人気アーティストによるプレイリストや、気分にあったプレイリストなどが推薦されます。このような機能を支えているシステムが推薦システムです。

推薦システム (recommender system) は、ユーザの嗜好に合ったアイテム

^{†1} EC サイト。

^{†2} 本書に記載の会社名、製品名は一般に各社の商標 (登録商標) です。本文中では TM、®マークは省略しています。

^{†3} 動画ストリーミングサービス。

^{†4} 音楽ストリーミングサービス。

(商品や映画、音楽、本、動画、画像、ニュース記事など)を提示するシステムです^{†1}。文献 126)では、推薦システムを「特定のユーザに最も興味をもたれそうなアイテムを提案するソフトウェアツールおよび技術」^{†2}と定義しています。文献によっては、推薦システムは、レコメンデーションシステムや推薦エンジン、レコメンドエンジンともよばれることもあります。

1.2 なぜ推薦システムが必要か

世の中には膨大なコンテンツが存在します。例えば、Google Books Search のブログ記事⁵²⁾によると、2010年8月の時点で世界には129 864 880冊の本が存在していたといわれています。現在はずっと増えていることでしょう。Apple Music¹¹⁾^{†3}では7500万曲以上^{†4}、Spotifyでは7000万曲以上²⁸⁾^{†5}の楽曲が配信されています。IMDb (Internet Movie Database)^{†6}では800万件以上の映像タイトルが登録されています⁶⁸⁾^{†7}。

また、あらゆる Web サービスのユーザ自身によっても、日々膨大なコンテンツが生み出されています。YouTubeでは、毎分500時間以上もの動画が投稿されています^{145), 153)}^{†8}。Twitterでは毎分456000ツイートが、Instagramでは毎分46740件の写真が投稿され、Facebookでは毎秒5件のFacebookプロフィールが新しく作られています⁴⁵⁾^{†9}。

^{†1} 推薦システムのテキストとして文献 72) や文献 125)、文献 5) があります。文献 72) については邦訳された文献 73) も出版されています。また、Coursera に Recommender Systems の専門講座³⁶⁾ が提供されています。本書はこれらの内容を参考にしています。また、推薦システムに関するサーヴェイや解説記事も多くありますが、早期に日本語でまとめられた解説記事として、文献 175) や文献 168)~170) が挙げられます。

^{†2} 原文 “Recommender Systems (RSs) are software tools and techniques that provide suggestions for items that are most likely of interest to a particular user”¹²⁶⁾

^{†3} 音楽ストリーミングサービス。

^{†4} 2021年11月現在。

^{†5} 2021年11月現在。

^{†6} 映像コンテンツのオンラインデータベース。

^{†7} 2021年11月現在。

^{†8} 2019年5月現在。

^{†9} 2018年5月現在。

これだけ膨大なコンテンツがあれば、この中にはきっと自身にとって興味のあるコンテンツ、面白いコンテンツ、感動するコンテンツが隠されていることでしょう。そして、そのようなコンテンツに巡り合えたときには、大きな喜びが得られることでしょう。しかし、問題はどのようにしてそのコンテンツに巡り合うかです。人のもち時間は限られていますので、すべてのコンテンツに目を通したうえで、興味のあるコンテンツを見つけることは不可能です。膨大なコンテンツの中には、興味のあるコンテンツも含まれているかもしれませんが、その一方で、興味のないコンテンツ、自分には関連のないコンテンツのほうが圧倒的に多く含まれています。このように、圧倒的多数の関連のないコンテンツの中に、本当に興味のあるコンテンツが埋もれてしまいます。このような状況を情報過多 (information overload)、あるいは、より具体的には、コンテンツ過多 (content overload)^{12), 65)} とよびます。私たちは、まさにコンテンツ過多の時代に直面しているといえます。せっかくのよいコンテンツも、それに興味のあるユーザに届かなければ、もったいないことです。

ここで救世主となるのが、推薦システムです。人は、世の中のすべてのコンテンツについて把握しておくことは不可能ですが、推薦システムは、すべてのコンテンツの情報をデータベースとして記憶しています。また、ユーザと推薦システムとのやりとり (インタラクション) の中で、ユーザの行動履歴 (購買履歴や閲覧履歴、評価履歴など) を蓄積します。その行動履歴を基に、ユーザの興味を推定します。そして、ユーザに代わって、ユーザの興味に合うコンテンツの候補を推薦リストとしてユーザに提示してくれます。例えば、候補を10件程度に絞ってくれば、ユーザがこの中から自分の興味に合ったコンテンツを選ぶことは容易です。

推薦システムの意義を理解するうえで、もう一つ重要な概念がロングテール (long tail)^{152), 162)} です。図 1.1 は、Web 上のあるコンテンツのダウンロード数の分布を表したものです。横軸は人気ランキングを表し、上位ほど左側に位置します。縦軸は各順位のコンテンツのダウンロード数を表します。上位のコンテンツはダウンロード数も膨大ですが、下位にいくにつれてダウンロード数

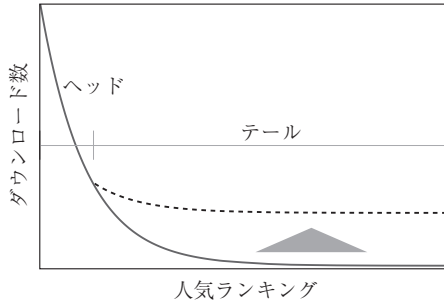


図 1.1 ロングテール

は減少していきます。しかし、完全に 0 になるということはなかなかなく、裾の長い分布になっています。これが長い尻尾のような形状をしていることから、このような分布はロングテールとよばれています。

ここで、上位の部分をヘッド、下位の部分をテールとよびます。ヘッドに位置するコンテンツはダウンロード数が膨大なことから、ヒットコンテンツといえます。一方で、テールに位置するコンテンツはダウンロード数は少なく、一部の特定のユーザにしか消費されないようなニッチコンテンツといえます。しかし、各コンテンツのダウンロード数はわずかであるとはいえ、テールに位置するコンテンツが全体の大部分を占めるため、軽視することはできません。むしろ、このテールにこそ隠れた需要が眠っているといえます。

そこで、推薦システムの出番です。ヒットコンテンツについては、メディアでもよく取り上げられたり、すでに人々に知られていたりすることもあることから、あえて推薦システムに頼るまでもないかもしれません[†]。しかし、ニッチコンテンツについてはもともと知名度が低く、ユーザ自身で興味のあるコンテンツを見つけ出すことは困難であるといえます。このようなときに、推薦システムにより、潜在的に興味をもつユーザに対して、そのようなニッチコンテンツを推薦することで、需要を引き出し、消費に結び付けられることが期待できます。つまり、推薦システムは図のように、テールを押し上げる役割を担います。

[†] もちろん、人気ベースの推薦も推薦システムの重要な戦略の一つではあります。

ユーザにとっては、自身では見つけられなかったような、興味のあるコンテンツに巡り合うことができるという点でメリットがありますし、サービス提供者側からみても、コンテンツに対する需要が増えることで収益の増大につながるというメリットがあります。さらに、コンテンツ制作者にとっても、自身のコンテンツを本当に興味のあるユーザに届けられるという喜びにつながります。

1.3 ユーザ, アイテム, 評価値

推薦システムを理解するうえで重要となるのが、ユーザとアイテムと評価値に関するデータです。本節では、それぞれについて数学的な表記も含めて定義しておきましょう。

1.3.1 ユーザ

推薦システムの利用者をユーザ (user) とよびます。表 1.1 には、例として Alice, Bruno, Chiara, Dhruv, Emi の 5 人のユーザを示しています。特に、推薦を受けようとしているユーザのことを対象ユーザ (target user) あるいはアクティブユーザ (active user) とよびます。例えば、Alice に対してクレープを推薦するという設定では、Alice が対象ユーザということになります。

推薦システムを利用しているすべてのユーザ集合を U と表すことにします。ここで、個々のユーザを識別するためにユーザ ID を用いることにします。特にユーザを一般化する場合には、ユーザ $u, v \in U$ のように記号で表すことに

表 1.1 ユーザに関するデータ

ユーザ ID	ユーザ名	年齢	性別
1	Alice	20	女
2	Bruno	22	男
3	Chiara	21	女
4	Dhruv	21	男
5	Emi	20	女

あ　と　が　き

本書では食をテーマにさまざまな推薦システムをみてきました。食を対象としただけでも、料理に対する好き嫌いだけでなく、栄養バランスを考慮したり、気分にあった料理を推薦するなど、さまざまな課題があります。また、その日の料理だけでなく、複数の料理を組み合わせた献立を推薦するという課題もあります。このような料理を対象とした推薦システムは、**料理推薦システム** (food recommender system) や**レシピ推薦システム** (recipe recommender system) ¹⁷⁴⁾ というトピックとして研究が進められています。

さて、Alice は Bruno と Chiara の 3 人グループで、さらに美味しいものを求め、ドライブがてら観光に出かけました。朝、どこかで朝食にしようとする近くの飲食店を探しました。Alice は焼肉のカルビが好きですが、朝からカルビを食べたいという気分ではありません。朝は、カフェでパンケーキが食べたいのかもしれない。このようにユーザの嗜好は時間帯などの状況に応じて変化します。このような推薦を受けようとしているときの状況をコンテキストとよび、コンテキストに応じて変化する嗜好をモデル化しようという問題が、**コンテキストアウェア推薦システム** (context-aware recommender system) ³⁾ のトピックで扱われています。

朝食も済ませ、目当ての観光スポットにも訪れました。しかし、思ったよりも時間に余裕ができましたので、どこかもう一か所巡りたいと思いました。すると、「このスポットを訪れた人は、こんなスポットにも訪れています」といくつかの近場の観光スポットが推薦されました。早速、立ち寄りたいたいと思い、カーナビに目的地をセットしました。すると、少々遠回りですが、Alice が好きな海沿いのルートが推薦されました。こうして景観を楽しみながら目的の観光スポットにたどりつきました。このように、観光旅行を支える推薦システムとし

て、観光情報推薦システム (tourism recommender system) やルート推薦システム (route recommender system) などの研究が取り組まれています¹⁸⁰⁾。特に、ユーザの現在地を基に推薦を提供しようとするシステムを位置情報ベース推薦システム (location-based recommender system)⁹⁴⁾ とよび、観光に限らずさまざまなドメインを対象に研究が進められています。

Alice たちはデザートにケーキを食べに来ました。ホールケーキを3人でシェアしようとしています。このとき、ショートケーキを推薦すると Alice と Chiara は喜ぶますが、Bruno はそれほどでもありません。一方で、Bruno の嗜好に合わせてブラウニーを推薦すると Bruno は大喜びですが、Alice と Chiara はブラウニーが好みかはわかりません。このとき、だれの嗜好を重視して推薦するのがよさそうでしょうか。このように、グループに対する推薦問題を対象としたトピックをグループ推薦システム (group recommender system)⁹⁹⁾ とよびます。結局、Bruno の好みに合わせてブラウニーをシェアすることにしました。結果、Alice と Chiara もブラウニーの美味しさに気付くことができ、大喜びでした。Alice と Chiara にとってはセレンディピティな体験ですね。

さて、観光旅行の最後にお土産屋を訪れました。しかし、お土産が多すぎてどれがよいのかよくわかりません。そこで、推薦システムに聞いてみました。

Alice 「学生 20 人でおすすめのお土産はありますか？」

推薦システム 「予算はどれくらいですか？」

Alice 「3000 円です」

推薦システム 「それでしたら、こちらのせんべいはいかがでしょう？ この土地ならではの素材の風味が楽しめます」

本書では知識ベース推薦システムについて学びましたが、それを発展させ、自然言語でのやりとりを目指そうというのが対話型推薦システム (conversational recommender system)⁴⁷⁾ です。さらに、最後に推薦システムが「この土地ならではの素材の風味が楽しめます」と推薦理由を提示することで、Alice にとって受け入れられやすい推薦になるかもしれません。どのような推薦理由を提示することがユーザの受け入れやすさ向上につながるか検証しようという取組み

が説明可能な推薦システム (explainable recommender system)^{143), 157)} のトピックで進められています。

お土産を買った Alice たちは帰路につきました。まだまだ旅の余韻に浸りたい気分です。すると、カーオーディオからはその気分合うような穏やかな楽曲を集めたプレイリストが流れました。このように楽曲を対象とした推薦システムを**楽曲推薦システム** (music recommender system)^{132), 164)} とよび、一つのトピックとして研究が進められています。楽曲推薦システムの中には、楽曲単位で推薦するものもあれば、プレイリスト単位で推薦するものもあります。また、ユーザの気分や感情に合った楽曲を推薦しようという試みもあります。

Alice は、今回の観光旅行を通して、その土地に興味をもちました。すると、「この土地に興味のある人におすすめの小説」として、その土地を舞台とした小説が推薦されました。観光と小説ではドメインが異なりますが、このようにあるドメインにおけるユーザプロフィールを基に、また別のドメインでの嗜好を予測しようという問題が**クロスドメイン推薦システム** (cross-domain recommender system)²⁹⁾ というトピックで扱われています。そして、その小説をきっかけに Alice はまた新たな土地に思いを馳せるのでした。

以上のように、推薦システムの研究には面白くワクワクするような課題がたくさん詰まっています。本書では、これら先端的なトピックについては扱いませんでしたが、いずれのトピックにおいても本書で学んだ内容が基礎となります。

世の中のコンテンツは数も種類も日々増大しており、人々の選択肢も無数に存在します。それにつれて人の嗜好もどんどん多種多様で複雑になってきます。そのような嗜好を果たして情報技術でどこまで予測できるのか、推薦システム研究はこの課題に挑戦しつづけていくことになるでしょう。さあ、推薦システムの探求への旅はいま始まったばかりです！

2022年5月

奥 健太

索引

【あ】	
アイテム	6
——アイテム類似度	
行列	112
アイテム因子	124, 162
アイテム因子行列	124, 161, 235
アイテム空間被覆率	276
アイテム制約	190
アイテムバイアス	109
アイテムベース協調	
フィルタリング	106, 126
アイテム変数	188
アイテム類似度	107
アクティブユーザ	5
アンケート	248
アンサンブル型ハイブリッド	221
暗黙のフィードバック	11
【い】	
いいね！ボタン	11
意外性	283
意外なアイテム	283
五つ星評価システム	10
【え】	
エコーチェンバ現象	274
エッジ	79
エポック	178
【お】	
オフライン処理	107
オフライン評価	240

オンライン処理	107
オンライン評価	251
【か】	
過学習	88, 166, 181, 242
学習	52, 131
学習モデル	52, 131
学習率	167
確信度	60, 136
確率	66, 141
確率的勾配降下法	173, 234
確率変数	66, 141
加重ハイブリッド	222
仮説	238
過大特殊化	183, 274
カタログ被覆率	275
カテゴリカル属性	7, 188
観測値	13
緩和	196
【き】	
偽陰性	260
機械学習	48
帰結部	56, 68, 135, 191, 216
擬似評価値	228
擬似評価値行列	228
既知・不既知の評価値	280
教師信号	49
偽陽性	260
偽陽性率	261
協調フィルタリング	92, 132, 161
協調ベース推薦システム	92, 181, 218

共分散	37, 121
行列因子分解	161, 234
行列補完問題	93, 133
寄与度	195
距離	29, 282
寄与率	45
距離に基づく新規性	281
近傍アイテム	31, 113
近傍ベース協調フィルタリング	94, 179
近傍ベース方式	18
近傍ユーザ	101
【く】	
クラス	54
クラス分類	54
クラス分類器	54
クリック率	252
訓練データ	49, 131, 153, 242
【け】	
欠損値	13, 93, 116, 132, 226
欠損評価値	13
決定木	79, 153
研究室実験	247
検証データ	242
【こ】	
交差検証法	245
構造化データ	7
勾配	167
勾配降下法	167, 234
効用	194

効用関数	207
誤差	254
コサイン類似度	22, 108
個人化	218
固有値	40, 122
固有ベクトル	40, 122
コールドスタート問題	
182, 185, 218, 226	
混合型ハイブリッド	236
コンテンツ過多	3
混同行列	260
コンバージョン率	252
【さ】	
再現率	263
最小確信度	60, 136, 216
最小支持度	58, 136, 216
最適化問題	166
残差	165
残差行列	165
【し】	
次元削減	35, 117, 154, 161
次元の呪い	35
指向的批評	211
嗜好予測問題	14, 254
自己重み	232
事後確率	69
指示関数	205
支持度	58, 136, 217
事象	66
二乗平均平方根誤差	257
事前確率	69, 145
実験参加者	247
実験参加者間計画	249
実験参加者内計画	249
実ユーザ	251
ジニ係数	81, 156
自明な推薦	127, 183, 273
修正案	196
収束	171
従属デフォルト	200
従属変数	52
重要業績評価指標	252

重要度重み	230
主成分得点	43, 123
主成分分析	36, 118
出現頻度	58, 137, 217
順序型評価値	10
順序効果	250
順序付け関数	14, 25
上位 K 推薦問題	15, 254
条件付き確率	68, 141
条件付き独立	71
条件部	
56, 68, 135, 191, 216	
情報過多	3
事例 (case)	202
事例 (instance)	49
事例ベース推薦システム	
201	
真陰性	260
新規アイテム	279
新規アイテム問題	182
新規性	279
新規ユーザ問題	182
真陽性	260
真陽性率	261
【す】	
推薦	14
推薦サイクル	16, 186, 201
推薦システム	1
推薦タスク	192
推薦知識ベース	192
推薦問題	14
推薦リスト	14
スイッチングハイブリッド	
225	
数値型評価値	10, 93
数値属性	7, 18, 187
スコア	14
スコア関数	
14, 24, 33, 63, 78, 89	
ステップ幅	167
スムージング	76, 148
スムージングパラメタ	
77, 149	

【せ】

正解データ	241, 254
正確性	127, 179, 183, 254
正規化	78, 109
正規化 MAE	259
正規化 RMSE	259
正クラス	54
正事例	54
正則化	166
正則化項	166, 235
正則化パラメタ	166
静的デフォルト	199
正の相関	39, 100
制約充足問題	192
制約ベース推薦システム	
186	
説明変数	52, 132, 153
セレンディピティ	
127, 183, 286	
—なアイテム	286
ゼロ頻度問題	76, 148
選言標準形	190
潜在因子	124, 155
潜在因子行列	124, 164, 235
潜在因子モデル	160, 234
選択クエリ	193

【そ】

相関行列	40
相関ルール	61, 135, 216
相関ルールマイニング	
61, 216	
疎行列	116, 226
属性	7
疎性	116
疎性問題	117, 161
損失関数	166

【た】

対象アイテム	6, 51
対象ユーザ	5
第 1 主成分	41, 125
第 2 主成分	42, 125

多クラス分類	54
多次元属性	8
多属性効用理論	194
多様性	276
単一批評	211
単純な	71
単純ベイズ仮定	71, 144
単純ベイズ協調フィルタ	
リング	140
単純ベイズ分類器	64, 140
単値型評価値	11
【ち】	
置換の批評	211
知識ベース推薦システム	185, 218
調整コサイン類似度	108
調和平均重み	230
直列型ハイブリッド	225
【て】	
定義域	187
適合アイテム	264
適合性制約	190
適合性フィードバック	26
適合率	263
—再現率曲線	265
テストデータ	242
データセット	240
デフォルト値	199
デモグラフィック属性	6
転置	8
【と】	
導出デフォルト	200
動的批評	215
特徴拡張ハイブリッド	227
特徴空間	19
特徴結合ハイブリッド	233
特徴ベクトル	8, 18, 48, 194
特微量	7
独立	70
独立変数	52

【な】	
内部ノード	79
内容因子行列	235
内容行列	234
内容ブーステッド協調	
フィルタリング	226
内容ベース推薦システム	18, 48, 181, 218
【に】	
二値型評価値	10, 133
二値属性	7, 48
人気度	280
—に基づく新規性	280
【ね】	
根ノード	80
【の】	
ノード	79
【は】	
ハイパーパラメタ	239
ハイブリッド型推薦	
システム	221
ハイブリッド相関重み	230
パーソナリティ	247
発火する	62, 139, 193
発火ルール	62, 139
発見性	274
葉ノード	79

【に】	
二値型評価値	10, 133
二値属性	7, 48
人気度	280
—に基づく新規性	280
【ね】	
根ノード	80
【の】	
ノード	79
【は】	
ハイパーパラメタ	239
ハイブリッド型推薦	
システム	221
ハイブリッド相関重み	230
パーソナリティ	247
発火する	62, 139, 193
発火ルール	62, 139
発見性	274
葉ノード	79

【は】	
ハイパーパラメタ	239
ハイブリッド型推薦	
システム	221
ハイブリッド相関重み	230
パーソナリティ	247
発火する	62, 139, 193
発火ルール	62, 139
発見性	274
葉ノード	79

【ひ】	
ピアグループ	101
ピアソンの相関係数	40, 97
非構造化データ	7
一つ抜き交差検証法	245
批評	210
批評パターン	215
被覆率	274
評価	238
評価指標	239, 253

評価値	9, 93, 133, 241, 254
評価値行列	13, 92, 116, 131, 161, 222, 241
評価目的	238
評価履歴	9, 131, 241
標準偏差	37
類出ルール	58

【ふ】	
フィルタ制約	191
フィルタバブル	274
不既知アイテム	279
複合批評	214
負クラス	54
不純度	81
負事例	54
不適合アイテム	264
負の相関	39, 100
プリミティブシステム	284
フロベニウスノルム	165
分散	37, 120
分散共分散行列	40, 119

【へ】	
平均逆順位	267
平均二乗誤差	257
平均絶対誤差	256
平均中心化評価値	98, 109, 229
平均中心化評価値行列	99, 109
平均適合率	269
ベイズの定理	69, 143
並列型ハイブリッド	222
ベクトル	8
ベースライン	239
変数	166, 187

【ほ】	
忘却アイテム	279
ホールドアウト法	243

【み】	
未観測値	13
未評価アイテム	279
【む】	
矛盾	196
無相関	39, 100
【め】	
明示的フィードバック	11
メモリベース協調フィルタ	
リング	94
【も】	
目的関数	166, 235
目的変数	52, 132, 153
モデルベース協調フィルタ	
リング	130, 160, 179
モデルベース診断	196
モデルベース方式	48
モノリシック型ハイブリッド	233
【ゆ】	
ユークリッド距離	30
ユーザ	5
— アイテム評価値行列	13

ユーザ因子	124, 162
ユーザ因子行列	
	124, 161, 235
ユーザ空間被覆率	276
ユーザ嗜好	9
ユーザバイアス	109
ユーザ評価	247
ユーザプロファイル	19, 52
ユーザベース協調フィルタ	
リング	94, 126
ユーザ変数	187
ユーザ要求	188
ユーザ類似度	96, 229
【よ】	
予測	53, 131
予測対象データ	
	51, 131, 153
予測被覆率	274
予測評価値	
	14, 51, 102, 114, 163, 223, 232
【ら】	
ラプラススムージング	
	76, 148
ラベル付きデータ	50
ラベルなしデータ	50

【り】	
リスト内類似度	277
リッカート尺度	248
利用ログ	247
【る】	
類似アイテム	112
類似性尺度	202
類似度	
	21, 96, 107, 203, 277
類似度関数	
	97, 108, 203, 277
類似ユーザ	101, 232
累積寄与率	45
ルール	56, 135, 190
ルールベース協調フィルタ	
リング	133
ルールベース分類器	
	54, 133
【れ】	
レベル	80
連言標準形	188
連続尺度	10
【ろ】	
ロングテール	3

【A】	
AUC	263
A/B テスト	251
【B】	
Beyond Accuracy	274
【D】	
d 次元特徴ベクトル	8
DCG	270

【F】	
F 値	265
【I】	
IDCG	272
【K】	
k 近傍法	28
k 分割交差検証法	245
【M】	
MAP	269

【N】	
n 段階尺度	10
nDCG	272
【R】	
ROC 曲線	261
Rocchio の式	26
【数字】	
2 クラス分類	54
5 段階リッカート尺度	248
7 段階リッカート尺度	248

— 著者略歴 —

- 2004年 大阪市立大学工学部土木工学科卒業
2006年 奈良先端科学技術大学院大学情報科学研究科博士前期課程修了
(情報生命科学専攻)
2009年 奈良先端科学技術大学院大学情報科学研究科博士後期課程修了
(情報生命科学専攻), 博士(工学)
2009年 立命館大学助教
2014年 立命館大学講師
2016年 龍谷大学講師
現在に至る

基礎から学ぶ推薦システム

— 情報技術で嗜好を予測する —

Fundamentals of Recommender Systems

— Predicting Preference Using Information Technology —

© Kenta Oku 2022

2022年7月21日 初版第1刷発行



検印省略

著者 奥 健 太
発行者 株式会社 コロナ社
代表者 牛来 真也
印刷所 三美印刷株式会社
製本所 有限会社 愛千製本所

112-0011 東京都文京区千石 4-46-10

発行所 株式会社 コロナ社
CORONA PUBLISHING CO., LTD.

Tokyo Japan

振替 00140-8-14844 · 電話 (03) 3941-3131(代)

ホームページ <https://www.coronasha.co.jp>

ISBN 978-4-339-02928-4 C3055 Printed in Japan

(森)



< 出版者著作権管理機構 委託出版物 >

本書の無断複製は著作権法上での例外を除き禁じられています。複製される場合は、そのつど事前に、出版者著作権管理機構 (電話 03-5244-5088, FAX 03-5244-5089, e-mail: info@jcopy.or.jp) の許諾を得てください。

本書のコピー、スキャン、デジタル化等の無断複製・転載は著作権法上での例外を除き禁じられています。購入者以外の第三者による本書の電子データ化及び電子書籍化は、いかなる場合も認めていません。落丁・乱丁はお取替えいたします。