

訳者まえがき

本書は“Statistical Methods for Recommender Systems”を翻訳したものであり、著者の大規模システムでの豊富な経験をもとに推薦システムにおける複雑な概念を具体的に解説した書籍である。著者である Deepak K. Agarwal は推薦システムやオンライン広告における研究開発に長年携わっており、Bee-Chung Chen はリーディングテクノロジストとして Yahoo! や LinkedIn で推薦システムの開発に携わってきた。

長年推薦システムの開発に携わってきた著者のノウハウが記述されており、理論だけでなく実践で役立つことが意識された内容となっている。本書の構成としては、第 I 部では推薦システムの古典的課題、課題解決のための主要なアイデアと背景知識、協調フィルタリングなどの古典的手法の紹介がされている。第 II 部ではまず推薦システムにおける一般的な課題設定、課題解決のためのシステム構成の解説がされ、徐々に洗練された手法の紹介が行われている。モデルの効率的なアップデートやコールドスタート問題などの現実的に直面する問題の対応についても述べられている。第 III 部では著者が独自に開発したアルゴリズムも含めた応用的手法の紹介がされている。古典的手法から応用的／実践的手法が順を追って解説されており、各手法の限界、理論的な仮定の緩和方法、手法の拡張方法を段階的に学ぶことができる。いくつか特筆すべき章をあげておく。第 4 章ではモデルの評価方法について詳細に述べられており、正しく機能するモデルを選択するために読んでおきたい。第 11 章では多目的最

適化について述べられている。他書では言及が少ないが、現実の世界では複数の利害関係の中から最適な解を選択することが必要である。その点、第 11 章は現実世界の問題を解くための足がかりとなる内容となっている。

推薦システムの理論を中心に解説した書籍はあるが、実アプリケーションに適用するためのノウハウおよびそのノウハウを数理的に解説している書籍は少ない。本書は理論と実践の橋渡しとなる書籍であり、推薦システムの構築を検討しているエンジニアが現実的な課題に対峙するための知識を得るのに本書がお役に立てれば幸いである。

2018年2月

訳者

まえがき

本書について

推薦システムは様々な文脈において自動でユーザに適したアイテムを選択するコンピュータプログラムである。このようなシステムは広く使われ、生活に不可欠なものとなっている。実例としては Amazon などにおけるユーザへの商品推薦、Yahoo! のようなウェブサイトを訪問するユーザへのコンテンツ推薦、Netflix における映画の推薦、LinkedIn のようなサイトでのユーザへの職業の推薦などがある。アルゴリズムは過去のユーザとアイテム間の相互作用から得られた大量の高頻度データをもとに構築される。このアルゴリズムは統計的なものであり、構築するにあたり逐次的な意思決定、高次元カテゴリデータでの相互作用のモデル化、スケーラブルな統計的手法の開発、などの分野における困難に挑戦することになる。この領域における新たな方法論はコンピュータ科学、機械学習、統計学、最適化、コンピュータシステム、そして当然ながら各ドメインにおける専門家との緊密な連携を必要とする。推薦システムはビッグデータの応用先として最も刺激的なものの1つである。

なぜ本書を書いたか

様々な領域で推薦システムに関する多くの書籍が著されてきた。これらはコンピュータ科学、機械学習、そして統計学などであり、問題の特定の側面に着目している。しかしあらゆる統計的な論点の包括的な取り扱いとそれらの相互

の関係性に関する記述を欠いている。Yahoo! と LinkedIn の実システムにおける経験を通じてこの事実を認識した。例えば統計学と機械学習は学習データにないサンプルの予測誤差を最小化するモデルを構築することに注力する。しかしながら実際に重要な問題の全ての側面には対処していない。統計学の観点では推薦システムは高次元の逐次過程であり、洗練された統計モデルを開発することと同等に、実験計画のような論点を研究することが重要である。実際、これら2つは強く関係している。効果的な実験計画は次元の呪いを柔らげるためのモデルを必要とする。また既存研究の大半は映画のレーティング、購買数やクリック率などの1つの量の応答モデルを構築しがちである。Facebook, LinkedIn や Twitter などのソーシャルメディアの出現により、複数の応答を考慮ができるようになった。例えばニュース推薦において、クリック率、シェアする率、ツイートする率を同時にモデル化したいこともあるだろう。このような複数の応答のモデル化は困難である。最終的にそのような複数の応答の予測が可能である場合、推薦を行うための利得関数をどのように構築したらよいのだろうか？ クリック率と比較してシェアする率を最適化することはどれだけ重要だろうか？ この種の質問に対する答えは有効なパラメータを導き出すことができるドメインの専門家との緊密な連携を行ったうえで、多目的最適化を使い得られる。

本書の目的は推薦システムの文脈で生じるこのような問題を包括的に議論することである。これに加えて適応的な逐次計画（多腕バンディット）、双線形ランダム効果モデル（行列分解）、分散計算環境を用いたスケーラブルなモデルの当てはめなど現在の最先端の統計手法を深く詳細に示す。本書を執筆した目的は産業の現場でのこのような大規模システムに関する広い経験を記し、統計学、機械学習、コンピュータ科学のコミュニティーにこの論点に対する関心を引き起こすことである。これは様々な点で有用であると信じる。特にウェブへの適用などにおける高次元で大規模データの研究の進展に役立つ。アカデミックでこのような研究を進めるうえで大量データに対し実行可能なソフトウェアが利用可能であることが必要であることは理解している。これを促進するためにオープンソースソフトウェアを補足として提供する (<https://github.com/beechnung/Latent-Factor-Models>)。)

また本書は理論と実践の溝を埋めるものとなることを信じている。プロジェクトの担当者には統計的な論点に関する理解の助けとなり、モデル担当者にはより複雑で実践的な問題に対し生じる統計学に関する争点について深い理解を与える。

本書の構成

本書は3部で構成される。

第I部では、推薦システムの問題設定、困難な点、その問題に立ち向かうにあたって主要なアイデアと必要となる背景知識を説明する。第2章ではこれまで推薦システムに使用されてきた古典的な手法の概要を示す。このような手法はユーザとアイテムを素性ベクトルという形で特徴付け、類似度関数、標準的な教師あり学習、または協調フィルタリングにもとづき（ユーザ、アイテム）対にスコアを付与する。これらの古典的な手法では推薦問題における探索-利用トレードオフは無視されることが多い。それゆえ第3章ではこの問題の重要性を示し、後の章で示す解決策の主要なアイデアを紹介する。技術的な点を掘り下げる前に、第4章では様々なアルゴリズムの性能評価手法を概観する。

第II部では一般的な問題設定における詳細な解決法を示す。第5章では様々な問題設定と、システム構成の例を示す。続く3つの章では3つの一般的な問題設定を説明する。第6章では特に探索-利用の側面に着目し、最も人気なものを推薦する問題を示す。第7章では素性ベースの個別化推薦を取り扱う。ここではいかにして最新のユーザ-アイテム相互作用のデータを利用して逐時的なモデル更新を行い、素早く良い解に収束するかに重点を置く。第8章は第7章で導入した素性ベース回帰を因子モデル（行列分解）に拡張し、同時に因子モデルを用いたコールドスタートへの自然な対処法を示す。

第III部では3つの高度なトピックを示す。第9章では潜在ディリクレ分配トピックモデルを使用する改良された行列分解モデルを通じてアイテムのトピックとトピックに対するユーザの親和性を同時に明らかにする因子分解モデルを示す。第10章では推薦されるアイテムがユーザに対して高い親和性をもつだけでなく、コンテキストに対しても適切であること（例えばユーザが現在読んでいるニュース記事に関連したアイテムの推薦）を必要とするコンテキスト依

存の推薦問題を対象にする。第 11 章では他の目的の損失を抑えながら（例えばクリック数の損失を 5% 以下にする）単一の目的（例えば収入）を最大化する、制約付き最適化にもとづいた複数目的最適化のための第一原理的なフレームワークを議論する。

制約事項

あらゆる書籍と同様に、本書の内容にも限界がある。大規模学習に使われる Spark などの現代のコンピュータパラダイムの深い所は対象としていない。伝統的な実験計画の手法では、ソーシャルネットワークをなすユーザに対してモデルのオンライン評価を適切に行うことはできない。ソーシャルグラフによる影響を補正する新技術が必要である。そのような進んだ話題は本書では扱わない。本書を通じて、回帰を主要な手段として応答を予測するというアプローチで推薦の問題に取り組む。この主な理由はモデルの出力する応答の予測は容易にそれに続くプログラムと組み合わせることができるからである。ランキング損失の直接最適化をもとにした手法について包括的には扱わない。2つのアプローチを比較することはまた議論に値するトピックであろう。

謝辞

Raghu Ramakrishnan, Liang Zhang, Xuanhui Wang, Pradheep Elango, Bo Long, Bo Pang, Rajiv Khanna, Nitin Motgi, Seung-Taek Park, Scott Roy, Joe Zachariah には、多くの洞察に満ちた議論と協働に対して感謝を送る。また Yahoo! と LinkedIn の同僚にも励ましと協力に対し感謝を送る。それらがいなければ多くのアイデアが日の目を見なかったであろう。