

早稲田大学大学院情報生産システム研究科

博士論文審査結果報告書

論 文 題 目

**Study on Personalized Recommendations Based on Behavioral
Sequence Modeling Using Deep Neural Networks**

申 請 者

Hangyu DENG

情報生産システム工学専攻
ニューロコンピューティング研究

2024 年 5 月

個別化推薦システム (Personalized Recommendation System) は、ユーザーのエンゲージメントと満足度を向上させる上で重要な役割を担っており、電子商取引やコンテンツ配信など様々な領域に影響を与えており、大きく注目されている。個別化推薦システムは、膨大な数のアイテムの中からユーザーが興味を持ちそうなアイテムを提供することを目的としている。最近の研究では、ユーザーとアイテムとのインタラクションを時間順に行動シーケンスとして整理することで推薦システムにとって有用な情報を抽出できることが示されている。すなわち、正確でタイムリーな推薦を行うためには、ユーザーの過去の行動シーケンスデータから得られる連続的なパターンを利用することが重要である。しかし、ユーザー行動シーケンスのモデル化には様々な課題がある。ユーザー行動シーケンスにはユーザーの複数の意図が含まれる可能性があるだけでなく、言語テキストなどの他のシーケンスのような厳密な構造的制約がないため、大きな不確実性が含まれる。さらに、ユーザー行動シーケンスデータは通常スパースであり、オーバーフィッティングの問題が発生しやすい。そのため、深層ニューラルネットワークによるユーザー行動シーケンスのモデリングは困難である。

そこで本研究では、深層学習法を用いて、単一の行動シーケンスのシーケンシャルな特性を利用するだけでなく、ユーザー行動シーケンスの本質的な構造も掘り下げて、サブシーケンス情報を抽出する。具体的には、第一に、シーケンスをアイテムペアに分解し、アイテム間の関係を抽出することでサブシーケンスを再構築する方法を、第二に、自己注意機構 (Self-Attention Mechanism) に基づいてサブシーケンス抽出をモジュール化する方法を提案する。そして第三に、他のユーザーの類似の行動シーケンスからのグローバルなアイテム遷移を利用する方法を、第四に、ユーザー行動のシーケンシャルな可予測性を識別・利用をする方法を提案する。このようにして、サブシーケンス情報を抽出し、サブシーケンスモデリングを用いて、高性能な個別化推薦システムを構築している。

以下に、本論文の構成と各章の概略について述べ、評価を与える。

第 1 章では、個別化推薦システムおよびその関連研究、特にユーザー行動シーケンスのモデリング手法を紹介し、本研究の動機・目的を明らかにしている。

第 2 章では、アイテムペアへのシーケンス分解とサブシーケンスの再構築に基づく、ユーザー行動シーケンスのモデリング手法の提案について述べている。提案法はまず、多層パーセプトロンに基づく関係抽出モジュールを構築し、ユーザー行動シーケンスを複数のアイテムペアに分解し、アイテム間の異なるタイプの関係を抽出する。次にこれらのアイテムペアをグループ化し、それらの時間順序情報に基づいてサブシーケンスを再構築する。これに基づいて、ユーザー行動シーケンスのモデリングを行い、個別化推薦システムを構築している。性能評価の数値例では、提案法を Amazon Electronics, Amazon Books, Taobao という 3 つのクリックスルー率のベンチマークデータセットに適用して、複数の最先端の従来法と比較した。LogLoss (Logarithmic Loss) と AUC (Area

Under the Curve) による評価では、すべてのデータセットにおいて提案法が従来法より優れていた。例えば、Amazon Books データセットに対して、従来の最良法である DIEN (G. Zhou et al., 2019) の AUC 評価は 0.8559 であるのに対して、提案法では 0.9001 であり、5.16% 向上した。これらのことから、提案法の有効性が明らかとなっている。

第 3 章では、自己注意機構を用いてサブシーケンス抽出のモジュール化に基づく行動シーケンスのモデリング手法について述べている。自己注意機構に基づいたトランスフォーマーは、次のアイテムがこれまでのアイテムの順序に依存する言語モデリングパラダイムにおいて、大きな成功を収めている。本論文では、自己注意機構を利用してサブシーケンス抽出モジュール化を行う。サブシーケンス抽出モジュールでは、自己注意ネットワークでアイテムのセマンティック類似性を考慮して、アイテムを異なるサブシーケンスに分配する。このプロセスは、アイテムを異なるサブシーケンスに効果的に分離し、関連するアイテムをグループ化することでサブシーケンス情報を抽出する。これにより、高性能な個別化推薦システムを構築することができる。性能評価の数値例では、提案法を Amazon Beauty, Amazon Games (W.C. Kang et al., 2018), Gowalla (E. Cho et al., 2011), ML-1M (F.M Harper et al., 2015) という 4 つのベンチマークデータセットに適用して、複数の最先端の従来法と比較した。HIT@k (トップ k ヒット率) では、すべてのデータセットにおいて提案法が従来法より優れていた。例えば、ML-1M データセットに対して、従来の最良法である LightSANS (X. Fan et al., 2021) の HIT@10 は、0.3023 であるのに対して、提案法では 0.3159 であり、4.50% 向上した。これらのことから、提案法の有効性が明らかとなっている。

第 4 章では、グラフニューラルネットワーク (Graph Neural Network: GNN) を用いたローカル情報とグローバル情報を利用した、行動シーケンスのモデリング手法について述べている。一般的なシーケンスモデルは、対象ユーザーの行動シーケンスのアイテムのみに着目するが、本論文では、ユーザーの潜在的な意図を明らかにするため、対象ユーザーの行動シーケンスというローカル情報だけでなく、他のユーザーの類似の行動シーケンスというグローバル情報も取り入れる。まず、学習セット内のすべての行動シーケンスを統合し有向グラフとして表現し、ユーザーが示す集合的な行動パターンを捉えて、アイテム間の重要な関係の識別を容易にする。次に、GNN をこの有向グラフに適用してサブシーケンスの抽出を行う。GNN のモジュールでは、GNN の学習によって、たくさんのユーザーそれぞれに対して、それぞれの潜在的な意図が「カプセル化」されて抽出される。これによって、サブシーケンスモデリングを行い、個別化推薦システムを構築している。性能評価の数値例では、提案法を前述の 4 つのベンチマークデータセットに適用して、複数の最先端の従来法と比較した。HIT@k による評価では、すべてのデータセットにおいて提案法が従来法より優れていた。例えば、ML-1M データセットに対して、従来の最良法である GC-SAN (C. Xu et al., 2019) の HIT@10 は、0.3044 であるのに対して、提案法で

は 0.3108 であり、2.10%向上した。これらのことから、提案法の有効性が明らかとなっている。

第 5 章では、アイテムの可予測性を利用したサブシーケンスのモデリング手法の提案について述べている。シーケンスにおける次のアイテムの予測のしやすさから、シーケンスの連続性に関する様々な有用な情報を得ることができる。例えば、予測しやすいアイテムは先行するサブシーケンスの継続を示唆し、予測しにくいアイテムは一般に他のサブシーケンスへの遷移を示唆する。この可予測性情報を有効に活用することで、サブシーケンスの検出・利用を改善することができる。提案法は二つの段階から構成される。初期段階では、高速で効率的なシーケンスモデルを用いてトレーニングセットの各アイテムの可予測性情報を検出する。後段では、この可予測性情報は、異なるサブシーケンス間の境界を識別するためと、関連性のないアイテム遷移を記憶しないように、予測しにくいアイテムの損失の重みを減らすために利用される。さらに、提案法を第 3 章で述べたサブシーケンス検出・モデリング手法に併せて用いることにより、より高性能な個別化推薦システムを構築している。性能評価の数値例では、提案法を前述の 4 つのベンチマークデータセットに適用して、複数の最先端の従来法と比較した。HIT@k による評価では、すべてのデータセットにおいて提案法が従来法より優れていた。例えば、ML-1M データセットに対して、従来の最良法である LightSANS (X. Fan et al., 2021) の HIT@10 は、0.3021 であるのに対して、提案法では 0.3199 であり、5.89%向上した。これらのことから、提案法の有効性が明らかとなっている。

第 6 章では、結論として、本研究の成果について総括し、今後の研究課題を論じている。

以上を要約すると、本論文では、1)多層パーセプトロンを用いてアイテム間の関係を探索し、2)自己注意機構を用いて潜在的なサブシーケンスの抽出をモジュール化し、3)GNN を用いてグローバルなアイテム遷移を利用し、4)ユーザー行動の可予測性の検出・活用をすることで、不確実性を伴うユーザー行動シーケンスに対して、サブシーケンス情報を抽出し、サブシーケンスモデリングを行い、個別化推薦システムを構築する方法を提案している。また、提案法を複数のベンチマークデータセットに適用しその有効性を明らかにしている。これらの成果は深層学習、ニューロコンピューティング分野の発展に寄与するところ大である。よって、本論文は、博士（工学）の学位論文として価値あるものと認める。

2024 年 4 月 6 日

審査員

主査	早稲田大学	教授	博士(情報工学)(九州工業大学)	古月 敬之
副査	早稲田大学	教授	工学博士(早稲田大学)	吉江 修
	早稲田大学	教授	博士(工学)(早稲田大学)	藤村 茂