

Lifelong User Action Sequence Modeling

Author: 西村直樹

背景

- 最近の推薦/検索システムでは、超長期にわたるユーザー行動履歴の活用がトレンドとなっている
 - e.g, [Instagram](#), [LinkedIn](#), [Pinterest](#), [Zalando](#)
- Indeedでは、超長期のユーザー行動履歴を埋め込み（Embedding）化した特徴量を下流の検索システムで利用することで、再ランキング（Re-ranking）性能が向上した事例がでてきた
- リクルートでも、同様の取り組みができないか検討したい

従来のユーザ行動履歴の活用法

- 集約: 過去7日、14日、90日以内の各属性のクリック数、応募数などを集計
- One-hot encoding

従来手法の課題

- 集約によって、ユーザー行動の順序情報が失われる
- 各属性を個別に集約するため、ジョブと属性の関連情報が失われる
- すべてのアクションを同じ重みで扱うため、ノイズが混入しやすくなる
- 属性の重複排除により、強調すべき属性の重要性が薄れてしまうことがある
- 「N日以内」といった古いデータの削除により、長期的な履歴情報が失われる

取り組みのゴール

- 現状の7日、14日、90日以内などのユーザー行動履歴の単純な集約やOne-hot encodingにとどまらず、すべてのユーザー行動履歴を活用し、行動ノイズを削減しつつ重要な行動が強調された埋め込みを作成することを目指す
- 作成した埋め込みを推薦モデルの入力とした際に、オフラインでの性能が改善することを確認する