

修士論文概要書

Summary of Master's Thesis

Date of submission: 02/29/2020

専攻名(専門分野) Department	電気・情報生命	氏名 Name	渡邊 遼真	指導教員 Advisor	村田 昇
研究指導名 Research guidance	情報学習システム	学籍番号 Student ID number	5324E110-4		
研究題目 Title	マルチモーダルな GNN における説明可能な影響力予測				

研究背景

近年、スマートフォンやソーシャル・ネットワーキング・サービス(SNS)の普及により、デジタルマーケティング市場は急速に拡大しており、2029年には国内市場だけで約2兆円規模に達すると予測されています。特にインフルエンサーマーケティングは企業のブランディングにおいて不可欠な要素となっていますが、すでに著名なトップインフルエンサーへの依頼はコストが高騰しており、競合他社との競争も激化しているという課題があります。

こうした背景から、現在はフォロワー数が少なくても将来的に人気が出る有望な人材を早期に発掘・契約する「青田買い」の手法が注目されています。これにより、企業は低コストで高い利益率を確保し、長期的なパートナーシップを築くことが可能になります。しかし、既存の先行研究におけるインフルエンサー予測モデルの多くは、予測プロセスがブラックボックス化しており、「なぜその人物が伸びると予測されたのか」という根拠が不明瞭でした。マーケティングの実務現場では、単なる予測スコアだけでなく、その予測に至った要因(説明可能性)が意思決定において極めて重要なため、高精度な予測と説明可能性の両立が求められています。

問題設定

本研究では、画像・テキスト・数値・グラフ構造といった「マルチモーダルデータ」を用いて、将来のエンゲージメント(影響力)を予測し、その成長要因を特定するタスクを設定しています。

対象データ

インスタグラムの投稿データ(画像、キャプション、タグ)やユーザー間の関係性(メンション、コメント)を含む異種混合グラフを使用します。このデータセットには完全なフォロー関係が含まれていないため、メンションやコメントの共起関係を用いて擬似的なネットワークを構築し、関係性的の代替指標としています。

タスクの定式化

過去 T 期間におけるグラフデータ G_t (ユーザーや投稿などのノード集合とエッジ集合) および特徴行列 X_t を入力とし、次期 $T+1$ における各ユーザーのエンゲージメントスコア \hat{y}_u を予測します。

最適化手法

本研究の目的は具体的な数値を当てることではなく

く、有望なユーザーを上位に抽出することであるため、回帰問題ではなく「ランキング学習(Learning to Rank)」として定式化しています。損失関数には、モデルの予測順位と正解順位の不一致に対してペナルティを与える関数を採用し、順位の整合性を学習させています。

提案手法

本研究では、時系列グラフニューラルネットワークによる高精度な予測と、公理的な特徴量帰属手法による説明性の付与を統合した以下のフレームワークを提案しています。

1. 予測モデル: Influencer Rank

- GCN (Graph Convolutional Networks): ユーザー、投稿、タグなどが混在する異種混合グラフに対し、隣接ノードからの情報を集約することで、単独の属性だけでなくネットワーク構造全体を考慮した特徴表現を獲得します。
- GRU (Gated Recurrent Unit) + Attention 機構: GCN で抽出された空間的特徴を時系列順に GRU に入力し、時間的な変化を捉えます。さらに Attention 機構を導入することで、全ての時点を均等に扱うのではなく、バズが発生した月など「予測にとって重要な時点」を自動的に強調して重み付けを行い、最終的なユーザー表現を生成します。

2. 説明可能性技術: Integrated Gradients (IG)

- 選定理由: 予測モデルのような非線形モデルにおいて、「Sensitivity(感度)」と「Implementation Invariance(実装不变性)」という2つの重要な公理を満たす信頼性の高い手法であるため採用されました。
- 算出方法: 情報量がゼロのベースライン(全ての特徴量が0の状態など)から実際の入力データまで特徴量を連続的に変化させ、その経路上の勾配を積分することで、各特徴量が予測結果にどの程度貢献したかを定量的に算出します。これにより、投稿頻度や画像の質などが成長にどう寄与したかを個別に評価可能になります。

応用例

Instagram のデータセット（2017 年 1 月～11 月を学習用、12 月をテスト用）を用いた評価実験を行い、提案手法の有効性を定性的・定量的な観点から検証しました。

実験 1：予測根拠の可視化（定性評価）

成長上位と予測されたユーザーに対して IG を適用し、成長要因を分析しました。

結果：ユーザーによって成長のドライバーが異なることが明らかになりました。例えば、ユーザー `london_thelug` や `mariahhydzik` においては「投稿間隔 (post_interval)」が正の貢献度を強く示しており、定期的な投稿が成長要因であると判断されました。一方、ユーザー `pdy` では「キャプションの長さ (caption_length)」や「過去の投稿履歴」が重要視されており、コンテンツの内容量が評価されていることが分かりました。これにより、単一の指標ではなく、ユーザーごとの戦略（頻度重視か、内容重視か）に基づいた分析が可能であることが示されました。

実験 2：予測精度の検証（定量評価）

評価指標：検索や推薦で重視される NDCG (Normalized Discounted Cumulative Gain) および、ユーザーの閲覧行動モデルに基づく RBP (Rank-biased Precision) を用いました。

結果：提案手法 (Influencer Rank) は、ベースライン手法と比較して全ての指標で高い精度を記録しました。具体的には、NDCG@50 において提案手法が 0.852 であるのに対し、ベースラインは 0.721、ランダム予測は 0.105 であり、将来有望なインフルエンサーを効果的に上位にランキングできていることが実証されました。

まとめ

本論文では、インフルエンサーマーケティングにおける「青田買い」を支援するための、説明可能な予測モデルの構築と検証を行いました。

- 研究の貢献：マルチモーダルな深層学習モデルを用いることで高い予測精度を実現しつつ、XAI 技術 (Integrated Gradients) を導入することで、これまでブラックボックスであった予測の根拠（投稿頻度、画像の輝度、テキスト長などの貢献度）を可視化することに成功しました。
- 実務的意義：これにより、マーケティング担当者は「誰が伸びるか」という結果だけでなく、「なぜ伸びるのか」という根拠を把握した上で、納得感を持って戦略的なインフルエンサー選定を行うことが可能になります。
- 今後の課題：ノードの特徴量（画像やテキスト）の貢献度は説明できましたが、GCN の特徴である「エッジ（誰とつながっているか）」の貢献度を IG で直接説明することは困難でした。今後は、グラフ構造そのものの重要度を評価できる GraphLIME などの新たな XAI 手法の導入が課題として挙げられています。