

2017 IEEE International Conference on Data Mining に参加して

松 谷 貫 司

Kanji MATSUTANI

電子情報学専攻修士課程 2017 年度修了

1. はじめに

2017 年 11 月 18 日から同月 21 日にかけて、アメリカ・ニューオーリンズで開催された 2017 IEEE International Conference on Data Mining に参加した。本会議には 50 の異なる国と地域から 778 件の論文が投稿され、そのうち通常論文 72 件とショート論文 83 件が採択された（採択率 19.9%）。私は 18 日に開催されたワークショップで研究成果^[1]をオーラル形式で発表した（図 1）。ワークショップには 261 件の論文が投稿され、我々の論文を含む 148 件が採択された（採択率 56.7%）。



図 1 発表の様子

2. 研究内容

本研究の目的は、ソーシャルメディアに投稿されたアイテム群が人気（以下、ポピュラリティと呼ぶ）を獲得していく過程の数理モデル化である。

2.1 研究背景

Facebook, Twitter, YouTube など、ソーシャルメディアサイトは Web 空間における人々の重要なコ

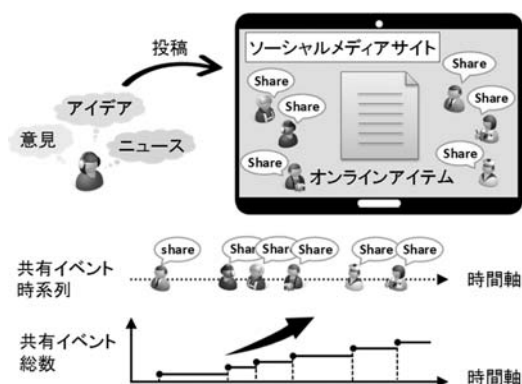


図 2 アイテムに対する共有イベント時系列

ミュニケーションの場として発展し続けており、ニュース・アイデア・意見などの多種多様な情報を容易に、世界に向けて発信したり評価したりすることができるようになってきた。投稿されたオンラインアイテムは、多くの人々に高く評価され共有されていくことによって、そのポピュラリティを獲得していく（図 2）。この現象は、人々の日常生活や社会のトレンド形成に大きな影響を及ぼしうることから、オンラインアイテムが共有され（以下、共有イベントと呼ぶ）ポピュラリティを獲得していく過程のモデル化が、近年ソーシャルメディアマイニングの分野で注目されている。そこで我々は、共有イベント時系列に基づいて、オンラインアイテム群における関係性を抽出し、さらに各アイテムの将来ポピュラリティを精度良く予測することを目指して、統計的機械学習理論に基づく新たな数理モデルを提案した。

2.2 提案モデル

共有イベントの発生過程は、相互エキサイテーション性を有すると考えられている。これは、任意のアイテムに対する過去の共有イベントが、別のアイテムに対する将来の共有イベントの発生を誘発するという性質である。この性質に基づき、計数過程の一種である多変量 Hawkes 過程を用いて、情報伝搬の背後に存在するソーシャルネットワーク（ユーザ間の交友関係）を推定する研究が数多くなされてき

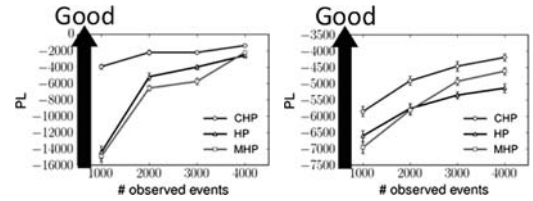
た。

ところで、投稿されたアイテムを誰が共有したかという情報は、プライバシーやセキュリティの関係上、必ずしも公開されているとは限らない。したがって、ソーシャルネットワークを介さず、アイテム間の影響関係を直接モデル化することが望ましい。しかし、ほとんどのソーシャルメディアでは、ユーザ数よりもアイテム数の方がはるかに多く、アイテム間の影響関係を網羅的に学習することは現実的でない。この課題を解決すべく、我々は、アイテム群における協調構造を多変量 Hawkes 過程に組み込んだ新たな確率モデルである、CHP モデルを提案した。CHP モデルは、共有イベント時系列に基づくアイテム群の潜在的なグループを抽出し、それらグループ間の影響関係をモデル化する。さらに我々は、グループ数を観測データから推定するために、ノンパラメトリックベイズ法と呼ばれる手法を用いた効率的な学習アルゴリズムを開発した。

2.3 評価実験

人工データと料理レシピ共有サイト Cookpad のデータを用いて、将来ポピュラリティ予測の観点から、CHP モデルと2つの既存モデル HP モデルおよび MHP モデルの性能を比較した (図3)。HP モデルはアイテム間の影響関係を考慮しないモデルであり、MHP モデルはアイテム間の影響関係を網羅的に学習するモデルである。図3より、CHP モデルが最も高性能で、MHP モデルがそれに続いていることがわかる。特に、観測データ数が少ない場合において CHP および MHP モデル間に顕著な性能差がみられる。この結果は、CHP モデルが過学習することなく相互エキサイテーション性を捉えられることを示唆している。

続いて、共有イベント時系列の観点から、Cookpad の料理レシピ群における協調構造を分析した。我々は、CHP モデルに観測期間のデータを学習させることで22個のグループを抽出した。代表的な



(a) 人工データ (b) Cookpad データ

図3 ポピュラリティ予測実験の結果

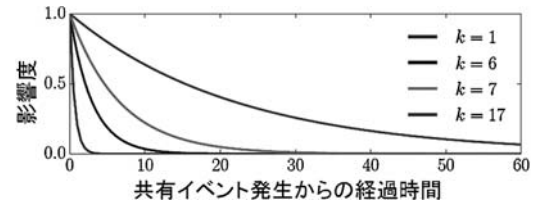


図4 抽出されたグループの影響の減衰

4つのグループにおける共有イベントの影響の減衰パターンを図4に示す。グループ $k=1$ はそうめんに関連したレシピ群であり、影響は全グループ中最も急速に減衰していた。 $k=17$ は、容易にレタスの芯を取り除く方法や、フルーツの皮を綺麗に剥く方法など、効率良く料理するための新規技法と関連したものであり、それらは長期的に影響を与え続けていた。

3. おわりに

国際学会での発表は2度目であったため、発表、質疑への応答ともに英語で上手く述べることができた。質疑・応答では多くの質問と意見をいただき、当該分野と我々の研究への関心の高さが感じられた。

最後に、今回の発表を行うにあたってご指導を頂いた木村昌弘教授、熊野雅仁実験講師、ならびに研究室の皆様に深く感謝致します。

発表論文

- [1] Kanji Matsutani, et al.: Discovering cooperative structure among online items for attention dynamics, *Proceedings of ICDMW 2017*, pp 1033-1041, IEEE, 2017.