

トピックモデルによるソーシャルネットワーク分析

東北大学大学院 経済学研究科 五十嵐未来*

東北大学 照井伸彦

1. はじめに

Twitter や Facebook といった現代のソーシャルネットワーク上には、ユーザー間のつながりを表すネットワーク情報に加え、ユーザーが生成するコンテンツとしてのテキスト情報が存在している。ユーザー同士が相互に関係しあい、企業への態度が変容していく現代社会において、これら 2 種類の情報を適切にモデリングしてソーシャルネットワークの特色を把握することは重要なマーケティング課題となっている。しかし、そのような研究は数が少ないのが現状であり、本研究ではそのミスマッチに対応すべく、トピックモデルを用いたソーシャルネットワーク分析手法を提案する。

2. モデル

本研究では、二人のユーザーの間にリンクが生成される際に 2 種類のプロセスを仮定する。一つがユーザー同士の興味に基づいて生成される場合と、もう一つがオフラインでの関係性の延長線上として生成される場合である。例えば、前者であればスポーツ好きという共通の趣味が要因となって生まれたリンクであり、後者であれば同級生や会社の同僚といった現実での関係性がソーシャルネットワーク上にも反映された際のリンクである。したがって、興味（関係性）に基づくリンクが生成される時、ユーザーは相手のテキスト情報（ネットワーク情報）を参照してリンク生成の有無を決定していると考えられる。

先行研究においては、テキスト情報のトピックモデルとして Latent Dirichlet Allocation [1] が、ネットワーク情報のトピックモデルとして Mixed Membership Stochastic Block-model [2] が代表的な手法として提案されており、提案モデルでは、それら二つのトピックを Correspondence Topic Model [3] の考え方に基づいて対応させている。

そして、ユーザーごとにこれら 2 種類のリンク生成過程のうち、どちらをどの程度重視しているかを示すパラメータを導入することにより、ソーシャルネットワーク分析のためのトピックモデル構築を試みる。

参考文献

- [1] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003a). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- [2] Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9(Sep), 1981-2014.
- [3] Blei, D. M., & Jordan, M. I. (2003b, July). Modeling annotated data. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 127-134). ACM.

* 980-8576 宮城県仙台市青葉区川内 27 番 1 号 東北大学大学院 経済学研究科
mirai.igarashi.s7@dc.tohoku.ac.jp