

## twitter における共通の関心を持つユーザのレコメンド

貴志 将考<sup>†</sup> 大沢 英一<sup>‡</sup>公立はこだて未来大学大学院システム情報科学研究科システム情報科学専攻<sup>†</sup>公立はこだて未来大学 システム情報科学部複雑系知能学科<sup>‡</sup>

## 1. はじめに

twitter は、ユーザ数が 5 億人を超えるマイクロブログサービスである。これを利用する上で、あるユーザの関心のある事柄を扱う他ユーザやつぶやきを検索する方法は、twitter が提供するレコメンド機能と、キーワード検索の 2 つの方法がある。レコメンド機能は、友人や知人などの既知の人物である可能性が高いことが知られている。キーワード検索に関しては、検索を行ったキーワードを使用したユーザが時系列順に表示される。

しかし、上記の方法ではあるユーザが共通の関心を持つ他のユーザを検索したい場合、手作業で効率的に探すことは難しいと考えられる。

本研究では、キーワード抽出とコミュニティ抽出を用いたレコメンドを行い、共通の関心を持つユーザのレコメンド方法の提案を目的とする。

## 2. 関連研究

田中らはフォロー関係や HITS アルゴリズムを利用することにより、特定の興味や関心について有用なつぶやきを行うユーザの推定を行った [2]。しかし、この研究ではつぶやき内容の考慮や、実際にレコメンドを行っての評価がされていない。したがって、本研究ではユーザ毎のつぶやき内容や実際のレコメンドとしての評価を考慮した方法を提案する。

## 3. アプローチ

本研究では、共通の関心を持つユーザのレコメンドを行うために、CNM 法というコミュニティ抽出手法を用いる。

CNM 法とはモジュラリティ  $Q$  値が最大になるようにノード間の結合を繰り返すことでコミュニティを抽出する手法である (図 1)。モジュラリ

ティは効用関数であり、「ノード同士が繋がる割合」から「リンクがランダムに配置された場合の期待値」を引いた値として定義される。

CNM 法は大規模なネットワークに対して有効な手法であることが知られているが、抽出対象のネットワークが多様な関係性を含む場合、ノード間の関連が弱いコミュニティが抽出されてしまうことが知られている [2]。

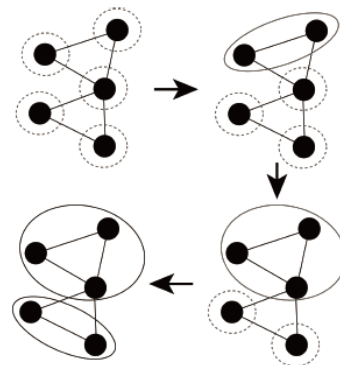


図 1. コミュニティ抽出

実際に twitter では友人知人、仕事仲間、共通の関心などといった様々な関係を含むネットワーク構造となっているため、ユーザをノード、フォロー関係をリンクとして捉えて CNM 法を適用した場合、ノード間の関連が弱いコミュニティが抽出されてしまう。この問題を解決するため、TF-IDF と呼ばれるキーワード抽出手法を用いて、ノード間の関係を特定の関心事に絞ることでこの問題が解決できると考えられる。

## 4. 提案手法

共通の関心を持つユーザのレコメンド方法について図 2 に概要を示す。図 2 では twitter を利用するユーザをノード、フォロー関係をリンクとして捉え、レコメンドを行うユーザを  $a$  とし、直接フォロー関係の無い距離 2 までのネットワークを構築する。距離 2 に居るユーザを  $\alpha \sim \theta$  とし、TF-IDF によってキーワードを抽出し、レコメンドを行うユーザが要求するキーワードに該当するキーワードが抽出されたユーザを検索する。検索されなかったユーザを取り除いた

Recommendation of users with the common concern in twitter  
<sup>†</sup>Masataka KISHI Graduate School of Systems Informations Science, Future University Hakodate.

<sup>‡</sup>Eiichi OSAWA Department of Complex and Intelligent Systems, Future University Hakodate

ネットワークは関連を共通の関心を持つユーザ同士で構築された関連の強いネットワークとなっている可能性が高く、このネットワークに対しコミュニティ抽出を行うことによって、共通の関心を持つユーザ同士のコミュニティが抽出出来ると考えられる。

レコメンドを行うユーザが含まれるコミュニティ内に存在するフォローしていない他のユーザを共通の関心を持つユーザとしてレコメンドを行う。

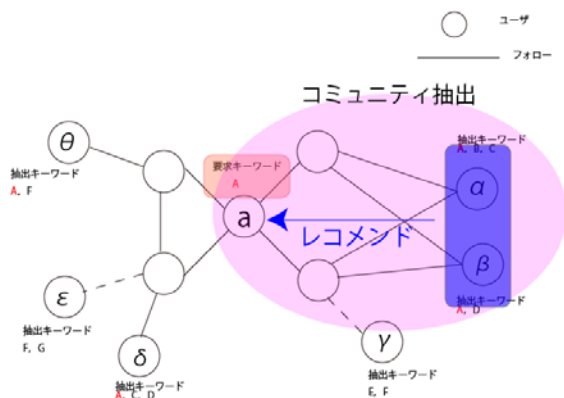


図2. 共通の関心を持つユーザのレコメンド

## 5. 実験

twitter を利用しているユーザ 10 名を対象にキーワードによるリンクの選定を行う場合と行わない場合でのレコメンドの有用性の評価実験を行なった。ここでの有用性とはレコメンドとして利用する価値の有無を有用性としている。

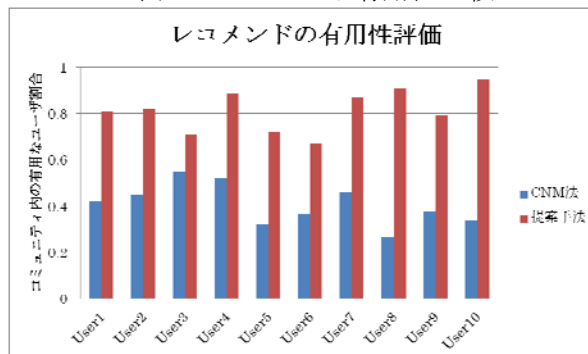
距離 2 のユーザ群の抽出キーワード数はユーザ毎の TF-IDF 値が高いもの 10 ワード、レコメンドを行うユーザの要求キーワード数 3 ワードをそのユーザが持つ関心事とした。キーワードによってリンクの選定を行なった場合と行わなかった場合でのレコメンドを行うユーザが所属するコミュニティのサイズの比較を表 1 に示す。

表 1. コミュニティサイズ比較

	フォロー数	CNM	提案手法
User1	142	378	24
User2	136	772	27
User3	107	984	21
User4	34	162	14
User5	37	124	15
User6	62	572	22
User7	212	2141	31
User8	127	1275	19
User9	138	1547	20
User10	357	2782	35

レコメンドとしての有用性評価の結果を表 2 に示す。また、CNM 法では対象となるコミュニティが大きいため、ランダムに 30 名を評価した。

表 2. レコメンド有用性比較



## 6. 評価, 考察

提案手法を用いることで表 1 からコミュニティの大きさがレコメンドとして利用するうえで有用なサイズとして抽出することが可能であることがわかった。

また、表 2 から共通の関心事によってリンクを選定することによって、レコメンドを行うユーザが所属するコミュニティ内に存在する距離 2 のユーザの有用な割合が大きく向上することがわかった。これはユーザ間の関連を絞り込むことでよりレコメンドを行うユーザが求めている他のユーザを含むコミュニティ抽出を行うことが出来たため、有用なユーザがコミュニティとして抽出されるためだと考えられる。

## 7. 結論

本研究では、twitter のフォロー関係からネットワークを構築し、キーワード抽出によって関連の強いユーザ同士に対しコミュニティ抽出を行うことで、共通の関心事を持つユーザを見出す方法を提案した。キーワードによるリンクの選定を行わない場合に比べ、提案手法はより有用性が高いコミュニティが抽出され、レコメンドとして有用であることを示した。

## 参考文献

- [1] S. Fortunato and M. Barthelemy. Resolution limit in community detection. in *Proc. Natl. Acad. Sci. USA*, Vol. 104, No. 1 pp. 36-41, 2007
- [2] 田中 淳史, 田島 敬史. twitter のツイートに関する分類手法の提案. DEMI Forum 2010