

ユーザー評点データによるネットワーク構造推定 Estimating network structure from score data

鈴木優伽[†] 伏見卓恭[†]
Yuuka Suzuki Takayasu Fushimi

武藤伸明[†] 斉藤和巳[†]
Nobuaki Mutoh Kazumi Saito

1. はじめに

近年様々な事象をネットワークで表すことが可能となった。例えばレビューサイトでは、ユーザーや商品をノードとし、購買活動をリンクとしたネットワークが構築できる。また、レビューサイトでは、ユーザー同士が互いのレビューを参考にしながら購買活動を行っているため、その関係をネットワークと捉えることが可能である。このようなネットワーク上におけるノード間の相互関係やネットワーク構造、ネットワーク上での現象に関する分析や研究は盛んに行われている [1] 仮に、あるノード間の関係やネットワーク構造を予測することは、それらノード間のリンクの有無を予測することに相当する。本研究では、ユーザー間のリンクの結合を、商品などのアイテム群に対するレビュー評点をから予測し、ユーザー間の潜在ネットワークの構築を目的とする。このようなネットワークを構築することで、そのレビューサイトを利用するユーザー群の類型化が図れるとともに、ユーザーの未レビューのアイテムに対する評点を高い精度で予測することが可能となる。ネットワーク上におけるノード間のリンク予測の既存研究として [2][3] などがある。これらの研究では、一部のノード間にリンクが与えられており、それらを用いて、ノードペア間にリンクがあるかどうかを予測している。しかし、レビュー評点を利用しないなど本研究とは構造が異なる。

本稿の構成は以下の通りである、2章で従来法によるネットワーク構築法について触れ、3章で提案法によるネットワーク構築法について説明する、4章で@cosmeのレビューデータを用いた評価実験を行い、最後に5章で本研究のまとめを述べる。

2. ネットワーク構築法

各オブジェクト(本稿では各ユーザー)をノードとし、リンクを張りネットワークを構築する。ノード集合を $V = \{u, \dots, v\}$ 、アイテム集合を $I = \{i, \dots, k\}$ 、ノード u がレビューしたアイテム集合を $I_u \subset I$ とし、ノード u のアイテム k に対する評点を $s(u, k)$ とする。また、ノード u と v の間にあるリンクを e_{uv} 、リンク集合を E とし、ノード集合 V とリンク集合 E からなるネットワークを $G = (V, E)$ とする。

2.1. 単純法-1

単純法-1では、ノード u, v 間のリンク e_{uv} の有無を、以下のように決定する。各ノードのレビューしたアイテム集合において、 $k \in I_u \cap I_v$ となるアイテム k が存在すれば、ノード u, v 間にリンク $e_{u,v}$ を付与する。全ノードペアに対し同様の手法でリンクを付与し、リンク集合 E を形成する。構築したネットワークを単純ネッ

トワーク-1 とする

2.2. 単純法-2

単純法-1と同様、ノード間のリンクの有無は各ノードのレビューしたアイテム集合の共通部分から決定するが、条件を追加する。ノード間に同じ評点がつけられた共通アイテム k が存在するとき、すなわち、 $\exists k \in I_u \cap I_v, s(u, k) = s(v, k)$ であるとき、ノード間にリンクを付与する。得られたリンクからリンク集合を形成し、構築したネットワークを単純ネットワーク-2 とする。

2.3. k -近傍法

k -近傍法(以後 k -NN 法と呼ぶ)では、ノード間のリンクの有無を隣接するノードとの類似度で決定する。まず、全ノードペア間の類似度を計算する。ただし本稿では、類似度として Jaccard 係数を用いる。次に、各ノードに対し、類似度で降順にペアとなるノードを並びかえる。 k の初期値を 1 とし、各ノードは、自身との類似度が k 位までのノードとリンクを結ぶ。全ノードが一つの連結成分になるまで k を増加させながら、上記のリンク結合作業を繰り返す。 k -NN 法で構築したネットワークを k -NN ネットワークとする、

3. 提案法によるネットワーク構築

グラフ G と評点データが与えられたとき、ノード u がアイテム k に評点 $s(u, k)$ を付ける確率は、ノード u の近傍ノードから以下のように推定できると考える。

$$P_u(s(u, k); G) = \frac{|W_u(k; G)| + q(k, s(u, k))}{|F_u(k; G)| + 1} \quad (1)$$

ただし、 $q(k, s(u, k))$ はアイテム k が評点 $s(u, k)$ となる事前確率であり、本稿では全データから推定する。また、 $F_u(k; G)$ はグラフ G において、ノード u と隣接するノードのうち、共通アイテム k をレビューしたノード集合であり、 $F_u(k; G)$ のうち、アイテム k にノード u と同じ評点を付けたノード集合を、 $W_u(k; G) = \{w; w \in F_u(k), s(u, k) = s(w, k)\}$ とする。この時、グラフ G における尤度関数は以下のように計算できる

$$L(G) = \log \prod_{u \in V} \prod_{k \in I_u} p_u(s(u, k); G) \quad (2)$$

提案法では、式 (2),(3) を利用し、すべてのノードペア (u, v) に対するリンク e_{uv} の有無を以下のように決定していく。また、得られたリンクからなるリンク集合を E_{pro} とする。

St1: リンク集合が空のグラフを $G_0 = (V, \emptyset)$ 、単一リンク e_{uv} のみからなるグラフを $G_{uv} = (V, \{e_{uv}\})$ とする。

St2: E_{pro} の初期値を $E_{pro} = \emptyset$ とする。

[†]静岡県立大学

St3: 各 (u, v) に対して, $L(G_{uv}) - L(G_0) > 0$ であるならば, そのノード間のリンクに有効性があると見出し, 保持する.

if $L(G_{uv}) - L(G_0) > 0$ then $E_{pro} \leftarrow E_{pro} \cup \{e_{uv}\}$

St4: E_{pro} からなるグラフを $G = \{V, E_{pro}\}$ とする.

構築したグラフ G を提案ネットワークとする.

4. 評価実験

4.1. データ概要

本研究では, 化粧品のレビューサイトである@cosmeに投稿されたレビューを利用した. 各レビューは, 投稿者, 投稿アイテム, 評点, 投稿時間で構成されており, 本研究では, 10403 ユーザー, 48548 アイテム, 0~7 評点でデータを構成した.

4.2. 評価方法

4手法によって得られたリンク集合を元に各ネットワークを構築する. 構築したネットワークごとに式(1)を計算し, 得られた値をもとに式(2)から, 尤度を求める. また, 次式で正解率を求めて評価する.

$$A(G) = \frac{1}{\sum_{u \in V} |I_u|} \sum_{u \in V} \sum_{k \in I_u} p_u(s(u, k); G) \quad (3)$$

4.3. 分析結果

各ネットワークにおけるリンク数, 尤度, 正解率を表1に表す. なお, $G_0 = (V, \emptyset)$ の評価は $p_u(s(u, k); G) = q(k, s(u, k))$ を用いて行う.

表1: 各ネットワークにおけるリンク数, 尤度, 正解率

ネットワーク	リンク数	尤度	正解率
単純-1	9636292	-4.83e+05	29%
単純-2	2697664	-2.49e+05	54%
10NN	104030	-4.07e+05	34%
100NN	1040300	-4.52e+05	33%
200NN	2080600	-4.65e+05	32%
$G_0(V, \emptyset)$	0	-3.83e+05	34%
提案	2088088	-1.80e+05	63%

表1を見ると, 単純法-1におけるネットワークが尤度, 正解率が共に一番低い結果となった. これらは, ユーザーのアイテムに対する評点を考慮せず構成したネットワークであるため, ノード間のリンクに有効性を持たないものが多く含まれていたと考えられる. また, k -NN法では, k の値を変化させても, 尤度, 正解率共に, 大きな変化は見られなかった. このことから, 単純にリンク数を増やすだけでは尤度, 正解率の改善には結びつかないことがわかる. さらに, G_0 における尤度, 正解率を見ると, 単純法-1, k -NN法と比べてあまり大差が見られず, 単純法-2, 提案法と比べると, 尤度, 正解率共に劣る結果となった. このことから, 事前確率だけ使いリンクを結ばないことは, 単純にリンク数を増やすことと同等であると確認できた. つまり, G_0 ではノード間に有効性の無いリンクの結びつきが発生することを懸念し

て, ノード間に一切リンクを結ばないネットワークを構築したが, それが, 必ずしも良い結果にはならないと考えられる. 一方, 評点を考慮した単純法-2と提案法で構築されたネットワークにおいては, 尤度, 正解率共に他の手法に比べて高い結果となった. これは, リンクを付与する際, 評点を考慮したことで, ノード間のリンクに有効性を持たないノードを排除できた為と考えられる. 提案法におけるネットワークをみると, 尤度, 正解率が一番高い結果となった. 提案法は, ノード間のリンクの有効性を確認してからリンクを付与しているため, より精密なリンク選択が行われたのではないかと考えられる. また, 単純法-2, 提案法によるネットワークの結果を比べると, 提案法の方が, 尤度, 正解率共に改善されている. このことから, 単に評点に着目するだけでなく, それらが発生する確率を考慮した提案法の方が, より精密なリンクの生成が可能であると考えられる.

5. おわりに

本研究では, 評点データからネットワークを構築する方法を提案し, 従来法で得られるネットワークと比較した. その結果, 提案法の有益性を確認できた. 今後は, 本研究で得られた知見を用いて, 今回とは異なるレビューサイトのデータを分析したり, より精度の高い手法を探求していく.

謝辞

科研費(C)(No.25330635)の助成を受けた.

参考文献

- [1] M.J.Salganik, P.S.Dodds, and D.J.Watts. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market", *Science* 10, pp.854-856, 2006
- [2] A.Clauset, C.Moore, and M.E.J.Newman. "Hierarchical structure and the prediction of missing links in networks", *Nature* 453, pp. 98-101, 2008
- [3] D.Liben-Nowell, J.Kleinberg. "The Link-prediction Problem for Social Networks", *Journal of American society for information science and technology*, 58, 2007.
- [4] M.E.J.Newman. "The structure and function of complex networks", *SIAM Review*, 45:167256, 2003.
- [5] M.Gozen-Rodriguez, J.Leskovec, A.Krause. "Inferring Networks of Diffusion and Influence", *ACM Transactions on Knowledge Discovery from Data*, Volume 5 Issue 4, Article No. 21, 2012