

レシピ共有サイトの潜在トピック構造分析

申成植†

熊野雅仁‡

小野景子‡

木村昌弘‡

† 龍谷大学 大学院理工学研究科 電子情報学専攻

‡ 龍谷大学 理工学部 電子情報学科

1 はじめに

近年、ウェブ空間に蓄積されたビッグデータが注目されており、人々の行動情報から有益な情報を得るための分析法が研究されている。本研究では、料理レシピ共有サイトにおいて、あるユーザが投稿した料理レシピを別のユーザが再現する行動に着目し、複数のユーザがどちらも再現するレシピの間には関係があると考えて、それら料理レシピの間にリンクを張ることにより料理レシピネットワークを構築し、料理レシピの潜在構造を分析する。具体的には、ソーシャルメディアにおけるユーザアクティビティと社会的紐帯の生成に関する Kamei[1] らの潜在空間モデルを、料理レシピと料理レシピ間リンクの生成に応用することにより、食材の利用頻度から料理レシピの潜在トピック (潜在的な料理の趣向) を抽出し、さらに、料理レシピ間リンクの形成における潜在的な料理の趣向どうしの特徴的な関係を抽出する。クックパッドのケーキレシピデータを用いた実験により、提案分析法の有効性を検証する。

2 分析法

提案分析法は Kamei らの潜在空間モデルに基づいている。そのモデルでは、 K 次元の離散確率分布で表現される、各料理レシピ u の潜在的な料理の趣向分布 $\pi_u = (\pi_{u,1}, \dots, \pi_{u,K})$ に基づいて、料理レシピ u と、料理レシピ u から料理レシピ v への料理レシピ間リンク (u, v) が生成されるとモデル化する。料理レシピ u は、それらに利用されている食材の頻度ベクトル $X_u = (x_{u,1}, \dots, x_{u,J})$ で表現される。ただし、 u における食材 j の頻度 $x_{u,j}$ は、すべての料理レシピデータにおける j の使用料の平均値と標準偏差をもとに Z 変換に基づいて計算される。食材 j は、LDA[2] に従い π_u を用いて生成される。具体的には、各 $j = 1, \dots, J$ に対して、料理レシピ u の潜在的

な料理の趣向 $z_{u,j}$ を分布 π_u から選択し、食材の頻度 $x_{u,j}$ を J 次元多項分布から、

$$j|z_{u,j}, (\phi_k)_{k=1}^K \sim \text{Mult}(\phi_{z_{u,j}})$$

と生成する。ここで、 J は食材の総種類数であり、 $J-1$ 次元単体の元 $\phi_k = (\phi_{k,1}, \dots, \phi_{k,J})$ は、ある J 次元ディリクレ分布から生成される。また、料理レシピ間リンク (u, v) は、 $(n^2 - n)/2$ 次元多項分布から、

$$(u, v)|w, \Pi \sim \text{Mult}(f(w, \Pi))$$

と生成される。ここに、 n はノード総数であり、

$$f_{u,v}(w, \Pi) = \sum_{k,l=1}^K \frac{\pi_{u,k} \pi_{v,l} w_{k,l}}{\sum_{v' \in V} \pi_{u,k} \sum_{v' \in V} \pi_{v',l}} (u \neq v),$$

$$0 < w_{k,l} < 1, w_{k,l} = w_{l,k}, \sum_{k,l=1}^K w_{k,l} = 1,$$

$$\Pi = \{\pi_u\},$$

$w_{k,l}$ は、料理レシピ間リンクの形成における、潜在的な料理の趣向 k と潜在的な料理の趣向 l との関係の強さを表すパラメータである。データからモデルパラメータを推定し、料理レシピの潜在構造を分析する。

3 実験

実験では、クックパッドのケーキデータを用いて、料理レシピネットワークを構築した。料理レシピネットワークを構築したところ、ノード数は 974、エッジ数は 7,381、総材料数は 199 であった。提案分析法を適用した結果を、表 1 と図 1(a), (b) に示す。推定された潜在的な料理の趣向の総数 \hat{K} は 11 であった。表 1 において、潜在的な料理の趣向名は人手でつけている。各潜在的な料理の趣向 k ごとに、材料 j の生起確率 $\hat{\phi}_{k,j}$ の値と、料理レシピ u の事後確率 $\hat{\pi}_{u,k}$ の値に基づいて、材料 j と料理レシピ u をランキングすることにより、潜在的な料理の趣向 k がどのようなものであるか分析を行った。図 1(a) は、 $\hat{w}_{k,l}$ の値の最大値が輝度 255(白)に、0 が輝度 40(濃灰色)に対応するように正規化して、グレースケールで \hat{w} を可視化した結果である。 $\hat{w}_{4,4} = 0.04$ が最も高

Seishoku SHIN†, Masahito KUMANO‡, Keiko ONO‡ and

Masahiro KIMURA ‡

†Division of Electronics and Informatics, Ryukoku University

‡Department of Electronics and Informatics, Ryukoku University

*1 <http://www.cookpad.com>

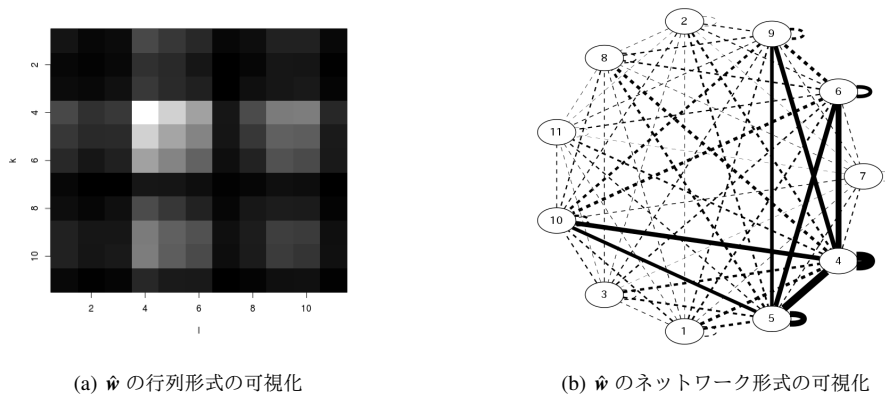


図 1: 潜在空間分析の結果

く、パウンドケーキ趣向同士で料理レシピ間リンクが形成されやすく、さらに、料理レシピ間リンクの結ばれやすさ d は

$$d = \sum_{l=1}^{\hat{K}} \hat{w}_{k,l}$$

パウンドケーキ趣向において最も高い値となった。すなわち、パウンドケーキ趣向は、最も他の趣向と料理レシピ間リンクが形成されやすいという傾向がある。図 1(b) は、潜在的な趣向の関係の強さの推定量 \hat{w} を可視化したものである。平均値 $\langle \hat{w}_{k,l} \rangle$ 以上の $\hat{w}_{k,l}$ を実線で、それ未満の $\hat{w}_{k,l}$ を点線で、それぞれ表している。ただし、強さに応じて太さを変えている。図 1(b) より、潜在的な料理の趣向同士の特徴的な関係が観察される。例えば、チョコレート趣向の間 $\hat{w}_{10,10} = 0.010$ よりも、チョコレート趣向とパウンドケーキ趣向の間 $\hat{w}_{4,10} = 0.022$ の方が約 2.2 倍料理レシピ間リンクが形成されやすい、すなわち、人々がどちらの料理レシピも再現するという傾向が示唆される。

4 まとめ

本研究では、料理レシピ共有サイトにおける潜在トピック構造を分析し、潜在的な料理の趣向を抽出し、趣向間関係の強さを調べ、人々がどちらの料理レシピも再現する趣向間の特徴的な関係を抽出した。今後は、料理レシピ共有サイトのユーザネットワークにおいて、ユーザの料理レシピ投稿行動及び、料理レシピ再現行動を分析していく予定である。さらに、多数のユーザからレシピ再現を受けて人気となる料理レシピを予測し、その予測結果の評価についても検討していきたいと考えている。

表 1: ケーキレシピから抽出した潜在的な料理の趣向

k		趣向名	
$k = 4$		パウンドケーキ	
$\phi_{k=4,j}$	材料名	$\pi_{u,k=4}$	レシピ名
0.17	卵	0.99	完熟バナナのバウンドケーキ
0.15	砂糖	0.99	カラメル林檎のヘルシーチーズケーキ
0.10	バター	0.99	柿 3 個サラダ油卵 1 個きな粉ケーキ
0.09	薄力粉	0.99	幸せの純生スフレロールケーキ
0.08	牛乳	0.99	ごろごろ渋皮煮入り栗のパウンドケーキ
0.04	レモン	0.99	ふわふわリコッタチーズパンケーキ
0.04	塩	0.99	バナナココナッツの米ぬかケーキ
0.03	バナナ	0.99	小麦ふすま入り塩バナナケーキ
0.03	りんご	0.99	りんご入りクルミとシナモンのケーキ
0.02	マーガリン	0.99	栗と南無ヨーグルトフランクランタンケーキ
$k = 10$		チョコレート系	
$\phi_{k=10,j}$	材料名	$\pi_{u,k=10}$	レシピ名
0.16	卵	0.99	バニラ風味のカップチーズケーキ
0.13	砂糖	0.99	大人のチョコバウンドケーキ
0.13	牛乳	0.99	簡単 Xmas デコカップケーキ
0.08	ココア	0.99	チョコレートチーズケーキ
0.06	マーガリン	0.99	ミルクチョコカップケーキ
0.06	バター	0.99	チョコバナナケーキ炊飯器編
0.05	チョコレート	0.99	しっとり満足バナナケーキ
0.05	バナナ	0.99	ふかふか強力粉のパンケーキ
0.05	薄力粉	0.99	HB で黒糖ホットケーキパン
0.03	チーズ	0.99	しっとり濃厚生チョコケーキ

参考文献

- [1] T. Kamei, K. Ono, M. Kumano, M. Kimura, "Predicting missing links in social networks with hierarchical dirichlet processes", IEEE International Joint Conference on Neural Networks(IJCNN), 2012
- [2] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent dirichlet allocation", Journal of Machine Learning Research, vol.3, pp.993-1022, 2003.