

料理レシピ共有サイトにおける 食材情報に基づいた人気レシピ予測

申成植¹ 熊野雅仁² 木村昌弘²

1. はじめに

人類は、太古より新たな食材と出会い、新たな組み合わせを試行し、新たな食材を生み出しながら、日々新しい料理を創造し、料理のフロンティアを拡大し続けている。近年、料理レシピ共有サイトが出現し、大量の料理レシピが Web 空間に蓄積・共有化され続けているため、食材ネットワークの構築に基づいて料理レシピを推薦する研究 [1] や、食材空間上で、文書のトピックモデルを用いて食材をカテゴリ化し、カテゴリに基づいて新たな食材の組み合わせを生成する研究 [2] だけでなく、さらには、料理レシピや食材情報から、新たな料理を創造する計算機の実現を目指す研究 [3] も登場しており、Web 空間上の料理レシピや食材情報がビッグデータとして注目されている。

ところで、レシピ共有サイトでは、日々、料理レシピが投稿されているが、新たに投稿するレシピは、既存の食材や料理法であるよりも、新たな試みを導入したレシピが投稿される傾向が強いものと考えられる。また、新たに投稿されたレシピのうち、その後、人気を得るレシピが存在し得る。しかし、レシピ共有サイトのユーザにとって、新たに投稿されたレシピが新規性を有するか、将来人気を得るか否かなどについては、レシピが投稿された時には予想がつかないものと考えられる。本研究では、新たな料理レシピが投稿されたとき、レシピに含まれる食材情報を入力するだけで、将来人気を得るレシピとなるかを予測する手法の構築に向けて、料理レシピ共有サイトの一つである、Cookpad の料理レシピ情報を利用し、食材情報の有用性を検証する。

2. 料理レシピ情報

本研究における料理レシピ情報は、食材情報とつくれば情報、ユーザ情報で構成される。つくれば情報とは、あるユーザが投稿した料理レシピに従い、別のユーザが料理を再現した際の画像と感想文を投稿する、Cookpad 固有

のユーザアクティビティ情報である。食材情報とは、料理レシピに記載される、料理を作る上で必要な食材名である。食材名を抽出するため、正規表現と形態素解析の後、人手による分類を行った。さらに、塩・粗塩・うす塩といった、細かな食材の違いを無視するため、名寄せ辞書を作成した。抽出された食材名を用いて、料理レシピの特徴量 $x_r = (x_{r,1}, \dots, x_{r,l})$ を構築した。ここで、 $x_{r,i}$ は、料理レシピ r において、食材 i が登場したか否かを表す、バイナリデータである。さらに、食材情報と、ユーザ情報を利用して、ユーザの特徴量 $z_u = (z_{u,1}, \dots, z_{u,l})$ を構築した。 $z_{u,i}$ は、ある期間にユーザ u が投稿した料理レシピ集合において食材 i を使用した料理レシピ数を表す。

本研究では、つくればを多く貰った料理レシピを人気レシピとする。ここでは特に、与えられた期間において、すべての料理レシピが得たつくれば数の平均 μ_r と、料理レシピ r が得たつくれば数 n_r を比較し、料理レシピ r が人気レシピか否かを示す以下のラベル y_r を付与する。

$$y_r = \begin{cases} +1 & (n_r > \mu_r) \\ -1 & (\text{otherwise}) \end{cases}$$

3. 予測法

本研究では、料理レシピの食材情報に着目した手法と、ユーザの重要度に基づく手法を提案する。さらに、ユーザの重要度を、ユーザの人気度を用いた手法と、流行食材に対するユーザの適合度に着目した手法とで予測を行う。これら、二つの観点に基づいた予測法を比較することで、人気レシピの予測における、食材情報の有用性を検証する。

レシピの食材情報に基づく手法 (PIM): PIMとは、学習期間において、食材情報で表現された料理レシピ x_r に対し、人気料理レシピラベル y_r を教師ラベルとして、食材空間上の識別面を学習し、テスト期間における任意の料理レシピが人気を得るか否か、判定する手法である。本研究では、識別面の推定に、ソフトマージン Support Vector Machine(C-SVM)を用いる。

¹ 龍谷大学 大学院理工学研究科 電子情報学専攻

² 龍谷大学 理工学部 電子情報学科

ユーザの適合度に基づく手法 (**UTIM**): UTIM は、過去と比較して、現在流行する食材を多く利用し、廃れた食材はあまり利用しない、食材の流行性に敏感なユーザ $u \in U$ の適合度を評価し、人気となる料理レシピを予測する手法である。過去の食材に関するナイーブベイズパラメータ $\theta = (\theta_1, \dots, \theta_I)$ の各要素

$$\theta_i = \sum_{r=1}^R x_{r,i} / \left(\sum_{i=1}^I \sum_{r=1}^R x_{r,i} \right)$$

から流行性のある食材を求める。学習期間とテスト期間における食材のナイーブベイズパラメータを、 θ^{train} と θ^{test} とすると、テスト期間におけるユーザの特徴量 z_u^{test} から、

$$UTIM(u) = \sum_i z_{u,i}^{test} * (\theta_i^{test} - \theta_i^{train})$$

をユーザスコアとして求める。また、テスト期間における $UTIM(u)$ の平均値 $\mu_{UTIM}^{test} = \sum_{u \in U} UTIM(u) / |U|$ を求め、ラベル \hat{y}_r をユーザ u_r に付与する。

$$\hat{y}_r = \begin{cases} +1 & (UTIM(u_r) > \mu_{UTIM}^{test}) \\ -1 & (\text{otherwise}) \end{cases}$$

$\hat{y}_r = 1$ を得たユーザは、流行食材を用いた料理レシピを投稿し、テスト期間で人気を得ると考えて予測を行う。

ユーザの人気度に基づいた手法 (**FUM**): FUM は、学習期間において、人気度の高いユーザ $u \in U$ が、テスト期間に投稿した料理レシピも人気を得るという仮説のもと、人気レシピを予測する手法である。ユーザ u がある期間に投稿した料理レシピ集合 R_u で得た平均つくれば数 $n_u = \sum_{r \in U} n_r / |R_u|$ に基づき、学習期間における、料理レシピの平均つくれば数 $\mu_r^{train} = \sum_{r \in train} n_r / |R^{train}|$ を求め、

$$u = \begin{cases} famous\ user & (n_u^{train} > \mu_{n,train}) \\ not\ famous\ user & (\text{otherwise}) \end{cases}$$

ユーザが *train* 期間における、人気ユーザか否か分類する。ユーザを分類した後、*test* 期間に投稿した料理レシピに対して、以下のラベル \hat{y}_r に基づいて人気レシピを予測する。

$$\hat{y}_r = \begin{cases} +1 & (u_r = famous\ user) \\ -1 & (\text{otherwise}) \end{cases}$$

FUM と **UTIM** を組み合わせた手法 (**FUM+UTIM**): FUM+UTIM 手法は、UTIM と FUM で求めたユーザのうち、共通のユーザを人気レシピを投稿するユーザと見なす手法である。

4. 実験

実験データとして、2014/5/26 に取得した、レシピ共有サイト Cookpad のケーキデータを表 1 に示す。ここで、dataset1 は訓練データが 2011 年、学習データが 2012 年の

データで、dataset2 は訓練データが 2012 年、学習データが 2013 年のデータである。図 1 は、各予測法における人気レシピの予測結果である。

表 1 実験データ

	dataset1		dataset2	
総食材種類数	451		443	
ユーザ数	980		1,028	
レシピ数 (訓練 学習)	2,656	2308	2943	2264
つくれば数 (訓練 学習)	66,405	34,115	40,024	23,134

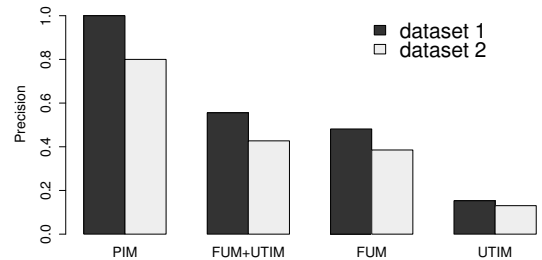


図 1 人気レシピの予測結果

図 1 より、食材情報を用いた PIM が、どちらのデータセットでも、最も高い精度を示した。すなわち、ユーザの重要度に基づいた手法より、料理レシピの食材情報に着目した手法の方が、人気レシピの予測精度が高かった。食材情報は人気レシピの予測に対して有用であると考えられる。

5. まとめ

本研究では、料理レシピ共有サイトにおける料理レシピ情報を用いて、人気レシピの予測を行い、食材の有用性を示した。本研究の本来の目的は、新たに投稿された料理レシピの食材情報のみを捉えることで、人気レシピ得るか否かを予測する手法の構築である。今後は、さらなる分析を行い、食材情報のみによる予測法の提案や、料理レシピの手順や画像といった、他のメタ情報を利用することで、精度の高い予測法の構築を探索する予定である。

参考文献

- [1] C.-Y. Teng, Y.-R. Lin, and L. A. Adamic: Recipe recommendation using ingredient networks, in Proc. 3rd Annu. ACM Web Sci. Conf. (WebSci'12), pp.298-307, 2012.
- [2] V. Nedovic: Learning recipe ingredient space using generative probabilistic models, in Proc. Int. Joint Conf. Artif. Intell. Workshops, pp.13-18, 2013.
- [3] L.R. Varshney, F. Pintel, K.R. Varshney, D. Bhattacharjya, A. Schorgendorfer, and Y. Chee: A Big Data Approach to Computational Creativity. CoRR arXiv:1311.1213. 2013.