

食嗜好の発展を理解するためのレシピ分析

聞 亦晨[†] 浅野 泰仁[†] 吉川 正俊[†]

[†] 京都大学大学院情報学研究科 京都市左京区吉田本町 〒606-8501

1 はじめに

食とは人間の生活を支えるものであり、食文化の発展は人類の歴史に重要な役割を果たしている。例えば農耕も、穀物を煮炊きする道具と技術によって、生では消化しづらい穀物の栄養を効率よく吸収できるようになったことから生まれたといえる。逆に、社会の変化や歴史的事件によって人々の食嗜好が変化し、長期的な食文化を変えていくことも多い。例えば戦後の日本でパン食が普及したことは米軍の占領政策と密接に関連している。このように、食嗜好の発展について学ぶことは人類の文化史を理解する上で重要であり、その将来を予測することも興味深い課題である。

かつては、食の変化は比較的緩やかであった。地域を越えた文化的交流や食材の流通の速度が現在と比べて遙かに遅かったからである。したがって、文献や実地調査といった手法によって食嗜好の発展を捉えることが可能であった。今や、食材の流通は世界規模で迅速になり、インターネットとウェブがその登場以前からは想像もできなかったほど文化的交流の速度を速めている。例えば、日本における Cookpad のように、誰でもレシピをアップロードできるサイトの登場によって、かつては知ることができなかった遠い地域の料理の作り方を知ることができるようになり、それを参考にして新しい料理が作られることも多くなった。結果としてそのようなレシピサイトには、日々生み出される新しい料理のレシピ、すなわち食嗜好の発展に係るデータが大量に蓄積されている。よって、現代は、コンピュータによるデータ解析なしには食嗜好の発展を捉えることは困難な時代になったと言える。また、レシピサイトが今後存続し続ければ、蓄積されたデータはその誕生以後の食文化の歴史の少なからぬ部分を反映するものになることは十分予測される。したがって、現段階でレシピサイトのデータから食嗜好の発展を分析する手法を構築することは、今後の食文化の歴史研究に大いに役立つと考えられる。

本論文では、Cookpad のレシピデータから食嗜好の発展を分析するために、レシピの古典度に基づくランキングを提案する。ある料理、例えば肉じゃがについて、どのレシピが古典的であり、どのレシピがそうでないかが分かれば、肉じゃがについて人々の食嗜好がどのように変わったのかを理解する助けになると考えられる。また、古典的なレシピを好むユーザーにとっ

ては、レシピ検索の助けとなる可能性もある。我々は古典的なレシピの使用食材に関する特徴から得られた二つの仮定に基づくレシピの古典度を定義し、それに基づいていくつかの料理についてレシピのランキングを試み、特に肉じゃがの例を挙げて考察した。

2 関連研究

本研究と関連するレシピデータ分析およびランキング手法について概説する。

レシピデータ分析に関する研究は、その多くがレシピの推薦または調理手順の分析を目的としている。例えば、山肩ら [3] は、エンティティ認識をはじめとする自然言語解析技術を用いて、レシピの特徴を抽出し、調理手順のグラフを作成することで、レシピの差異を分析している。一方、レシピデータを分析して文化や歴史の理解に役立てようという研究はほとんど行われていない。

本研究で用いているレシピの古典度に基づくランキング手法は、情報検索に用いられている様々なランキング手法と関連していると考えられる。例えば VisualRank [1] は、「多くの典型的な画像と類似している画像の典型度は高い」という仮定を用いて、典型度に基づく画像のランキングを行っている。この手法は、類似度が定義できる対象の典型度を計算することも可能であるから、将来的にはレシピの典型度を求め、我々の目的に役立てられる可能性がある。

3 レシピの古典度

今回、我々はレシピ中の「食材」に着目してレシピの古典度を計算することを考えた。いくつかの料理について、多くのレシピを観察したところ、古典的なレシピの多くが以下の二つの仮定を満たしているという結果が得られた。

- (A) ある料理の古典的なレシピは、その料理に使われることが少ない食材をほとんど用いていない。
- (B) ある料理の古典的なレシピは、その料理に良く使われている食材を欠くことはあまりない。

例えば肉じゃがという料理では、よく使われている食材(共通食材と呼ぶ)としては肉(牛肉または豚肉)、じゃがいも、醤油、たまねぎ、にんじん等が挙げられ、使われることが少ない食材(稀食材と呼ぶ)としては、トマト、各種スパイス等が挙げられる。

以下、上記の仮定に基づいて、ある料理 f のあるレシピ a の古典度を求める手法を提案する。まず、 f の名前(例えば肉じゃが)をキーワードとして Cookpad

Recipe Analysis for Understanding Evolution of Food Preference

Yichen Wen[†], Yasuhito Asano[†] and Masatoshi Yoshikawa[†]

[†] Graduate School of Informatics, Kyoto University

Yoshidahonmachi, Sakyo-ku, Kyoto 606-8501

を検索し、レシピの集合を得る。そのレシピの総数を $n(f)$ とする。次に、これらのレシピで用いられている食材を抽出する。Cookpad の書式は比較的自由であるため、同じ食材でも様々な表記が存在するが、形態素解析と料理用語オントロジー [2] を用いて食材名を整理している。料理 f の「共通食材」集合 $C(f)$ は、各食材についてそれを使用しているレシピの数をカウントし、上位 $|C(f)|$ 件の食材を共通食材と見なすことで得るものとする。なお、 $|C(f)|$ としては、料理 f の各レシピ a の食材数 $n(a)$ の最頻値を用いている。また、レシピ a に用いられている各食材を a_i ($1 \leq i \leq n(a)$) で表すことにし、先ほどカウントして得られた、各 a_i が用いられているレシピの数を $d(a_i)$ とする。

上記の数値を用いて、レシピ a の古典度を以下の式で表す a の稀食材使用度と a の共通食材使用度の積で定義する。

$$\text{稀食材使用度: } 1/\lg\left(\frac{\sum_i^{n(a)} \log_2 \frac{n(f)}{d(a_i)}}{n(a)}\right) \quad (1)$$

$$\text{共通食材使用度: } 1/\lg\left(\log_2 \frac{16}{\sum_{a_i \in C_f} \frac{d(a_i)}{n(f)}}\right) \quad (2)$$

稀食材使用度 (1) は、仮定 (A) を反映している。全体の分母に現れる $\frac{n(f)}{d(a_i)}$ は食材 a_i が料理 f のレシピに使用されることが少ないほど大きくなるため、稀食材を使うほど、(1) の値は小さくなる。したがって古典的なレシピではこの値が大きくなると期待される。共通食材使用度 (2) は、仮定 (B) を反映している。共通食材の中で用いていないものがあると、その食材が料理 f によく用いられていればいるほど、この値は小さくなるからである。したがって古典的なレシピではこの値が大きくなると期待される。古典度は稀食材使用度と共通食材使用度の積で表されるため、仮定 (A)(B) を両方満たしているような古典的なレシピの値が大きくなると期待される。

4 予備実験と考察

前節で定義した古典度に基づいて、いくつかの料理についてランキングを行い、著者らが評価したところランキング上位に現れたレシピは古典的なレシピが多く、ランキング下位に現れたレシピはそうでないものが多かった。今後多くの人間による評価を行い、提案手法の妥当性を定量的に検証する予定であるが、今回は「肉じゃが」について得られた結果を示す。

図 1 は、ランキング 1 位のレシピの写真と使用食材 (原文まま) であり、図 2 は、ランキング最下位のレシピのものである。その食材と調理法を見たところ、前者は後者に比べてより古典的な肉じゃがのレシピであると思われる。特に後者は、かなり前衛的なレシピであるように思われる。以下は、ランキング上位 5 件および下位 5 件のレシピの URL のリストである。

上位 5 件 - cookpad.com/recipe/{2142539,2383119,1726529,1795744,2629254}

下位 5 件 - cookpad.com/recipe/{2574168,2099138,1788481,1995584,881282}



図 1: 1 位 (cookpad.com/recipe/2142539)



図 2: 最下位 (cookpad.com/recipe/2574168)

食材: ジャがいも, 牛肉, とろけるチーズ, クッキングイモ, 豚肉, 水, 砂糖, 醤油 トマトソース, 塩, ナツメグ

5 まとめと今後の課題

本論文では、レシピサイトのデータを用いて食嗜好の発展を分析するために、レシピの古典度に基づくランキング手法を提案した。本手法は食材のみの情報、とくに共通食材と稀食材に注目したものであるが、予備実験を行った料理に関しては、ある程度良いランキングが得られていることが確認できた。食材については、化学調味料などの人工的食材の検出を行い、それを古典度に反映させる手法も考えられる。ありふれた食材であっても、その組合せが新しいこともあり得るし、食材の使用量が古典的でない場合もある。さらに、食嗜好の最近の傾向を反映させることも考えている。例えば、レシピの投稿数を見ると、最近牛肉の肉じゃがは減少傾向であり、豚肉の肉じゃがが増加傾向である。また、レシピの古典度は食材だけで決まるものではなく、調理道具や調理手法も大いに関係すると考えられる。今後はこういった要素について検証し、それらを反映する手法を構築して、複数の人間による評価を行い、手法の性能を検証していくのが一つの課題となると考えられる。また、今回の予備実験の結果、古典的でないレシピには、前節の図 2 で挙げたような、将来の食嗜好の発展につながりそうな前衛的なレシピも含まれていたが、そうでないもの (単に残り物を混ぜたりするだけのレシピなど) も多かった。こういったものを分類し、将来の食嗜好の発展を予測することも今後の重要な課題と考えられる。

参考文献

- [1] Jing, Y. and Baluja, S.: VisualRank: Applying PageRank to large-scale image search, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 11, pp. 1877–1890 (2008).
- [2] Nanba, H., Takezawa, T., Doi, Y., Sumiya, K. and Tsujita, M.: Construction of a cooking ontology from cooking recipes and patents, *Proc. Ubicomp 2014*, pp. 507–516 (2014).
- [3] Yamakata, Y., Imahori, S., Sugiyama, Y., Mori, S. and Tanaka, K.: Feature extraction and summarization of recipes using flow graph, *Proc. SocInfo 2013*, pp. 241–254 (2013).