

## 重複レシピ投稿者の検出方法に関する研究 ～レシピランキングサービスの向上を目指して～

### 研究成果のポイント

1. 料理レシピサイトへの重複投稿を検出する手法を開発しました。
2. この手法により、重複レシピユーザーを精度良く検出できることを明らかにしました。
3. ユーザーフレンドリーなレシピランキングサービスの提供につながります。

国立大学法人筑波大学 図書館情報メディア系 関 洋平 准教授らの研究グループは、楽天株式会社ならびに楽天技術研究所との共同研究の中で、レシピランキングサービスにおいて、内容に一部重複があるかたちで記述された低品質なレシピが現れる問題への対処として、調理手順の類似度と料理画像の類似度とを組み合わせることによって重複するレシピの投稿者を検出する手法を開発しました。

インターネット上のユーザにより投稿される多数の料理レシピは、主婦層や一人暮らしの学生を中心に盛んに利用されています。多数のユーザにより投稿されたレシピは、さまざまなバリエーションに富んでおり、調理をするユーザの好みや手持ちの食材などの制約に応じた多様なレシピを提供することが望まれています。たとえば、「肉じゃが」を調理するにしても、手持ちの材料が豚肉か牛肉か、それぞれの食材に応じてどのような味付けが美味しいのかなど、多様な料理レシピの候補を検索できることが、ユーザにとっては好ましいと考えられます。したがって、調理の目的に対して本質的ではない改変でしかないレシピの品質は高くなく、検索結果からは排除することが期待されます。

本研究では、こうした再利用レシピを重複レシピと呼び、投稿したレシピとそれ以前に投稿されたレシピ集合との間で、調理手順間の類似度と料理画像間の類似度とを組み合わせることによって検出する手法を提案し、その有効性を実験により示しました。また、短期間に多数のレシピを投稿するユーザほど、重複レシピを投稿するケースが多いことを明らかにしました。

本研究の成果は、2017年8月20日(日本時間9時30分)付「Proceedings of 9th Workshop for Cooking and Eating Activities in conjunction with the 2017 International Joint Conference on Artificial Intelligence (IJCAI 2017)」で公開されました。

\* 本研究の一部は、楽天株式会社が助成する筑波大学共同研究「スパムレシピのフィルタリングに関する研究」(研究期間:平成28～29年度)によって実施されました。

### 研究の背景

共同研究をしている楽天株式会社が提供するサービス(楽天レシピ)において、インターネット上のユーザにより投稿される料理レシピは、主婦層や一人暮らしの学生を中心に盛んに利用されています。多数のユーザによって投稿されたレシピは、さまざまなバリエーションに富んでいることが期待され、調理をするユーザの好みや手持ちの食材などの制約に応じた多様なレシピの提供が望まれます(図1)。たとえば、「肉じゃが」を調理するにしても、手持ちの

材料が豚肉か牛肉か、それぞれの食材に応じてどのような味付けが美味しいかなど、一通りではない多様な料理レシピの候補を検索できることが、ユーザにとっては好ましい要件となります。料理レシピについても Google に代表されるサーチエンジンと同様の検索サービスが提供されますが、料理名をクエリ（検索語）としたときに、それぞれのユーザに対して適切と考えられるレシピを提供するためには、検索結果の多様性（Diversity）が重要となります。

また、ユーザのなかには、レシピの投稿記事を作成する手間を省くために、以前に投稿したレシピの調理手順の一部や料理画像を再利用して投稿するケースが数多く見受けられます。このような場合、調理の内容に関して本質的ではない改変をした投稿では、レシピの品質は高くないため、検索結果からは排除することができれば、レシピサイトの利用者に質の高い情報を届けることが可能となります。

### 研究内容と成果

本研究ではまず、短期間に多数のレシピを投稿したユーザを、重複レシピを投稿するユーザの候補と見なしました。そしてそのユーザが投稿したレシピと、それ以前に投稿されたレシピ集合との類似度の計算によって類似レシピを多数投稿したことが判明したユーザを、重複レシピ投稿ユーザと見なして検出する手法について検証しました(図2)。実験の結果、調理手順(テキスト)の類似度と、料理画像(色素)の類似度とを組み合わせることにより、重複レシピユーザを精度良く検出できることを明らかにしました。また、短期間に多数のレシピを投稿するユーザほど、重複レシピを投稿するケースが多いことを確認しました。

### 今後の展開

今後の課題として、レシピランキングサービスに本研究提案手法を取り込み、より独創性の高いレシピを上位にランキングするなどの取組みを検討しています。また、類似度の計算に単語の分散表現や Deep Learning などの最新の人工知能技術を取り入れることにより、検出精度のさらなる向上が見込める可能性について検討を続ける予定です。

### 参考図

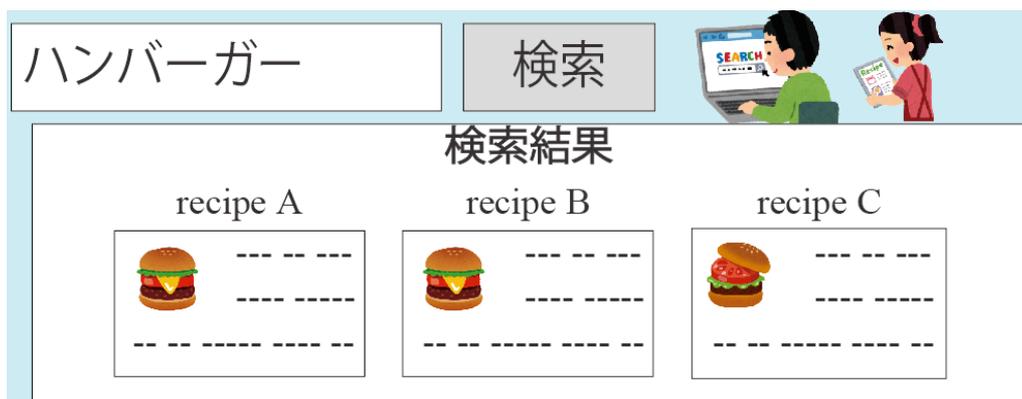


図 1 料理レシピの検索結果における多様性

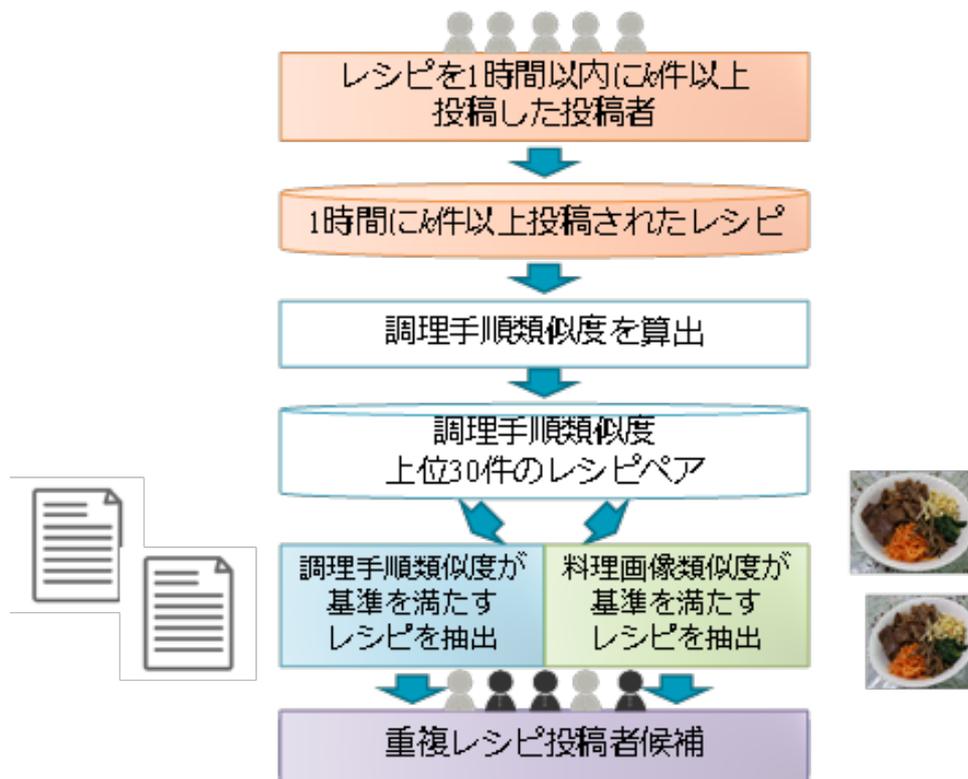


図 2 重複レシピ投稿者の検出

#### 参考文献

久保遥, 関洋平: 投稿型レシピサイトを横断した重複レシピの判別, 第 8 回データ工学と情報マネジメントに関するフォーラム(DEIM2016), C8-3, 2016 年 3 月(優秀インタラクティブ賞受賞).

#### 掲載論文

【題名】 Method for Detecting Near-duplicate Recipe Creators Based on Cooking Instructions and Food Images

(調理手順と料理画像の類似度に基づく重複レシピ投稿者の検出方法)

【著者名】 Masaki Oguni, Yohei Seki, Risako Shimada, and Yu Hirate

【掲載誌】 Proceedings of 9<sup>th</sup> Workshop for Cooking and Eating Activities in conjunction with the 2017 International Joint Conference on Artificial Intelligence (IJCAI 2017)

#### 問い合わせ先

関 洋平(せき ようへい)

筑波大学 図書館情報メディア系 准教授