

調理工程と材料を用いた類似レシピの提案手法

D-8 Proposal method of similar recipes using cooking process and materials

岩瀬 祐大† 天野 一幸†
Yuta IWASE† Kazuyuki AMANO†

†群馬大学大学院理工学部

† Graduate School of Science and Technology, Gunma University

1. はじめに

近年、インターネットやスマートフォンアプリなどの影響で、様々なレシピがオンライン上に混在している。その中で、類似な内容・ジャンルのレシピを手軽に得ることは極めて難しい。本研究では、{ レシピ文書・材料・調理工程数 } を特徴量とした類似レシピの抽出を提案する。

2. 抽出方法

類似度算出のみなら Doc2Vec[1] を用いて、各レシピ文書における単語をベクトル化し、レシピ間の類似度を算出することができる。しかし、Doc2Vec のみでは同ジャンルのレシピの抽出率が低い場合が存在する。よって、以下で示す3つの観点から各レシピにおける類似レシピの抽出を行う。本実験では、「クックパッド」*1のデータセットを使用した。

2.1 レシピ文書

今回使用するレシピは“名詞”、“動詞”、“形容詞”が文書中に15単語以上出現する文書に限定している。

Doc2Vecによりレシピ文書の側面からレシピ間の類似度を計算し、算出された値は[-1, 1]の範囲で表される。このDoc2Vecで算出された値を主軸として考える。

2.2 材料

各レシピにどの材料が使用されているのかを one-hot 表現でベクトル化し、Jaccard 係数で材料の側面からレシピ間の類似度を算出する。Jaccard 係数で算出された値は[0, 1]の範囲で表されるが、今回は Doc2Vec で算出される値の範囲との均一を図るために[-1, 1]の範囲で正規化を行う。

2.3 調理工程数

調理工程数とは、調理工程を指す単語(例: 切る, 焼く, 煮る)がレシピ文書内において何度出現したかをカウントした数値とする。調理工程数においては、検索レシピと被検索レシピそれぞれで生じる差分を特徴量として算出する。各レシピでの差分にはばらつきがあるので、上に同じ理由で[-1, 1]の範囲で正規化を行う。

2.4 類似度算出

2.1~2.3により算出された値から、各レシピにおいて、

(レシピ文書での値, 材料での値, 調理工程数での値)

のような3次元ベクトルを形成する。検索レシピのベクトル: $\mathbf{p} = (p_{レ}, p_{材}, p_{調})$, 被検索レシピのベクトル: $\mathbf{q} = (q_{レ}, q_{材}, q_{調})$ とすると、2点間でのユークリッド距離 ($d(\mathbf{p}, \mathbf{q})$), つまりは、

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_{レ} - q_{レ})^2 + (p_{材} - q_{材})^2 + (p_{調} - q_{調})^2}$$

を測ることで、レシピ間の類似度を算出する。

3. 実験

本実験を行うにあたり、[肉じゃが, カレー, ハンバーグ, オムライス, 唐揚げ, 親子丼, コロッケ, 野菜炒め, 和え物, 酢の物] というような10種類のレシピジャンルを使用する。10(ジャンル) × 1600(レシピ) = 16000(レシピ) をデータとして扱い、その中で、10(ジャンル) × 10(レシピ) = 100(レシピ

) を検証におけるテストデータとする。

1. 単純に Doc2Vec で算出した類似レシピ抽出
2. 新しい提案手法での類似レシピ抽出

の2パターンで実験を行い、各検索レシピにおける類似度上位10レシピがどのレシピジャンルに該当するのかを統計化し、結果を比較する。

4. 結果

パターン1, パターン2それぞれの結果を以下の図1, 図2に示す。赤色で囲われた枠は最大出現数の被検索レシピジャンルを表している。

被検索レシピジャンル

	肉じゃが	カレー	ハンパ	オムラ	唐揚げ	親子丼	コロッ	野菜炒	和え物	酢の物	計	
検索レシピジャンル	肉じゃが	49	5	5	3	11	8	2	10	2	5	100
カレー	13	14	9	7	15	7	7	9	11	8	8	100
ハンパ	8	21	24	15	3	7	9	2	6	5	100	
オムラ	8	11	9	18	12	7	6	12	4	13	100	
唐揚げ	17	6	3	23	36	22	4	14	4	3	100	
親子丼	16	14	13	3	9	17	5	16	4	3	100	
コロッ	6	6	4	11	20	9	20	3	15	6	100	
野菜炒	6	10	13	9	11	6	6	29	3	7	100	
和え物	3	10	13	7	6	10	10	7	22	12	100	
酢の物	9	8	14	9	9	7	6	6	19	13	100	

図1: 単純に Doc2Vec で算出した類似レシピの抽出結果。

被検索レシピジャンル

	肉じゃが	カレー	ハンパ	オムラ	唐揚げ	親子丼	コロッ	野菜炒	和え物	酢の物	計	
検索レシピジャンル	肉じゃが	82	1	1	5	0	3	6	2	0	0	100
カレー	11	55	5	2	0	9	6	11	1	0	0	100
ハンパ	0	13	66	8	0	3	7	3	0	0	0	100
オムラ	1	1	13	81	0	2	1	1	0	0	0	100
唐揚げ	1	5	1	0	92	0	1	0	0	0	0	100
親子丼	2	1	2	13	4	76	0	2	0	0	0	100
コロッ	3	5	4	5	7	2	74	0	0	0	0	100
野菜炒	7	4	1	2	9	0	6	60	9	2	0	100
和え物	0	4	2	3	5	5	2	12	50	17	0	100
酢の物	1	0	3	0	0	2	0	1	9	84	0	100

図2: 新しい提案手法での類似レシピの抽出結果。

5. まとめ

{ レシピ文書・材料・調理工程数 } を特徴量とした類似レシピの抽出手法の結果から、同ジャンルの類似レシピの抽出率と精度の向上が確認できた。課題としては、成果評価のためにユーザー評価の必要などが挙げられる。

謝辞

本研究を遂行するにあたり、クックパッド株式会社と国立情報学研究所が提供する「クックパッドデータ」を利用した。ここに記して謹んで感謝の意を表する。

参考文献

[1] Quoc V Le and Tomas Mikolov. 'Distributed representations of sentences and documents. In International Conference on Machine Learning', Vol.14, pp.1188 - 1196, 2014.

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