

Doc2Vec を用いた料理レシピのベクトル化に基づく 自動カロリー推定システムの構築

加藤 頌大[†] 上田真由美^{††} PanoteSiriaraya^{†††} 中島 伸介^{†††}

[†] 京都産業大学コンピュータ理工学部 〒 603-8047 京都府京都市 北区上賀茂本山

^{††} 流通科学大学経済学部 〒 651-2188 兵庫県神戸市西区学園西町 3-1

^{†††} 京都産業大学情報理工学部 〒 603-8047 京都府京都市 北区上賀茂本山

E-mail: [†]1544386@cc.kyoto-su.ac.jp, ^{††}Mayumi_Ueda@red.umds.ac.jp, ^{†††}spanote@gmail.com,

^{†††}nakajima@cc.kyoto-su.ac.jp

あらまし 近年、料理レシピサイトが数多く存在しており、多くのユーザが料理の献立を検討する際に料理レシピサイトを利用している。また、生活習慣病の予防をはじめとした健康意識が高まり、各料理レシピのカロリーを把握したいという要求が高まっている。このような背景から、料理レシピのカロリー自動計算方式を実現することの意義は高いといえる。そこで、本研究では、既存の料理レシピ提供サイトで提供される料理レシピから、食材や調理方法を抽出し、Doc2Vec でレシピのベクトル化を行う。文章ベクトルとカロリーデータを使い、scikit-learn と SVR(サポートベクトル回帰) で構築した自動カロリー判定器で未知のレシピに対してカロリーを推定する。

キーワード 料理レシピ, Doc2Vec, 自動カロリー判定器, scikit-learn, SVR

1. はじめに

人間は生命維持のために食事をしエネルギーを摂取する必要があるが、栄養を考えず食事を行っても肥満や生活習慣病のような健康を害する原因となりうる。ダイエットをされている方や筋トレをされている方は、一日の摂取カロリーや食材に気をつけながら食事をとっている。そこで、健康管理を行う上で、食事を作る際に摂取するカロリーを把握することができれば自分で栄養管理することも容易になる。

近年は、cookpad [2] や楽天レシピ [3] のような様々な料理レシピ提供サイトが存在する。その他にも、Delish Kitchen [4] やクラシル [5] のような調理工程の動画を紹介するレシピサイトもスマートフォンのアプリケーションで提供されている。味の素レシピ大百科のような管理栄養士といった食の専門家が携わったレシピには、すでにカロリー情報が記載されている。その他にもタンパク質や塩分など豊富な情報が添えられている。しかし、まだまだカロリー情報が記載されたレシピの数は少ない。cookpad のような一般のユーザが料理レシピを投稿するようなサイトでは、毎日レシピが増え続けるため膨大な数のレシピが存在するが、一般のユーザがそれぞれのレシピに対して、栄養情報を記載することは難しい。

そこで本研究では、大量の料理レシピを用いて Doc2Vec [6] により料理レシピのベクトル化を行い、カロリー情報付きの料理レシピを学習データとした機械学習によるレシピカロリー判定器の構築に基づく、自動カロリー計算方式の構築を目標としている。

2. 関連研究

近年、クックパッドや楽天レシピなどの料理レシピサイトを

利用するユーザが増加している。このような料理に関するメディアが増加すると共に、食メディアに関する研究は盛んに行われるようになった。

會下らは、料理をカメラで写真を撮り、CNN を用いて画像からカロリー量を推定するシステムを提案している。食事カテゴリと食材には強い相関関係があるためにカテゴリを学習している。Multi-task CNN を用いることで1種類の食品のカロリーとカテゴリを学習することが可能である。[7]。

崔らは、ユーザのレパートリー不足とメニュー決めが困難というユーザの悩みを解決することを目的とし、定番レシピとの差を考慮した料理レシピ検索手法を提案している。具体的には、メニュー別に頻出する食材の組合せを用いて定番レシピを定義し、これを用いて他のレシピの定番料理との差を明らかにしている。[8]

岩本らは、調理の難易度を考慮したレシピ検索を提案している。クックパッドなどの料理レシピサイトから HTML データを取得して、各レシピの HTML タグを利用して調理手順を抽出している。各調理動作の難易度を階級に分け、料理の難易度を定める。[9]。

従来の研究において、画像からの推定など各料理に関する情報を使用したレシピのカロリー推定は存在せず、本研究を遂行することの意義は大きいと考えている。個々のレシピに対してカロリーを推定できるので、一般のユーザが投稿する楽天レシピやクックパッド等の料理レシピサイトでも利用できると考えている。本研究のように、今後も膨大に追加されていくレシピに対しても対応できるようなシステムの構築は重要な取り組みである。

3. 食材や調理動作を表すキーワード分析に基づく手法 (先行研究) とその課題

先行研究 [10] において、調理方法を考慮したカロリー算出システムについて検討した。カロリー算出システムは、図 1 のように、材料・手順抽出モジュール、単位変換モジュール、カロリー計算モジュールの 3 つのモジュールと、食材名辞書、単位変換辞書、カロリー辞書の 3 つの辞書により構成することを想定している。

ここでカロリー算出方式の算出手順を図 2 に示す。材料・手順抽出モジュールは、入力された既存レシピから、材料部分および調理手順部分の抽出を行う。この際、食材名辞書を参照することにより、表記ゆれに対処し、以降の処理に用いる標準的な食材名に変換する。

通常、料理レシピ中には、「人参 1 本」や「醤油大さじ 1」といった食材特有の単位で表記されているため、抽出した「食材名および分量」を「食材名および分量 (グラム)」に変換する。ここでは、以降の操作で用いる日本食品標準成分表 [11] に対応するため、各食材の分量をグラム単位へ変換する。

その後、文部科学省による日本食品標準成分表に記載された各食材のエネルギーに基づき、各食材を調理手順に記載された調理方法で調理した際のエネルギーを算出し、料理レシピに対するカロリーを算出する。

しかしながら、以下に示すような、料理レシピ特有の問題から、すべてのレシピに自動的に対応することは困難であった。

- 「しょうゆ、みりん、お酒を混ぜます」といった複数の食材に対して一つの調理動作となる場合の対応
- {塩, ソルト} や {青じそ, 大葉} といった、意味が同じだが表記が違う場合の対応
- 複数食材を記号 (例えば, A) で表記し、「ここで A の食材を混ぜます」といった記号で表現される場合の対応

すなわち、レシピ内に出現するキーワード頻度に基づく特徴ベクトルでは、同様の意味のキーワードであってもその表記の揺れ等によって全く異なるものとして扱われることになり、本来のレシピ間類似度の判定を適切に行えないことがある。一方、近年注目を集めている Doc2Vec では、文書中のキーワードの表記ではなく、意味をベクトル化することが可能であり、上述の問題を解決できる可能性がある。そこで、次節にて、Doc2Vec を用いたレシピのベクトル化に基づく自動カロリー計算方式について説明する。

4. Doc2Vec を用いた料理レシピのベクトル化に基づく自動カロリー推定システム

Doc2Vec を用いた料理レシピのカロリー自動算出方式について述べる。Doc2Vec は、文書中のキーワードの表記そのものではなく、キーワードの前後関係も含めた意味をベクトル化することが可能である。すなわち、同じ意味であっても表記の揺れ等に対して全く異なることレシピ上の表記の違いについてはそれほど大きな影響はないといえる。以下に、Doc2Vec を用いた料理レシピのカロリー自動算出方式の大まかな処理手順を示す

(図 3.)。

- (1) できるだけ多くのレシピ文書 (本研究では約 7500 件) を準備し、レシピテキストのクリーニングを行い、使用食材、分量、調理手順以外の不要な情報をできるだけ取り除く。
- (2) 1. にてクリーニングしたレシピデータを用いて、Doc2Vec によりレシピ空間における適切なベクトル次元を学習する。
- (3) 学習した文章ベクトルと料理のカロリー情報を使い、scikit-learn にて機械学習を行い、SVR (サポートベクトル回帰) によって判定器を構築する。
- (4) 3. にて学習した判定器により、カロリーが未知のレシピに対し推定したカロリーを出力する。

以上の通り、提案手法では、一般的なユーザ投稿型レシピサイトにて掲載されている、食材名、分量、調理手順といった情報から、カロリー値を推定する手法の開発を目指している。味の素のレシピサイトである“レシピ大百科”には料理名、材料、調理方法、カロリーなどの栄養情報が記載されている。そこで本研究ではこの“レシピ大百科”に掲載されているレシピおよびカロリー値を学習データとする。

レシピから材料、調理方法を抽出した上で、形態素解析システム (Mecab) を用いて、3 つの品詞 {名詞, 動詞, 形容詞} を抽出した。抽出した単語リストに対して、200 次元とした Doc2Vec によりベクトル化を行う。料理レシピには、それぞれ 4 人前や 2 人前のようにレシピによって作る量が変わる。それに伴い食材の分量も変化する。食材の分量はカロリーにとって有用な情報である。1 人前のカロリーを推定する際に、分量によって変化するのを防ぐため、ベクトル化を行う前に全てのレシピの分量を 1 人前の分量に変換する。

自動カロリー判定器を scikit-learn と SVR (サポートベクトル回帰) を使い、機械学習を行う。機械学習には文章ベクトルと味の素の“レシピ大百科”の料理のカロリーを正解データとしたものを使用する。

5. 評価実験

5.1 実験方法

今回、提案した手法を基にレシピの自動カロリー判定器を構築したものを使い、学習していないレシピに対してカロリーの推定を行なった。提案手法に基づいて構築された自動カロリー判定器が精度よく未知のレシピに対してもカロリーを自動で推定できているかということを実験に基づき考察を行なった。手法としては、Doc2Vec でベクトル化した文章ベクトルとレシピ大百科から取得したカロリーデータを scikit-learn で機械学習を行い構築した自動カロリー判定器を使う。SVR (サポートベクトル回帰) で実際のカロリーと推定したカロリーの誤差をどの程度の範囲なのかを調べた。

実際に使用したデータは味の素が提供している“レシピ大百科”から約 7500 レシピを使い、レシピに記載されている文章とカロリー情報を使用した。使用したレシピのうち 4/5 を学習データとして扱い、残りの 1/5 をテストデータとして扱う。

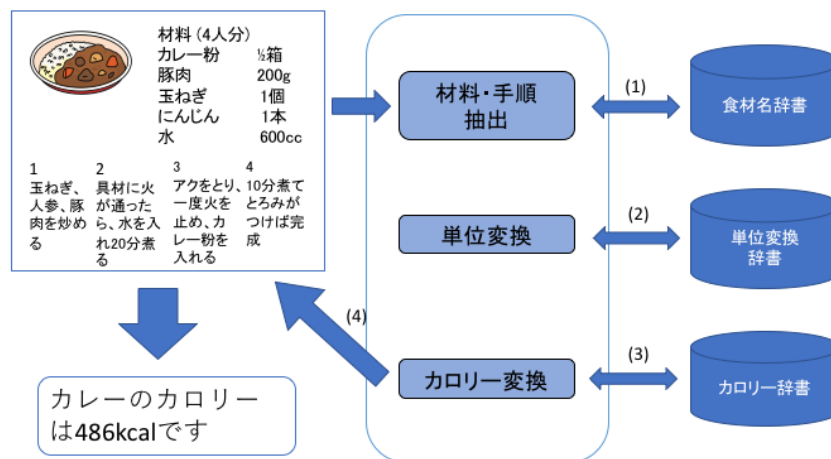


図1 カロリー辞書におけるシステム図

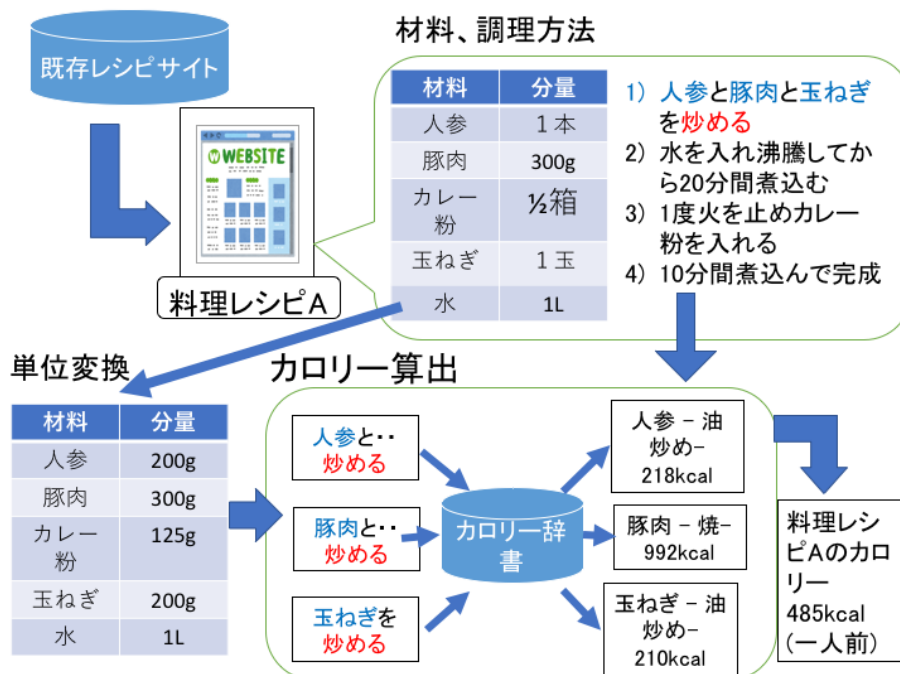


図2 調理方法を考慮したカロリー算出手順

自動カロリー判定器の精度を検証することで、各レシピに対してどの程度のカロリーの誤差で収まるのかを確認することができる。

5.2 自動カロリー判定器によるカロリー推定の評価および考察

前節で説明した実験方法に基づいて、自動カロリー判定器により実際のカロリーと推定したカロリーとの誤差を求めた。実際のカロリーと誤差が少なかった上位7個のレシピを表1で示し、誤差が大きかった上位7個のレシピを表2で示した。

カロリーの誤差が少ないものに関しては、全て誤差が1kcal以下に収まっている。誤差が1以下に収まっていることから正確に推定できていると言って良い。カロリーの誤差が大きいものに関しては、600kcal以上の誤差が出ている。誤差が大きいものの傾向として、「本格チキンカレー」が1128kcal、「豚の角煮」

が1040kcalなど実際のカロリーが大きいものに対して誤差が大きく出ている。

今回の結果では、全てのレシピに対して156kcal 227kcalの間に収まっていた。学習データとテストデータの中身を変えて実験を行なったが、似たように一定の値の間に収まっていた。

今回学習で使用した7500レシピの中には3000kcalや15kcalのように、200kcalに近い値だけを使用したのではなく、カロリーに偏りがようレシピを使用した。正確にカロリーを判定できているレシピが存在しているが、1000kcalのように大きな値に対してはできていないため、対応できるよう実験を進める必要がある。今回はscikit-learnで機械学習を行い、カロリー推定の精度を実験を行う上で、約7500レシピを使用した。データ数としては十分とはいえないと考えており、今後は少なくとも数万件のレシピデータを使って実験を行う必要があると考えて

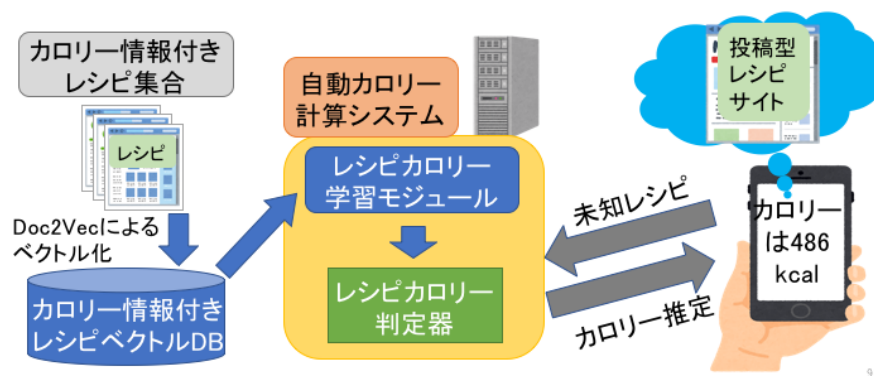


図 3 Doc2Vec を用いた自動カロリー判定器の構成

表 1 正解のカロリーと誤差が少ない上位 7 レシピ

料理名	正解カロリー (kcal)	カロリー推定値 (kcal)	誤差 (%)
餃子鍋	216	216.09	0.004
チョコレートムース	218	217.92	0.035
鰯の黒酢鍋照り	223	223.15	0.068
シャキシャキたけのこと卵の塩味炒め	218	217.84	0.073
ブロッコリーのミモザサラダ	212	212.19	0.088
きのこごぼうとにらのかき揚げ	226	225.48	0.232
ひじき入り豆腐ハンバーグ風	224	224.65	0.290

表 2 正解のカロリーと誤差が大きい上位 7 レシピ

料理名	正解カロリー (kcal)	カロリー推定値 (kcal)	誤差 (%)
本格チキンカレー	1128	225.70	79.99
豚の角煮	1040	216.63	79.17
グリル玉ねぎ、ゆでじゃがいも、ソーセージのコンソメハーブクリーム添え	998	223.79	77.58
チキンかつトーストサンド	987	226.49	77.05
サーモンマリネの洋風すし	975	219.77	77.46
ホクホク長いも入りビーフカレー	976	222.96	77.16
ジンジャーブレッド シナモン風味の生クリーム添え	926	220.97	76.14

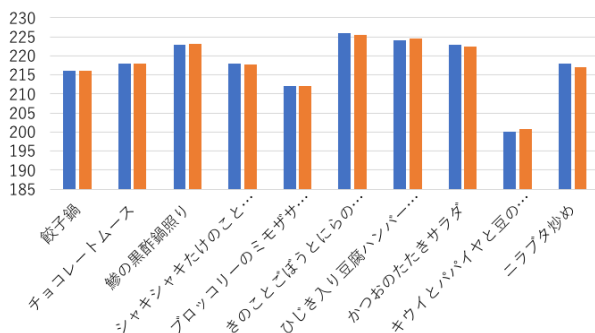


図 4 正解のカロリーと誤差が少ない上位 10 レシピ

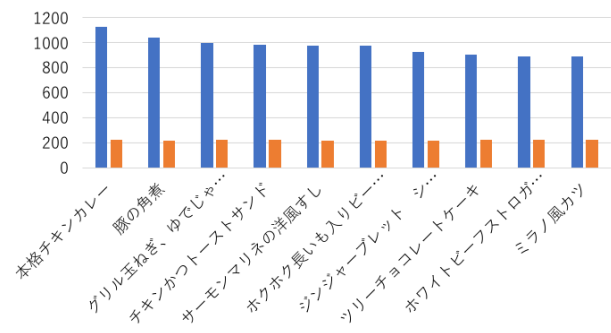


図 5 正解のカロリーと誤差が大きい上位 10 レシピ

いる。

6. まとめと今後の課題

我々は、大量の料理レシピを用いて Doc2Vec により料理レ

シピのベクトル化を行い、カロリー情報付きの料理レシピを学習データとした機械学習によるレシピカロリー判定器の構築に基づく、自動カロリー計算方式の構築し、料理レシピの正確なカロリー推定を目標としている。そこで、本稿は提案手法を基

に自動カロリー判定器を構築し, 味の素のレシピサイトのレシピを使いカロリーの推定の実験を行なった. ある一定のレシピに対しては正確にカロリーが推定できていることが確認できた. しかし, 1000kcal などカロリーが大きいものや 15kcal のような小さいカロリーに対して精度よくカロリー推定できていない. 今後は, Doc2Vec を使用しベクトル化した文章ベクトルや scikit-learn を使った機械学習を見直し, 今回対応できていなかった料理レシピに対しても正確なカロリーが推定できるよう自動カロリー判定器の精度向上を目指す.

文 献

- [1] 味の素レシピ大百科, <https://park.ajinomoto.co.jp/recipe/>
- [2] cookpad, <https://cookpad.com/>
- [3] 楽天レシピ, <https://recipe.rakuten.co.jp/>
- [4] DELISH KITCHEN, <https://delishkitchen.tv/>
- [5] Kurashiru, <https://www.kurashiru.com/>
- [6] Tomas Mikilov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean, “Distributed representations of words and phrases and their compositionality”, Advances in neural information processing systems, pp.3111-3119, 2013.
- [7] 會下拓実, 柳井啓司, Multi-task CNN による食事画像からのカロリー量推定, DEIM Forum 2017 B6-2, 2017.
- [8] 崔 赫仁, 塩井隆円, 楠 和馬, 波多野賢治, メニューごとの共通食材に着目した料理レシピ検索手法の提案, DEIM Forum 2017 C1-1, 2017.
- [9] 岩本純也, 宮森 恒, 調理の難易度を考慮したレシピ検索システムの提案, DEIM Forum 2012 E1-3, 2012.
- [10] 加藤頌大, 上田真由美, 中島伸介, “調理方法を考慮した料理レシピ自動カロリー計算方式”, データ工学と情報マネジメントに関するフォーラム (DEIM2018), P1-6, 2018.
- [11] 日本食品標準成分表, 文部科学省, http://www.mext.go.jp/a_menu/syokuhinseibun/