

大規模レシピに基づく食事記録のカテゴリ推定による 食物摂取頻度解析：アスリートを対象にした実践

Food frequency quest by category estimation of meal record based on large scale recipe Practice for athletes

川原田美雪[†] 唐澤弘明[†] 坂本美沙子[‡] 天野宗佑[§] 山肩洋子[†] 相澤清晴[†]
Miyuki Kawarada Hiroaki Karasawa Misako Sakamoto Sosuke Amano Yoko Yamakata Kiyoharu Aizawa

1. はじめに

アスリートにとって、食事管理はパフォーマンス向上やコンディション維持のために非常に重要である。

食事管理の手法としては、食事記録法^[1]、24時間思い出し法(24hDR)^[2]、食物摂取頻度調査法(Food Frequency Questionnaire, 以後 FFQ)^[3]といった手法が広く利用されている。食事記録法、24hDRは、比較的正確に食習慣の調査ができる一方で、専門の管理栄養士によるフィードバックを前提とした手法であり、大学の運動部に所属する学生など専門の管理栄養士による支援が得られないアスリートにとっては利用が困難な手法である。

FFQは、リスト化された個々の食品について、過去の一定期間の摂取頻度や摂取量を回答するという手法である。FFQは先述の食事記録法や24hDRと比較して簡便であり、同時に複数人の食事調査ができる手法として期待されている。また、日本においてはアスリートを対象にしたFFQは存在しないが、砂見らによって大学生アスリートに栄養管理における妥当性が示されており^[4]、管理栄養士による支援がない学生アスリートにとって非常に有益な手法と考えられる。しかし、正確に回答するためには回答者自身が食の知識を十分有することが必要となるため、より正確に回答するための支援が求められる。

そこで、本研究では十分に栄養や食に関する知識がないスポーツ選手であっても、正確に食物摂取頻度調査を行うことができるような支援システムの構築を目標とする。

本研究では、東京都にある大学の運動部に所属する学生6名の協力の元、12週間分の食事記録を収集した。そして、食事記録と最も一致するレシピを楽天レシピ^[5]から探索し、レシピカテゴリをマルチに推定することによってメニュー内に含まれるFFQの回答項目となる食品を抽出した。また、食事名のカテゴリの分類精度に関する評価実験を行った。

2. 関連研究

本研究では、食事名のテキストから食事が所属するカテゴリ予測を行い食習慣の分析を行った。また、予測には単語の分散表現を用いた。

したがって本節では、食事名の解析に関する研究、レシピカテゴリの予測に関する研究、単語の分散表現に関する研

究、および食習慣の分析に関する研究について説明する。

2.1 食事名の解析に関する研究

天野らは、類似するレシピ名の WordGraph から最小パスを探索することによって食事名を要約し、ワードの出現頻度から食習慣を表現するという手法を提案した^[6]。この手法では、食傾向が出現頻度順の単語で表現でき視覚的に食習慣が分かりやすい一方で、正確な食習慣表現という点では不十分であると考えられる。

また、Kumierczykらは食事名から、トピックモデルを基にクラスタリングと線形回帰を行うことによって、栄養素や所属クラスタを予測する手法を提案した^[7]。食事名の情報のみから詳細な食傾向の分析が可能となる手法であるが、アスリートの繊細な食習慣を分析する上で十分な精度とはいえない。

2.2 レシピカテゴリの予測に関する研究

レシピカテゴリの予測に関して、様々な研究が行われている。Kichererらは、ロジスティック回帰と決定木を用いてレシピを自動分類する手法を提案した^[8]。また、確率論的エントロピーを用いてレシピの複雑さ、レシピ間の類似度を推定し、レシピネットワークを設計する手法も提案されている^{[9][10]}。

これらの手法はレシピ間の類似度を図るという点で非常に有用であるが、本研究での入力情報はアスリート自身が入力した短い食事名のテキストであり単語数が非常に少ないため、分類は困難である。

2.3 単語の分散表現に関する研究

Mikolovらは、単語を実数ベクトルで表現する Skip-gram という表現手法を提案した^[11]。この手法は word2vec と呼ばれるツールとして公開され、広く利用されている。この手法では、単語を密な実数ベクトルに変換することでベクトル間の距離を定義づけることができるため、単語間の類似度の推定に応用されている^{[12][13]}。

Skip-gram において、入力はコーパス、出力は各単語を表現する実数ベクトルである。入力となるコーパスは長さ T の文字列としてみなす。単語列の位置 t の単語を w_t とすると、Skip-gram では以下の目的関数を最大化するように単語ベクトルを最大化する。

$$\frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j} | w_t)$$

[†] 東京大学 The University of Tokyo

[‡] ドームアスリートハウス DOME ATHLETE HOUSE

[§] foo.log 株式会社 foo.log Inc

また、食に関する分野ではレシピに登場する材料を Skip-gram における周辺語とみなし食材や料理名をベクトルで表現する food2vec^[14]や、Diet2vec^[15]が提案されている。

2.4 食習慣の分析に関する研究

合田らは、一食ごとの栄養素をクラスタリング解析し、それぞれのクラスターに所属する食事数の割合を食習慣ベクトルと定義した^[16]。数日間の食事記録を食習慣ベクトルに変換することによって年齢や性別の推定を行い、管理栄養士以上の精度を示した。しかし、全ての食事に対して 31 項目の栄養素が必要となり、登録できる食事データに制限がある。

3. 提案手法

本章では、我々が提案する食事名が所属するマルチカテゴリの推定手法について説明する。我々が提案する手法では、まず食事名を単語ごとに分解する。次に、レシピデータベースに所属する全てのレシピに対して、食事名に含まれるワードと一致するワードの数による類似度と、単語の分散表現に基づく食事名同士のコサイン類似度の 2 つをもとに、食事名とレシピ名間の類似度を求める。最後に、食事名との類似度が上位 K 番目のレシピに対して順位に基づく重み付き多数決を行い、一定の閾値を超えたものを食事名の所属カテゴリとする。提案手法の概要は図 1 に示す。

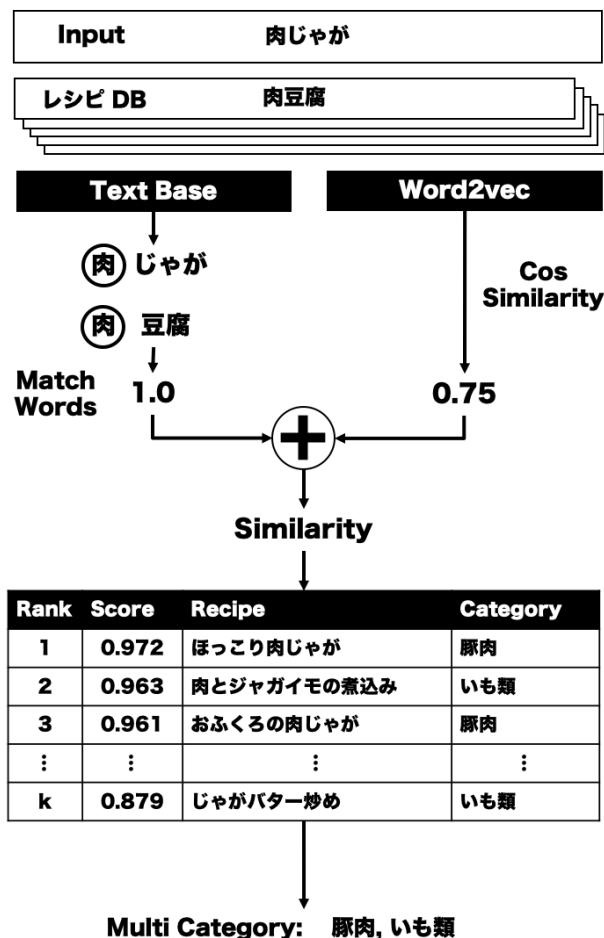


図 1. 提案手法の概要. 食事名とレシピデータ間の類似度を

テキストベース、単語の分散表現の 2 つの指標で表し、上位 K 番目までの重み付き多数決によってカテゴリを決定する。

3.1 単語分割

日本語の食事名を単語に分割するために、日本語形態素解析ツール Mecab^[17] を用いた。実験では、Mecab での分割を行ったのちに不要な単語を省き、名詞、動詞、形容詞のみを食事名として用いた。

3.2 レシピ名と食事名の類似度の推定

レシピ名 (R) と食事名 (F) の間の類似度の推定には、一致するワードの個数 s_{and} と、word2vec を用いた名前の近似度 s_{word} の 2 つの指標の重み付き線形和を用いる。本研究では、 $\lambda = 0.3$ とした。

$$similarity(R, F) = (1 - \lambda)s_{and}(R, F) + \lambda s_{word}(R, F)$$

3.2.1 一致するワードに関する類似度

単語ごとに分割された食事名とレシピ名に対して、一致したワード数を類似度とする。例えば、“ミックス天井”と“ミックスフライ定食”との間の類似度を求める場合には、“ミックス”が一致しているため類似度は 1 となる。

$$s_{and}(R, F) = n(R \& F)$$

3.2.2 Word2vec に関する類似度

Wikipedia データ^[18]と基本料理知識ベース^[19]と楽天レシピ^[5]を基にコーパスを形成し、word2vec のモデルを構築する。食事名を単語ごとに分割し、単語のうち名詞、動詞、形容詞のみを抽出して 200 次元の実数ベクトルに変換し、それぞれの単語ベクトルの線形和を食事名のベクトルとする。比較するレシピデータセットに対しても同様の処理を行い、2 つのベクトルのコサイン類似度を単語の分散表現に基づく類似度の指標とする。

$$s_{word}(R, F) = \frac{vecR \cdot vecF}{\|vecR\| \|vecF\|}$$

3.3 重み付き多数決によるマルチカテゴリの推定

3.2 の手法を用いて、全てのレシピデータセットに対して類似度の計算を行い、類似度が上位 K 番目のレシピを抽出する。K 個のレシピが所属するカテゴリごとの類似度の平均を求め、類似度が閾値を超え、かつ K 個内のレシピが多く所属するカテゴリを出力する。

4. 実験データセット

4.1 食物摂取頻度調査表 (FFQ)

本研究では、レシピが所属するカテゴリを食物摂取頻度調査表 (FFQ)^[3]に基づいて集計する。FFQ では、食材の摂取に関して 120 の項目が存在するが、そのうち飲料、調味料等の質問や、朝食の頻度等の質問を除いた 87 個の食材カテゴリを用いる。

4.2 楽天レシピ

食品名のデータセットとして、楽天レシピ^[5]に 2013 年 4

月までに公開された 441,927 件の投稿レシピのうち、飲み物、ソース・調味料・ジャム、お弁当、季節・イベントのカテゴリに所属するレシピを除く 407,632 件を用いた。

また、楽天レシピは 1 つのレシピにつき 1 つのカテゴリに所属している。本研究で用いるレシピは、楽天レシピが定める 262 種類のカテゴリのいずれかに所属している。

4.3 食品成分表

食品名のデータセットとして、楽天レシピとともに、文部科学省が公開する日本食品標準成分表^[20]から食品名 3,307 件を利用した。() 内に含まれる単語など、一部の単語は事前に削除し実験を行った。

4.4 Eat Smart データ

評価実験に用いる食事名のデータとして、Eat Smart^[21]に一般料理として登録された 1,870 件の食事名のうち、調味料、飲料を除いた 1,618 件を用いた。

4.5 Word2vec のコーパスに用いたデータセット

Word2vec の構築には、Wikipedia^[18]のコーパスを用いた。また、食品に関する単語の相関を考慮するために基本料理知識ベース^[19]、楽天レシピデータ^[6]もコーパスに追加した。基本料理知識ベースとは、食事名に材料や調理法、属性が付与されたデータベースである。具体例を表 1 で示す。

表 1. 基本料理知識ベースの一例。「あさりの酒蒸し」。

<p>【あさりの酒蒸し】 [材料] <あさり> ガーリック.3 ネギ.3 生姜.2 酒 油.6 醤油.5 コショウ.2 食塩.2 [調理法] <蒸す> [属性] おさけ(102/材料関連).5 国産のあさり(102/材料関連).5 タウリンが豊富(103/栄養素) おさけのお供に(105/付け合せ).5 和懐石料理(121/料理ジャンル).5 居酒屋のメニュー(221/場所).</p>
--

4.6 アスリート食事記録データ

東京都にある大学の運動部に所属するアスリート 6 名の 12 週間 (2017/2/1 - 2017/4/30) の食事記録を利用した。食事記録は毎日全ての食事を記録しており、食事名は全て Eat Smart の一般料理データの中から抽出されている。今回の実験においては、サプリメントや飲料の摂取記録は考慮しないこととする。

5. 評価実験

食事名のカテゴリ予測に関する精度評価を行うために、Eat Smart の一般料理に登録された全 1,618 件の食事名に対してカテゴリ推定を行った。食事名が所属するカテゴリの分類は必ずしも一意には定まらないため、1 つのレシピに対して最大 3 つのカテゴリを正解カテゴリとした。

評価の指標には正解率(Accuracy)と Dice 係数(DSC)を用いた。正解率は、推定したカテゴリが正解カテゴリの部分集合となる場合を正答として計算する。また、Dice 係数は、以下の式で表される指標である。

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

テストデータセットは、Eat Smart に一般料理で登録された 1,618 件全ての食事名 (All) と、実際にアスリートが登録を行った食事名 (Athlete) の 2 つを用いて実験を行った。

実験結果を表 2 に示す。表 2 より、全ての食事意を対象にした場合よりも実際にアスリートが食べる一般食のみを対象にした方が高精度に分類できていることがわかる。これは、実際の食事はパターンも少なく、データベース内に該当するレシピ数も多いためと考えられる。

表 2. カテゴリ推定精度の評価実験の結果。

	Accuracy	DSC
All	0.756	0.634
Athlete	0.773	0.670

表 3. カテゴリ推定における正答の一例

食事名	カテゴリ
にぎり寿司 (あなご)	精白ご飯 白身の魚
マカデミアナッツ	—
焼き魚 (さば)	青身の魚
エビの春巻き	エビ・カニ フライ・揚げ物
カプレーゼ	トマト チーズ

表 4. カテゴリ推定における誤答の一例

食事名	カテゴリ
カップラーメン	—
めんたいこ	スパゲティ
マシュマロ	食パン・ロールパン 洋菓子
バナナ	洋菓子
ソフトクリーム	—

また、実際のカテゴリ推定の結果を表 3、表 4 に示す。表 3 はカテゴリを全て予測できた例、表 4 には誤答の例である。

表 3 において、マカデミアナッツは FFQ の回答項目に該当するカテゴリが存在しないため正解カテゴリは存在しない。実際のカテゴリ推定実験においても、十分な閾値となるカテゴリが存在しなかった。

表 4 のカップラーメン、ソフトクリームの例では、インスタント麺やアイスクリームのカテゴリに所属するレシピがデータベース内にあまり存在しなかったために十分な類似度となるレシピ数が少なく、カテゴリ推定ができなかったと考えられる。また、めんたいこ、マシュマロ、バナナの例も同様に単品食材に関するデータが少なかったため、食材と

相関の高いカテゴリが抽出されていることがわかる。

また、実際のアスリートの食事記録に対してカテゴリ推定を行った場合の集計結果を表6で示す。6人の食事は栄養素ベースでは非常に類似しているが、表6のようにカテゴリベースで食傾向を解析するとその違いが可視化できることがわかった。

表6. アスリート6人の食事記録に対するカテゴリ推定結果の出現順上位10位。

No.1	No.2	No.3	No.4	No.5	No.6
精白ご飯	洋菓子	洋菓子	精白ご飯	精白ご飯	精白ご飯
洋菓子	精白ご飯	精白ご飯	洋菓子	洋菓子	豆腐
淡色野菜	淡色野菜	淡色野菜	淡色野菜	淡色野菜	淡色野菜
食パン	食パン	豆腐	卵	大根	洋菓子
豚肉	大根	漬物	漬物	食パン	食パン
鶏肉	豆腐	白身の魚	菓子パン	豚肉	卵
豆腐	鶏肉	卵	食パン	和菓子	牛肉
大豆	中華麺	大根	大根	鶏肉	大根
大根	調理パン	さけ	大豆	キャベツ	鶏肉
調理パン	漬物	鶏肉	豚肉	カレー	豚肉

6. まとめと展望

本研究では、アスリートの食事管理における食物摂取頻度調査表の記入簡便化のために、食事名から食事カテゴリをマルチに推定する手法を提案した。単語の一致による類似度と単語の分散表現による類似度を組み合わせることで、大規模レシピデータベースから類似したレシピを抽出し、多数決を行うことによって、食事名のカテゴリを推定できることを実験により示した。また、カテゴリ推定によって繊細な食傾向の違いも可視化できることが明らかになった。

今後の課題として、飲料や調味料の分類にも取り組む必要があると考える。本研究ではレシピをデータベースとしたためこれらの分類を行うことはできなかったが、実際にFFQに回答する場合には飲料、調味料に関する設問も多く存在する。より分類精度を向上させ、調味料や飲料まで分類が行うことによって、さらに簡便にFFQに基づく正確な食習慣分析が可能になると考えられる。

文献

- [1] Taguchi M., Tatsuta W., Higuchi M., Resting Energy Expenditure of Female Athletes in Different Types of Sport. *Jpn. J. Nutr. Diet.* 68, 5, pp.289-297, 2010
- [2] Sunami A., Sasaki K., Suzuki Y., Oguma N., Ishihara J., Nakai A., Yasuda J., Yokoyama Y., Yoshizaki T., Tada Y., Hida A., Kawano Y., Validity of a Semi-Quantitative Food Frequency Questionnaire for Collegiate Athletes. *J Epidemiol.* 26, 6, 284-291, 2016
- [3] Kokubo Y., Yokoyama Y., Kisara K., Ohira Y., Sunami A., Yoshizaki T., Tada Y., Ishizaki S., Hida A., Kawano Y., Relationship between Dietary Factors and Bodily Iron Status among Japanese Collegiate Elite Female Rhythmic Gymnasts. *Int J Sport Nutr Exerc Metab.* 26, 2, pp.105-113, 2016
- [4] 砂見綾香, 鈴木良雄, 安田純, 多田由紀, 日田安寿美, 川野因, 大学生アスリートにおける10食品群の摂取頻度と食物摂取重量との関連, *日本食育学会誌*, 11巻, 1号, pp. 3-11, 2017
- [5] 楽天レシピ, <https://recipe.rakuten.co.jp/>, (2019/06/18 閲覧)
- [6] 天野宗佑, 相澤清晴, 小川誠, 短縮による食事名データの要約表現, *電子情報通信学会技術報告*, vol.114, no.485, pp.147-152, 2015
- [7] Kusmierczyk, Tomasz & Nørnvåg, Kjetil., Online Food Recipe Title Semantics: Combining Nutrient Facts and Topics. *10.1145/2983323.2983897*, 2016
- [8] Hanna Kicherer, Marcel Dittich, Lukas Grebe, Christian Scheible, Roman Klinger, What you use, not what you do: Automatic classification and similarity detection of recipes, *Data & Knowledge Engineering*, Vol. 117, pp. 252-263, 2018
- [9] Su-Do Kim, Yun-Jung Lee, Hwan-Gue Cho, Seong-Min Yoon, Constructing Cookery Network based on Ingredient Entropy Measure, *Indian Journal of Science & Technology*, Vol.9, Issue 26, 2016
- [10] Young-joo Chung, Finding food entity relation -ships using user-generated data in recipe service, *CIKM, Proceedings of the 21st ACM international conference on Information and knowledge management*, pp.2611-2614, 2012
- [11] Tomas Mikolov, Ilya Sutskever, Kai Cjen, Greg Corrado, Jeffrey Dean, Distributed Representations of Words and Phrases and their Compositionality, *Neural Information Processing Systems 2013*, pp. 3111-3119, 2013
- [12] 城光英彰, 松田源立, 山口和紀, “文脈限定 Skip-gram による同義語獲得に関する研究”, *言語処理学会第22回年次大会 発表論文集*, pp. 717-720, 2016.
- [13] 有賀 竣哉, 鶴岡 慶雅, “単語のベクトル表現による文脈に応じた単語の同義語拡張”, *言語処理学会第21回年次大会 発表論文集*, pp. 752-755, 2015.
- [14] food2vec-Augmented cooking with machine intelligence, <https://jaan.io/food2vec-augmented-cooking-machine-intelligence/>, (2019/6/18 閲覧)
- [15] Tansey, Wesley and Lowe Jr, Edward W and Scott, James G, “Diet2Vec: Multi-scale analysis of massive dietary data,” *The 2016 NIPS Workshop on Machine Learning for Health*, 2016.
- [16] 合田悠治, 天野宗佑, 山肩洋子, 相澤清晴, 大規模食事記録データと深層学習モデルによる食習慣ベクトル化手法の検討, *映情学技報*, vol. 43, no. 5, ME2019-35, pp. 115-120, 2019.
- [17] Mecab: Yet Another Part-of-Speech and Morphological Analyzer, <https://taku910.github.io/~mecab/>, 2015
- [18] Wikipedia, <https://www.wikipedia.org>, (2019/6/18 閲覧)
- [19] 清丸寛一, 黒橋禎夫, 遠藤充, 山上勝義: 料理レシピとクラウドソーシングに基づく基本料理知識ベースの構築, *言語処理学会第24回年次大会*, 2018.
- [20] 日本食品標準成分表 2015年版, http://www.mext.go.jp/a_menu/syokuhinseibun/1365295.htm, (2019/06/18 閲覧)
- [21] EatSmart, <https://www.eatsmart.jp/>, (2019/6/18 閲覧)