卒業論文

タイトルはここ

english title

富山県立大学大学院 工学研究科 電子・情報工学専攻
2355020 水上和秀
指導教員 António Oliveira Nzinga René 講師
提出年月:

目 次

図一覧 **ii**

表一覧 **iii**

記号一覧 **iv**

第 1 章　はじめに **1**
§ 1.1 本研究の背景 1
§ 1.2 本研究の目的 2
§ 1.3 本論文の概要 3

第 2 章　関連研究 **4**
§ 2.1 他分野および音楽分野における推薦システムの事例 4
§ 2.2 感情分析に関する研究例 7
§ 2.3 (音楽と) ポジティブ心理学について 8

第 3 章　理論の説明 **8**
§ 3.1 BERT による文章のベクトル化と感情分析 8
§ 3.2 まだ未定 11
§ 3.3 未定 11

第 4 章　提案手法 **12**
§ 4.1 テキストマイニングによるレビューの取得 (レビューの前処理) 12
§ 4.2 12
§ 4.3 提案システムの概要 12

第 5 章　数値実験並びに考察 **14**
§ 5.1 数値実験の概要 14
§ 5.2 実験結果と考察 14

第 6 章　おわりに **15**

謝辞 **16**

参考文献 **17**

図一覧

2.1 レコメンドシステムのイメージ 5
2.2 情報推薦システムの分類 5

3.1 BERT の流れ 9
3.2 BERT の入力部分 9

表一覧

記号一覧

以下に本論文において用いられる用語と記号の対応表を示す.

用語	記号
j 人目の使用者の名前	ϵ_j
j 人目の身長	α_j
j 人目の体重	β_j
j 人目の基礎代謝量（下限）	B_j^L
j 人目基礎代謝量（上限）	B_j^H
j 人目のアレルギー情報	x_j
j 人の有する生活習慣病	z_j
対象の日数	D
レシピの数	R
食材の数	Q
栄養素の数	N
データベース上の食材数	S
データベース上の食材番号	$d : 1, 2, 3, \cdots, S$
日の番号	$k : 1, 2, 3, \cdots, 3D$
栄養素の番号	$l : 1, 2, 3, \cdots, N$
材料の番号	$m : 1, 2, 3, \cdots, Q$
レシピの番号	$i : 1, 2, 3, \cdots, R$
i 番目のレシピの名前	y_i
i 番目のレシピの献立フラグ	r_{ki}
i 番目のレシピの主菜フラグ	σ_i
i 番目のレシピの調理時間	T_i
i 番目のレシピの摂取カロリー	C_i
i 番目のレシピの調理コスト	G_i
i 番目のレシピの m 番目の材料の名前	q_{im}
i 番目のレシピの m 番目の材料量	e_{im}
i 番目のレシピの l 番目の栄養素の名前	n_{il}
i 番目のレシピの l 番目の栄養素の量	f_{il}
d 番目の食材名	Z_d
d 番目の食材の販売単位	W_d
d 番目の食材の値段	M_d

第1章

はじめに

§ 1.1　本研究の背景

§ 1.2　本研究の目的

く

§ 1.3　本論文の概要

本論文は次のように構成される．

第 1 章 本研究の背景と目的について説明する．背景ではを述べる。目的ではを述べる。

第 2 章 関連研究について述べる。

第 3 章 本研究で用いる

第 4 章 本研究の提案手法について述べる

第 5 章

第 6 章 本研究で述べている提案手法をまとめて説明する．また，今後の課題について述べる．

第2章

関連研究

§ 2.1　他分野および音楽分野における推薦システムの事例

近年、ソーシャルメディアの発展やスマートフォンなどのICT機器の普及により、多くの人々がインターネットに触れる機会が増えている。その中でも、YouTube[1]をはじめとする動画配信サイトや、Amazon[2]をはじめとするEC（Electronic Commerce：電子商取引）サイトなどを利用する機会が増加している。これらのサービスでは、動画や商品を閲覧している際に、関連性の高い別の動画や商品を提示されることが一般的である。このような技術は「情報推薦システム」と呼ばれ、現在その重要性が高まっている。また、情報推薦システムによりユーザーは膨大な選択肢の中から適切な情報や商品を効率的に発見できるようになり、現代の情報過多の時代において重要な役割を果たしている。レコメンドシステムのイメージを図 2.1 に示す.

推薦システムの目的
推薦システムにはいくつかの手法が存在し、

ルールベース推薦

ルールベース推薦とは、システムの運営者があらかじめ「特定の行動をとった人や特定の属性を持つ人に対してどのような商品や情報を提供するか」というルールを設定し、そのルールにしたがってレコメンドを行う手法である。例えば、パスタを購入しようとしている人にチーズやソースを推薦するように、利用者の行動や嗜好を予測して適切だと思われるルールを設定する。この手法は比較的低コストで容易に実装可能であり、透明性が高いという利点がある。運営者が設定したルールに基づいて商品を推薦するだけなので、推薦の結果を簡単に説明できる。しかし、あらかじめ設定したルールに依存するため、利用者の多様なニーズや嗜好を細かく反映することが難しいという欠点がある。また、利用者の行動パターンが変化した場合に対応するには、ルールを手動で更新する必要があるため、運用コストが高くなる可能性がある。

コンテンツベースフィルタリング

コンテンツベースフィルタリングは、アイテム（利用者に推薦する対象やコンテンツ）の特徴量を使用し、利用者の嗜好に近いアイテムを推薦する手法である。例えば、映画を推薦するシステムを考えてみる。この場合、利用者が「アクション映画」や「サス

[1] https://www.youtube.com
[2] https://www.amazon.co.jp

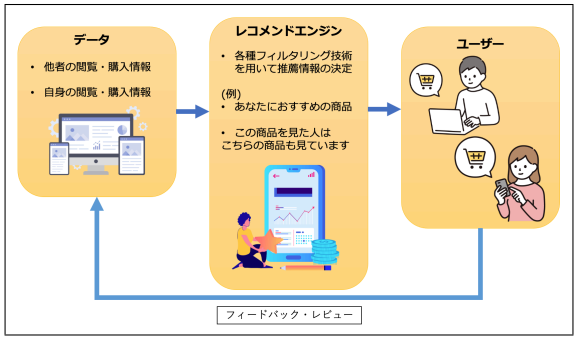

図 2.1: レコメンドシステムのイメージ

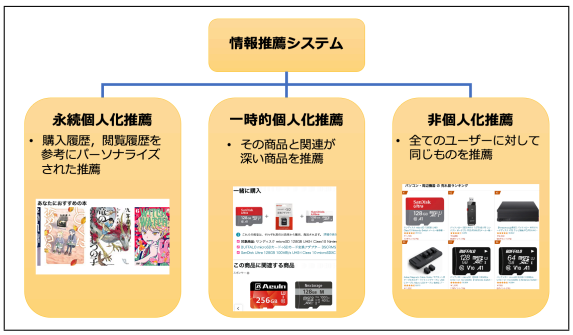

図 2.2: 情報推薦システムの分類

ペンス」といったジャンルを検索すると、システムはそれらの特徴量を分析し、ジャンルやテーマが類似する映画を選び出して推薦する。また、過去に視聴した映画のデータを基に、類似した内容や雰囲気の映画を提案することも可能である。この手法は、利用者の過去の嗜好に基づき、個人に特化した精度の高い推薦ができるという利点がある。利用者が以前に好んだアイテムと特徴の類似度を計算しアイテムを優先的に推薦するため、個々の利用者に特化した推薦が可能である。一方、この手法は過去の嗜好に基づいて推薦を行うため、利用者に同じようなアイテムが繰り返し推薦され、新しいアイテムの発見が難しいという欠点がある。コンテンツベースフィルタリングには、利用者が自分の好むものを直接指定する「直接指定コンテンツベースフィルタリング」と、利用者の嗜好データからプロファイルを作成し、アイテムデータベースと比較することで嗜好を測る「間接指定コンテンツベースフィルタリング」が存在する。

推薦システムには永続個人化推薦と一時的個人化，非個人化推薦の 3 種類に分類することができる [?]．それぞれのイメージを図 2.2 に示す．

協調フィルタリング

協調フィルタリングは、システム使用前に利用者の嗜好データを収集し、それを基に利用者の好みや傾向（例えば、どのアイテムが好まれ、どれが嫌われるか）を分析する手法である。この手法では、嗜好が類似している利用者同士が、共通して好むアイテムや嫌うアイテムがあると仮定し、似たような嗜好を持つユーザーを見つけ出します。その後、そのユーザーが好むと思われるアイテムを推薦する。

協調フィルタリングは大きく分けてユーザーベース協調フィルタリングとアイテムベース協調フィルタリングの二つに分類される。ユーザーベース協調フィルタリングとは似た嗜好を持つ他のユーザーを見つけ、そのユーザーが高く評価したアイテムを推薦する手法のことであり、アイテムベース協調フィルタリングとはアイテム同士の類似性に基づき、ユーザーが過去に好んだアイテムに似たアイテムを推薦手法のことである。これら手法は他の利用者の行動データを活用するため、利用者の過去の評価や行動に基づいて適切な推薦ができるという特徴がある。しかし、協調フィルタリングはシステムに新規ユーザーや新規アイテムが追加された場合に過去のデータがないため推薦することが難しくなるという「コールドスタート問題」やユーザーとアイテムの評価データが非常に少ないためにゆーザー間の類似性を正確に算出できず推薦精度が低下してしまう「スパース性問題」などの課題が

ある。これらの問題を解決するために、コンテンツベースフィルタリングやハイブリッドフィルタリングと組み合わせて使われることが多い。

ハイブリッドフィルタリング

ハイブリッドフィルタリングとは、コンテンツベースフィルタリングと協調フィルタリングを組み合わせた手法である。コンテンツベースフィルタリングは推薦がパターン化されやすく，協調フィルタリングは，新規ユーザーや新規アイテムに対して機能しにくいといった欠点があるが，ハイブリッドフィルタリングはそれらの欠点が補われ，幅広いユーザーに適切な推薦を行うことができる．

推薦システムは、現在、様々な分野で活用されており、その適用事例も増えている。例えば、EC サイトの分野では商品の推薦、教育分野では学習コンテンツの提示、利用者のニーズや好みに応じた提案を行う仕組みが使われている。こうした他分野での具体的な推薦システムの事例について紹介し、その後、音楽分野における推薦システムの事例を紹介する。

他分野での推薦システムの事例

1. Amazon の推薦システムの事例
   1

2. Coursera における推薦推薦システムの事例
   2

音楽分野分野での推薦システムの事例

他分野での事例を書く

1. YoutubeMusic での推薦システムの事例
   い

2. AppleMusic における推薦システムの事例
   あ

本研究への関連性

他分野での事例を書く

§ 2.2　感情分析に関する研究例

感情分析とは自然言語処理の分野の一つであり、テキストや音声、画像などのデータから人々の感情的な状態や糸を抽出し、それを定量的・定性的に分析する技術のことである。主に、

感情分析の場合は

ネガポジ (ネガティブ・ポジティブ) 分析とは感情分析の手法の一つであり、

感情分析の手法

感情分析の手法は大きく分けて 2 つのアプローチがある。

感情辞書による手法

この手法は感情極性辞書を用いて、単語やフレーズに割り与えられたポジティブまたはネガティブスコアを集計することでテキストの感情を推定する手法である。感情極性辞書を用いた感情推定の流れを説明する。まず、対象となる文章を形態素解析と係り受け解析により単語に分割する。次に、分割した単語を感情極性辞書に基づいてポジティブまたはネガティブに振り分ける。最後に、各単語の情報を集計し、文章全体の感情を推定する。

形態素解析とは、文章を言語において意味を持つ最小単位 (形態素) に細分化することである。形態素解析で文章を形態素に分割し、係受け解析により各形態素の係り先を決定する。日本語の感情分析で使われる感情極性辞書は以下の二つがあげられる

単語感情極性対応表

単語感情極性対応表は、東京工業大学の高村教授が公開した、スピンモデル (SPIN Model) を用いて作成された感情辞書である。スピンモデルとは物理学のスピン理論を応用し、単語の感情極性を文脈に基づいて推定する手法である。単語が同じ文脈で共起する場合、それらの単語は同じ感情極性を持つと仮定し、語彙ネットワークを構築する。このネットワークを用いて、単語同士の感情的な関係性を計算し、感情極性を決定している。

日本語感情極性辞書

日本語感情極性辞書は、東北大学の乾研究室が公開した極性辞書であり、約 8,500 の名詞および複合名詞に感情極性情報を人手で付与している。名詞の評価極性は、評価・感情、状態、行為などを基準に決定されている。

深層学習による手法

・それぞれの利点について書く

感情分析のアプローチ

他の機械学習を用いた感情分析の事例

感情分析の応用分野

§ 2.3　(音楽と) ポジティブ心理学について

第3章

理論の説明

§ 3.1　BERT による文章のベクトル化と感情分析

BERT の技術的説明を書く。

BERT

BERT は，Google が提案した自然言語処理モデルの一つであり，Transformer の Encorder 部分を基盤とし，Attention メカニズムを用いて単語間の関係性をモデル化している．

BERT による処理の流れを図??に示す．持っていることが最大の特徴であり、文脈を考慮して単語の意味を理解することができる。また、BERT は、テキスト内の単語の位置情報を学習することができるため、単語の順序を考慮した文脈理解が可能である。これにより、文書分類、質問応答、言語翻訳など、様々な自然言語処理タスクに応用されている。BERT は、大規模なコーパスから事前に学習されたモデルであり、学習済みのモデルを転移学習することで、より小規模なデータセットでも高い精度で処理することができる。

結果として，多くの NLP タスクで従来手法を上回る精度を達成しており，BERT は分散表現と転移学習において大きな進展をもたらした．

Transformer

近年，翻訳などの入力文章を別の文章で出力するというモデルは，Attention を用いたエンコーダー，デコーダ形式の RNN が主流であった．しかし RNN は逐次的に単語を処理しているため、訓練時に並列処理ができないという欠点があった。それに対し Transformer は，RNN や CNN を用いず Attention のみを用いている．Transformer は，再帰も畳み込みも一切行わないので並列化が容易であり，他のタスクにも汎用性が高いという特徴がある．Transformer はベクトル化された文章を入力とし、Attention を多数並列に配置した Multi-Head Attention が用いて分析を行っている。

BERT における入力部分

言語を用いたタスクを解く際には、モデルが言語を扱えるように数値化する必要がある。BERT では、まず MeCab を使用して文を単語に分割し、その後 WordPiece を用いて単語をさらにトークンに分割する。BERT の日本語モデルでは、32,000 個のトークンが定義されており、各トークンには固有の ID が割り振られている。BERT への入力時には、このトー

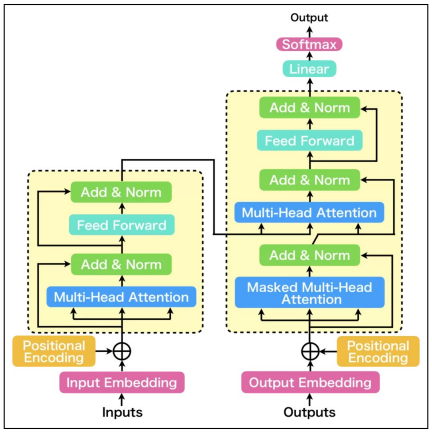

図 3.1: BERT の流れ

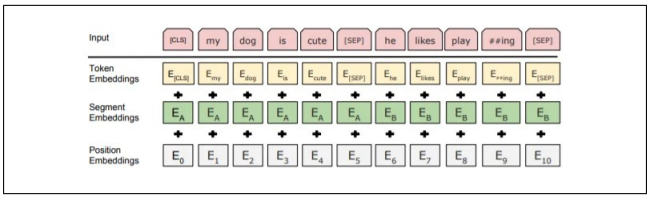

図 3.2: BERT の入力部分

クン ID が使用される。トークン ID に変換されたデータは、BERT モデルに入力される前に、以下の 3 種類の埋め込みを加えることで、モデルに適した数値ベクトルとして表現される。**1. トークン埋め込み**

トークンごとに、事前学習された埋め込み行列 $\mathbf{E} \in \mathbb{R}^{|V| \times d}$ を用いてベクトル表現に変換する。ここで、$|V|$ は語彙サイズ、d は埋め込みベクトルの次元である。各トークン ID t_i に対応する埋め込みベクトル $\mathbf{e}_i$ は次のように定義される。

$$\mathbf{e}_i = \mathbf{E}[t_i] \tag{3.1}$$

このベクトル $\mathbf{e}_i$ は、語彙内のトークンの意味的な特徴を学習した固定長ベクトルである。

2. 位置埋め込み

トークン埋め込みを行っただけでは入力の順序に関する情報を持たないため、文章を正しく扱えなくなる可能性がある．そのため、トークン列内での順序情報をモデルに追加するため、位置埋め込みを加える。位置埋め込み行列は $\mathbf{P} \in \mathbb{R}^{n \times d}$ として定義され、各位置 i に対応するベクトル $\mathbf{p}_i$ は次のように計算される

$$\mathbf{p}_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \tag{3.2}$$

$$\mathbf{p}_{(pos,\ 2i+1)} = cos\left(\frac{pos}{10000^{\frac{2_i}{d_{model}}}}\right) \tag{3.3}$$

ここで、 pos はトークンの位置（0, 1, 2, ...） 、 i は埋め込み次元のインデックスである。これにより、固定長の埋め込みベクトルを用いながら、トークンの相対的な順序や距離をモデルが学習可能となる。

3. セグメント埋め込み

BERT では、1 つの入力が単一文か複数文かを区別するために、セグメント埋め込みを使用する。各トークンには、対応するセグメント ID s_i（文 1 なら 0、文 2 なら 1）が割り振られ、埋め込みベクトルは以下のように計算される。

$$\mathbf{s}_i = \mathbf{S}[s_i] \tag{3.4}$$

ここで、$\mathbf{S} \in \mathbb{R}^{2\times d}$ はセグメント埋め込み行列であり、s_i が 0 または 1 に応じて適切なベクトルが選択される。

4. 埋め込みベクトルの統合

最終的に、トークン埋め込み、位置埋め込み、セグメント埋め込みを加算して、モデルに入力するベクトル $\mathbf{x}_i$ を生成する：

$$\mathbf{x}_i = \mathbf{e}_i + \mathbf{p}_i + \mathbf{s}_i \tag{3.5}$$

これにより、各トークンには、その意味（トークン埋め込み）、位置（位置埋め込み）、文区別（セグメント埋め込み）の情報が含まれるベクトルが追加される。

Attention

Attention は、入力の各トークンが他のすべてのトークンにどれだけ関連しているかを学習するメカニズムであり、この機構は、一般的に自己注意（Self-Attention）として知られている。Self-Attention では、各トークンの埋め込みベクトルを Query、Key、および Value という 3 つのベクトルに変換し、その相関関係を計算して、重み付けされた値を集約する。Query、Key、および Value は入力単語 x_i にそれぞれの重み W_Q、W_Q、W_Q を用いて以下の式で定式化される。

$$Q_i = W_Q \cdot x_i, \qquad K_i = W_K \cdot x_i, \qquad V_i = W_V \cdot x_i \tag{3.6}$$

また、3 つの相関関係は各トークンのクエリとキーとの内積を計算することで求めることができる。

$$score(Q_i, K_j) = \frac{QK^T}{\sqrt{d_k}} \tag{3.7}$$

ここで、d_k は Query と key の次元数である。得られた式にソフトマックス関数を適応し、V_i を付加重することによって Attention を求めることができる。

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3.8}$$

Multi-Head Attention

Attention は一つの計算を逐次的に行っているため、1 つの視点からしか文脈を読み取ることができない。これに対し、Multi-Head Attention は複数の Attention を並列に連結した構造をしているため、各ヘッドが独自の視点から情報を捉え、最終的にそれらを結合して出力を得ます。このアプローチにより、情報の多様な側面を同時に捉えることができる。Multi-Head Attention は以下の式で定式化される。

$$Multi\text{-}HeadAttention(Q, K, V) = Concat(head_1, head_2, \cdots, head_h)W_o \tag{3.9}$$

$$\text{where } head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{3.10}$$

モデルの学習方法

- **事前学習**
  事前学習では、BERT は Masked Language Modeling (MLM) と Next Sentence Prediction (NSP) という 2 つのタスクを使ってモデルを訓練する。MLM では、入力文の中でランダムに選ばれた単語を「[MASK]」トークンで置き換え、モデルにその単語を予測させる。MLM において、Multi-Head Attention は文脈を双方向的に考慮し、マスクされた単語を予測する。Multi-Head Attention は、文全体の意味的および構文的な関係を理解し、正しい予測を行うための情報を集約する。

- **ファインチューニング**
  ファインチューニングをする

§ 3.2　まだ未定

§ 3.3　未定

第4章

提案手法

§ 4.1　テキストマイニングによるレビューの取得 (レビューの前処理)

§ 4.2

§ 4.3　提案システムの概要

第5章

数値実験並びに考察

§ 5.1　数値実験の概要

§ 5.2　実験結果と考察

第6章

おわりに

謝辞

本研究を遂行するにあたり，多大なご指導と終始懇切丁寧なご鞭撻を賜った富山県立大学工学部電子・情報工学科情報基盤工学講座の António Oliveira Nzinga René 講師，奥原浩之教授に深甚な謝意を表します．最後になりましたが，多大な協力をしていただいた研究室の同輩諸氏に感謝致します．

2025 年 2 月

水上和秀

参考文献

[1] 公益社団法人 千葉県栄養士会，“生活習慣病の予防、食生活 生活習慣病の予防と食事”，https://www.eiyou-chiba.or.jp/commons/shokuji-kou/preventive/seikatusyukan/，閲覧日 2023.1.7.

[2] 国立研究開発法人 国立循環器病研究センター，“食事療法について ”，https://www.ncvc.go.jp/hospital/pub/knowledge/diet/diet02/，閲覧日 2023.1.7

[3] ソフトム株式会社，“ソフトム通信　第 79 号「給食業界におけるＡＩ活用」”，https://data.nifcloud.com/blog/food-service-provider_ai-use-case_01/，閲覧日 2022.12.28.

[4] 貝沼やす子，江間章子，“日常の献立作りの実態に関する調査研究（第１報）”，日本調理学会誌，Vol.30，No. 4，pp. 364-371，1997.

[5] 株式会社おいしい健康，“おいしい健康”，https://oishi-kenko.com/，閲覧日 2022.10.16.

[6] 総務省統計局，“小売り物価統計調査による価格調査”，https://jpmarket-conditions.com/，閲覧日 2022.10.11.

[7] J. W. Ratcliff and D. Metzener，“Pattern Matching: The Gestalt Approach”，*Dr. Dobb’s Journal*，p.46，1988.

[8] C. A. Coello Coello and M. S. Lechuga，“MOPSO: a proposal for multiple objective particle swarm optimization，”*Proceedings of the 2002 Congress on Evolutionary Computation (CEC’02)*，Vol. 2，pp. 1051-1056，2002.

[9] Q. Zhang and H. Li，“MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition”，*IEEE Trans. Evolutionary Computation*，*Vol. 11*，No. 6，pp. 712–731，2007.

[10] LeftLetter，“多目的進化型アルゴリズム MOEA/D とその改良手法”，https://qiita.com/LeftLetter/items/a10d5c7e133cc0a679fa，閲覧日 2023.1.6.

[11] J. H. Holland，“Adaptation in Natural and Artificial Systems”，1975.

[12] K. Deb，A. Pratap，S. Agarwal and T. Meyarivan，“A Fast and Elitist Multi-objective Genetic Algorithm: NSGA-II”，*IEEE Tran. on Evolutionary Computation*，Vol. 6，No. 2，pp. 182-197，2002.

[13] D. E. Goldberg，“Genetic algorithms in search，optimization and machine learning ”，*Addison-Wesly*，1989.

[14] メディカル・ケア・サービス株式会社，“制限食にはどんな種類があるの？”，健達ネット，https://www.mcsg.co.jp/kentatsu/health-care/12106，閲覧日 2023.1.6.

[15] ときわ会栄養指導課，“減塩について”，栄養指導，
http://www.tokiwa.or.jp/nutrition/diet/low-salt.html， 閲覧日 2023.01.15

[16] 全国健康保険協会，“ちょっとした工夫で脂質をコントロール”，
https://www.kyoukaikenpo.or.jp/g4/cat450/sb4501/p004/， 閲覧日 2023.01.15

[17] 厚生労働省，“日本人の食事摂取基準 (2020 年度版)”，
https://www.mhlw.go.jp/content/10904750/000586559.pdf， 閲覧日 2023.01.15

[18] 東京医科大学病院，“カリウムは調理のくふうで減らせます”，内臓内科，
https://articles.oishi-kenko.com/syokujinokihon/dialysis/05/， 閲覧日 2023.01.15

[19] 厚生労働省，“糖尿病”，https://www.mhlw.go.jp/content/10904750/000586592.pdf，
閲覧日 2023.01.17

[20] 厚生労働省，“慢性腎臓病”，https://www.mhlw.go.jp/content/10904750/000586595.pdf，
閲覧日 2023.01.17

[21] 腎臓内科，“慢性腎臓病の食事療法”，東京女子医科大学，
https://www.twmu.ac.jp/NEP/shokujiryouhou.html， 閲覧日 2023.01.17

[22] 厚生労働省，“脂質異常症”，https://www.mhlw.go.jp/content/10904750/000586590.pdf，
閲覧日 2023.01.17

[23] 厚生労働省，“高血圧”，https://www.mhlw.go.jp/content/10904750/000586583.pdf，
閲覧日 2023.01.17

[24] 厚生労働省，“食べ物アレルギー”，アレルギーポータル，
https://allergyportal.jp/knowledge/food/， 閲覧日 2023.01.17

[25] J. Blank，“pymoo: Multi-objective Optimization in Python ”，
https://www.egr.msu.edu/ kdeb/papers/c2020001.pdf，閲覧日 2023.1.22.

[26] 和正敏，“多目的線形計画問題に対する対話型ファジィ意思決定手法とその応用 ”，電子情報通信学会論文誌 Vol. J 65-A， No. ll， pp. 1182-1189， 1982.

[27] 厚生労働省，“日本人の食事摂取基準 (2020 年版) ”，
https://www.mhlw.go.jp/content/10904750/000586553.pdf， 閲覧日 2022.12.26.

[28] 農林水産省，“一日に必要なエネルギー量と摂取の目安”，
https:/www.maff.go.jp/j/syokuiku/zissen_navi/balance/required.html， 閲覧日 2023.1.22.