

Skip-gram を用いた SNS からの料理名抽出と コンテキストウェアなレシピ推薦への一考察

落合 桂一[†] 深澤 佑介[†]

[†] 株式会社 NTT ドコモ 〒 239-8536 神奈川県横須賀市光の丘 3-6

E-mail: †{ochiaiike,fukazawayu}@nttdocomo.com

あらまし 食事は健康のために重要な要素の 1 つであり、食事の内容を決めることは重要である。食事内容を決める際には個人の好みや季節、費用など様々な要因が影響する。従来研究では要因の影響を調査するために少人数でのユーザ調査が行われている。しかしながら、長期的に同一ユーザの食事の履歴を観測することはコストがかかるため行われていない。そこで、我々はソーシャルメディアへの投稿から料理名を抽出することで低コストに食事傾向を観測することを提案する。ソーシャルメディアを活用する上で、各投稿文章から料理名を抽出する必要がある。料理名を抽出する方法として料理名辞書を用いたキーワードマッチによる方法が考えられる。この方法では、料理名の辞書を精度よく作れば料理名も精度よく抽出できるが、辞書構築に人手や時間のコストがかかる。そこで本研究では、辞書ベースの方法に対して、Skip-gram を利用して半自動で辞書を拡張することを提案する。あらかじめシードとなる一定量の料理名辞書を構築し、料理名の分散表現のベクトルから平均的な料理名を表すベクトル（平均料理名ベクトル）を生成する。料理名は名詞のサブセットであると考え、SNS の投稿を形態素解析し名詞を抽出後、平均料理名ベクトルとのコサイン類似度が高い単語を料理名として抽出することで料理名辞書を拡張する。料理名辞書とのキーワードマッチにより判定する方法と機械学習を用いた固有表現抽出により料理名を抽出する方法をベースラインとして評価を行った。

キーワード 料理名抽出, Word2vec, Skip-gram, 固有表現抽出

1. はじめに

食事は健康のために重要な要素の 1 つである。そのため食事の内容を決めることは重要である。しかしながら、日本では主婦の多くが献立決めで悩まされている。例えば、株式会社ミニーが全国の 20 代～40 代の主婦約 4,000 名を対象に行った料理に関するアンケートでは約 70 % が「毎日、何を作るか迷う」と献立に悩んでいる^(注1)。食事内容を決める際には栄養バランスや個人の好み、季節、費用など様々な要因が影響する [1], [2]。そこで、コンテキストに応じて食事内容の推薦を行うことでユーザを支援することが考えられる。

推薦を行う際の 1 つの要素として季節や曜日などの時間的なコンテキストを考慮することが考えられる。時間的な要因として曜日や季節の影響がアンケートで調査されている [3], [4]。しかしながら、これらの研究では対象となるユーザ数が数十名程度と限られていたり、数日分の食事内容について質問が行われており長期的なデータを得ることは難しい。

そこで、我々はソーシャルメディア（以下、SNS）に着目する。SNS では料理に関しての投稿も行われており、投稿から各ユーザが「いつ」「何」を食べたのが取得できる。SNS を利用して食事行動を分析するには、各投稿の文章から料理名を抽出する必要がある。料理名を抽出する方法として大きく以下の 2 つの方法が考えられる。

(1) 料理名辞書を使った抽出（辞書ベース手法）

(2) 機械学習による固有表現抽出（機械学習ベース手法）

(1) の方法は、料理名辞書を作れば精度よく抽出できるが、辞書構築にコストがかかる。一方 (2) の方法は辞書構築のコストは不要であるが、通常の固有表現抽出をそのまま使えないため学習データの構築にコストがかかる。

そこで本研究では (1) の辞書ベース手法に対して、単語の意味表現を獲得できる Skip-gram [10], [11] を利用して料理名を表す単語を拡張することを提案する。提案手法では、あらかじめ一定量の料理名辞書を構築し、料理名の分散表現のベクトルから平均的な料理名ベクトル（平均料理名ベクトル）を生成する。料理名は名詞のサブセットであると考え、SNS の投稿を形態素解析し名詞を抽出後、平均料理名ベクトルとのコサイン類似度が高い単語を料理名として抽出する。料理名辞書とのキーワードマッチにより判定する方法と機械学習を用いた固有表現抽出により料理名を抽出する方法をベースラインとして評価を行った。提案手法の有効性を評価するため、辞書ベース手法、固有表現ベース手法をベースラインとして検証を行った。その結果、辞書ベース手法や固有表現抽出を利用する手法より高性能に料理名抽出できることを確認した。

本研究の貢献は以下の通りである。

- Skip-gram を利用して特定ドメインの名詞を半自動で抽出する手法を提案した。
- 辞書ベース手法および機械学習ベース手法をベースラインとして提案手法の有効性を定量的に確認した。

(注1): <https://woman.mynavi.jp/article/140227-44/>

• 提案手法により SNS から抽出した料理名データが食事行動の分析に利用できる可能性を定性的に示した。

2. 関連研究

2.1 料理名を対象とした固有表現抽出

人名や地名、組織名などの特定の名詞句を抽出するタスクは固有表現抽出と呼ばれ多くの研究が行われている [5]。Mori らはレシピ用語に特化した固有表現抽出に関する研究を行っている [6], [7]。この研究では料理や食材、調理道具、調理の動作など 8 種類のレシピ固有表現を定義し抽出を行っている。レシピテキストを対象に評価を行い、F 値約 0.85 という性能で抽出できることを報告している [6] 本研究では SNS の投稿を対象としており、レシピテキストより崩れた日本語が使われる可能性があり、既存手法がどの程度の性能で利用できるか明らかではない。

2.2 Word2vec (分散表現) による類義語抽出

Mikolov らは Continuous Bag-of-Words (CBOW) および Skip-gram という単語のベクトル表現手法を提案した [10], [11]。この手法では従来よく用いられていた Bag-of-Words では単語を One-hot ベクトルで表現してところを、密なベクトルで表現することにより単語の意味を柔軟に表現可能となった。これらの手法を word2vec というツールとして公開している (注2)。CBOW や Skip-gram を利用することにより、意味が近い単語はベクトルの値として近くなるため類義語判定に応用されている [8], [9], [12]。

2.3 分散表現の単語以外への適用

単語を One-hot ベクトルから密ベクトルに変換するという考え方を商品など他のデータに応用する研究が行われている。Barkan ら [16] は情報推薦における各アイテムを単語と見なし、Skip-gram を適用することで各アイテムのベクトル表現を取得し推薦を行った。また、1 つのレシピに出現する材料を Skip-gram における周辺語とみなし食材や料理をベクトル表現するという研究も行われている [14], [15]。これらの研究では料理名や食材はあらかじめ抽出されており、その後のベクトル化を行っている。

3. 提案手法

3.1 Skip-gram

Skip-gram [10], [11] の入力はコーパス、出力はコーパスに出現する各単語の分散表現 (単語ベクトル) である。入力となるコーパスは長さ T の単語列と見なせる。単語列の位置 t の単語 w_t に対して前後の単語を文脈とする。このとき、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学習する。

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

3.2 提案手法

提案手法では、あらかじめシードとなる一定量の料理名辞書

Algorithm 1 Procedure of Extracting Meal Name from Tweet

Input: 大規模コーパス L , 平均料理名ベクトル v_{avg} , SNS の投稿 t

Output: 抽出された料理名リスト e

```
1: Word List  $w \leftarrow \text{morphological\_analysis}(t)$ 
2: extract NOUN word from  $w$ 
3: for  $i = 0$  to  $\text{length}(w)$  do
4:   Get word vector from  $L$  for  $w[i]$  (i.e.  $L[w[i]]$ )
5:   Calculate cosine similarity  $sim = \text{cos\_sim}(v_{avg}, L[w[i]])$ 
6:   if  $sim \geq \text{threshold}$  then
7:     append  $e \leftarrow w[i]$ 
8:   end if
9: end for
10: return  $e$ 
```

と汎用的な大規模コーパスから、skip-gram により料理名辞書の単語も含む全単語のベクトル表現を学習しておく。また、料理名辞書に出現する単語ベクトルの平均ベクトルを計算する。これを平均料理名ベクトル v_{avg} とし、以下の通りコーパス L において料理名辞書に出現する単語の平均ベクトルとして計算する。

$$v_{avg} = \frac{1}{|M|} \sum_i L[m[i]]$$

ここで、 $L[w]$ はコーパス L で学習した単語 w の単語ベクトル、 m は料理名辞書に含まれる料理名、 M は料理名辞書 ($m \in M$) とする。その後、SNS の各投稿文章を Algorithm 1 に示す手順で処理を行い料理名を抽出する。Algorithm 1 では料理名は名詞のサブセットであると考え、以下の手順で処理を行う。

- SNS の投稿を形態素解析 (1 行目)
- 名詞を抽出 (2 行目)
- SNS に含まれる名詞と平均料理名ベクトルとのコサイン類似度を計算 (5 行目)
- 類似度が閾値以上であれば該当の単語を料理名として抽出 (6~7 行目)

料理名は「カレー」「うどん」などの一般名詞だけでなく、名詞句で構成される料理も多く存在する。本稿では料理名抽出のファーストステップとして 1 つの単語で表現される料理名を拡張することを主題とし、名詞句として抽出する処理は対象外とする。

4. 評価実験

4.1 実験データ

Skip-gram の学習コーパスには日本語 Wikipedia の全記事のダンプを利用した。Skip-gram の実装には word2vec^(注3) を利用した。ベクトルの次元数は 200、文脈を見るウィンドウサイズは 5 とし、出現回数が 5 回未満の単語は削除した。

料理名辞書は NHK の「きょうの料理」で紹介されたレシピ

(注2): <https://github.com/svn2github/word2vec>

(注3): <https://github.com/svn2github/word2vec>

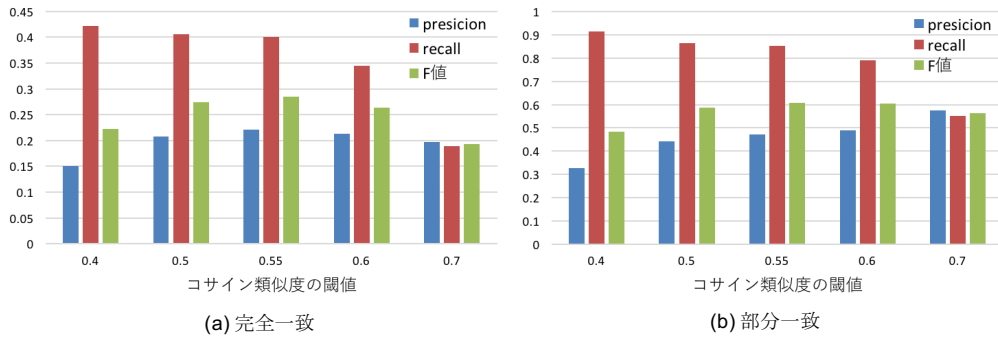


図1 性能評価結果

を投稿しているNHKの公式Twitterアカウント^(注4)の投稿から取得した。具体的にはこのアカウントでは【料理名】本文という形式で投稿を行っているため【料理名】の部分で正規表現により抽出した。一部の投稿では料理名以外も【】の部分に記載されているため、正規表現による抽出後に目視でクレンジングを行った。料理名の抽出には2015年1月1日から2017年9月20日までの11,780件の投稿を利用した。これらの投稿から約6,000件の料理名を抽出し、料理名辞書とした。

本研究ではSNSとしてmiilという料理写真を共有するSNSを対象とした^(注5)。miilへの投稿は、miilからTwitterに連携して投稿されているツイートをTwitter API^(注6)により取得した。miilへの投稿を抽出するためハッシュタグに#miilを含む投稿を抽出した。また、本研究では献立推薦を目的の1つとしているため、食事をした場所が自宅の投稿を抽出した。miilから連携されるTwitterの投稿では「at 自宅」「at Home」のように場所が明示されているため、正規表現によりこれらの文字列を含む投稿を抽出した。対象期間は2016年1月1日から2016年12月31日とした。その結果、投稿数61,702件、ユニークユーザ数2,482ユーザの投稿を収集した。

4.2 ベースライン手法

ベースライン手法として料理名辞書との単語の一致により判定する辞書マッチ手法(ベースライン1, B1)と、機械学習による固有表現抽出で料理名を抽出するSasadaらの手法[6](ベースライン2, B2)の2つと比較を行った。

ベースライン1: 辞書マッチ

辞書マッチ手法では、あらかじめ形態素解析の辞書に料理名を形態素として登録しておく。そして、判定対象の文章を形態素解析し、形態素と料理名辞書の完全一致により判定する。

ベースライン2: 機械学習による固有表現抽出 本研究では、Sasadaらの機械学習によるレシピ用語抽出[6]を機械学習による固有表現のベースラインとした。

4.3 評価指標と評価データ

定量的な評価指標には再現率(recall)、適合率(precision)およびF値(F-measure)を利用する。再現率は人手で抽出した料理名に対して、抽出できた料理名数の割合、適合率は抽出

表1 評価データの例1

本文	お肉屋さんのフライ盛り合わせとブロッコリーとミックスビーンズのサラダ at 自宅 https://t.co/WVWgYY5ypV #miil
料理名	フライ盛り合わせ、ブロッコリーとミックスビーンズのサラダ

した料理名のうち、人手で作成した正解データに含まれる割合である。F値は再現率と適合率の調和平均によって求められる。計算方法は以下の通り。

$$F \text{ 値} = \frac{2 \times \text{再現率} \times \text{適合率}}{\text{再現率} + \text{適合率}}$$

なお、正解の判定では、料理名が完全一致した場合と、部分一致している場合があるため、完全一致と部分一致のそれぞれに対して上記の指標を算出した。

評価データには、前述のmiilへの投稿からランダムサンプリングした235件の投稿から人手により料理名を抽出し424個の正解データを作成した。評価データの例を表1に示す。正解データには名詞と名詞句の両方が含まれる。

4.4 実験環境

計算機環境にはLinux Ubuntu 16.04, 実装はPython 3.6.2, 形態素解析にはMeCab^(注7)を利用し、辞書データにはmecab-ipadic-NEologd [17]^(注8)と料理名辞書、ベースライン2である文献[6]の実装にはSasadaらの実装^(注9)を利用した。形態素解析にはKyTea^(注10)を利用し、レシピ用語抽出の学習コーパスとしてはMoriらのデータ[7]を利用した。

4.5 実験結果

本節では、以下3点の結果を示す。

(1) 提案手法では平均料理名ベクトルと各単語のコサイン類似度が閾値以上の場合に料理名であると判定するため閾値の設定を変えることによる性能変化の結果(パラメータチューニング)

(2) 定量評価としてベースライン手法と提案手法の性能比較結果(ベースライン手法との比較)

(注4): https://twitter.com/m_kyounoryouri

(注5): <http://miil.me/>

(注6): <https://dev.twitter.com/rest/public>

(注7): <http://taku910.github.io/mecab/>

(注8): <https://github.com/neologd/mecab-ipadic-neologd>

(注9): <http://www.ar.media.kyoto-u.ac.jp/how-to/recipe-NLP/>

(注10): <http://www.phontron.com/kytea/>

表 2 ベースライン手法との比較結果

手法	完全一致			部分一致		
	適合率	再現率	F 値	適合率	再現率	F 値
辞書マッチ (B1)	0.642	0.165	0.263	0.945	0.243	0.386
文献 [6] (B2)	0.160	0.283	0.204	0.383	0.679	0.490
提案手法	0.222	0.401	0.286	0.472	0.854	0.608

(3) 定性評価として抽出結果の実例から提案手法の性質を考察

4.5.1 パラメータチューニング

パラメータチューニングの結果を図 1 に示す。図 1 (a) は完全一致で各指標を算出した結果、図 1 (b) は部分一致で各指標を算出した結果を示す。全体的に閾値が高くなるほど適合率は高くなり再現率は低下するという一般的に想定される傾向となった。完全一致、部分一致ともにコサイン類似度の閾値を 0.55 とした場合が F 値がもっとも高くなった。そのため、次のベースライン手法との比較では閾値を 0.55 として比較を行う。

4.5.2 ベースライン手法との比較

ベースライン手法と提案手法の性能の評価結果を表 2 に示す。太字は各指標の中でもっとも数値が高いことを示す。完全一致、部分一致ともに適合率は辞書マッチがもっとも高く、再現率および F 値は提案手法がもっとも高い結果であった。辞書マッチでは辞書に料理名が正確に登録されていれば抽出も正確に行えるため適合率が高いと考えられる。機械学習による固有表現抽出を使った文献 [6] の手法は他の手法より性能が高くない結果であった。これは、文献 [6] ではレシピのテキストを対象としているが、本研究では SNS の投稿を対象としており料理名の表記方法に違いがあることが原因と考えられる。具体的には本研究で対象としてテキストでは表 1, 3 に示すように料理名が助詞や句読点などで区切られて連続して出現することがある。しかし、レシピのテキストではそのような表記がなされることは稀であるため学習データに含まれないと考えられる。そのため、機械学習の学習データを増やすことで性能が向上すると考えられる。

本研究では献立推薦への応用を考えており、料理名抽出の結果を直接ユーザに提示することが想定される。そのため、適合率が高いことが望ましいが、その観点では辞書マッチがもっとも適している。一方、推薦の際は抽出された料理名はユーザに提示せず、推薦結果のみを提示すれば適合率は低くても再現率が高い手法が適している可能性がある。そのため、今後推薦までを含めて精度評価を行う必要がある。

4.5.3 抽出結果の実例

表 3 に評価データにおける各手法の抽出結果の例を示す。辞書マッチ (B1) では 1 件しか抽出が行われていない。これは例えば「ひじき」という単語は、料理名辞書には「ひじきの煮物」など複数の料理名の一部としては登録されているが、「ひじき」のみでは登録がないことが原因である。また、文献 [6] (B2) では、「今週」「cm」など料理とは無関係の単語も抽出されている。英数字に関しては、正規表現により削除して判定するこ

表 3 料理名抽出結果の例

本文	今週の常備菜。もやしナムル、冬瓜のそぼろ煮、ひじきのミックスビーンズサラダ、ガーリックブロッコリー、紅茶鶏。 at 自宅 https://t.co/cm88jBKmsQ #miil
正解データ	もやしナムル、冬瓜のそぼろ煮、ひじきのミックスビーンズサラダ、ガーリックブロッコリー、紅茶鶏
辞書マッチ (B1)	ナムル
文献 [6] (B2)	今週、ナムル、冬瓜、ひじき、ミックスビーンズサラダ、ガーリックブロッコリー、紅茶、cm
提案手法	常備菜、もやし、ナムル、冬瓜、そぼろ、ひじき、サラダ、ガーリック、ブロッコリー、紅茶

とで精度向上できると考えられる。「ミックスビーンズサラダ」「ガーリックブロッコリー」など名詞が連続する料理名では機械学習が有効に機能していると考えられる。一方、提案手法では料理と関連する単語が抽出されている。しかしながら、「ガーリック」「ブロッコリー」のように提案手法では一つの料理名として抽出されるべき単語が分割されて抽出される。このような単語の結合に関しては機械学習の手法との組み合わせにより精度向上できると考えられる。

4.6 レシピ推薦への活用に向けた食事傾向の分析

本節では提案手法により SNS の投稿から抽出した料理名がレシピ推薦に活用できるか分析を行う。表 4 に月別に抽出された料理名に対して TFIDF を計算した結果を示す。ここでは、各料理名の出現回数を TF、文書単位を月別とし IDF は出現した月数の逆数とした。1 月にお節料理や七草粥、2 月の恵方巻き、7 月のうなぎなど各季節の代表的と考えられる料理が SNS への投稿から抽出できている。表 4 の結果では、一般的に知られている傾向が抽出されているため、今後曜日ごとの傾向と組み合わせるなど、時間的なコンテキストを踏まえたより深い分析を行い推薦への応用を検討する必要がある。

5. まとめと今後の課題

本稿ではソーシャルメディアへの投稿から低コストに食事傾向を観測することを目的としてソーシャルメディアの投稿から料理名を抽出する手法を提案した。提案手法では、Skip-gram を利用して学習した単語の分散表現とシードとなる一定量の料理名辞書から平均的な料理名を表すベクトル (平均料理名ベクトル) を生成し、平均料理名ベクトルとの類似度を計算することで料理名を判定した。料理名辞書とのキーワードマッチにより判定する方法と機械学習を用いた固有表現抽出により料理名を抽出する方法をベースラインとして評価を行った。その結果、ベースライン手法に対して、完全一致では 2.3 %、部分一致では 11.8 % の精度向上を確認した。

今後の課題として、現在の精度ではレシピ推薦への応用には不十分であるため精度向上が考えられる。精度向上のため平均料理名ベクトルとその他の料理名の分散表現空間での可視化や誤り事例の分析を行う。平均料理名ベクトルではなく料理名辞書の各単語と類似度が高い単語を拡張する手法との比較や名詞

表 4 月別に抽出された料理名に対して TFIDF を適用した結果

順位	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
1	雑煮	恵方巻き	もち麦	メンマ	フレーク	梅酒	うなぎ	ライム	すだち	松茸	安納芋	シュトーレン
2	七草粥	恵方巻	葉っぱ	沢庵	ユッケ	ミント	ぶっかけ	ゴーヤー	ゴハン	パンブキン	クレソン	年越しそば
3	酒粕	チューハイ	桜餅	かき菜	ソーダ	青じそ	土用	冷や汁	生野菜	ピオーネ	ボジョレー	ローストチキン
4	七草	カツ丼	酢味噌	ちゃんぽん	タルタルソース	梅干	蒲焼き	カレーパン	油淋鶏	茶漬け	スイートポテト	クリスマスケーキ
5	ポタージュ	モツ	ツナ缶	ひとくち	お揚げ	牛タン	鰻井	冷製パスタ	ほうじ茶	カップヌードル	メロンパン	フライドチキン
6	お節	ホワイトチョコ	ししゃも	土佐煮	バルメザンチーズ	モヒート	カオマンガイ	棒棒鶏	ガーリックトースト	けんちん汁	スンドゥブ	クリスマスディナー
7	数の子	ヒラメ	若布	ヒジキ	ヤングコーン	カブレゼ	缶詰	和牛	片栗粉	シナモン	銀杏	カステラ
8	栗きんとん	巻き寿司	塩味	ペビーリーフ	絹さや	寒天	ゴマだれ	メカジキ	チンゲン菜	ぶどう	ハッシュドビーフ	カリフラワー
9	いかにんじん	カカオ	スパイシー	クロワッサン	スコーン	佐藤錦	アオサ	ピッツァ	アーモンド	スタチ	麵つゆ	豆乳鍋
10	鶏レバー	太巻き	山かけ	若竹煮	パンケーキ	煮豆	うなぎ	空芯菜	ゴーヤー	甘露煮	いわし	ショートケーキ

句としての抽出，評価データの拡充なども今後の課題である．
 また，提案手法は一般化して考えると特定ドメインの名詞を抽出する手法と見なせるため，料理名以外への適用を検討する．
 加えて，SNS の投稿から抽出した料理名に基づいて献立やレシピの推薦を行うことも今後の課題である．

文 献

[1] Furst, T., Connors, M., Bisogni, C. A., Sobal, J., and Falk, L. W. "Food choice: a conceptual model of the process" *Appetite*, 26(3), 247-266, 1996.

[2] Giskes, Katrina and Kamphuis, Carlijn BM and van Lenthe, Frank J and Kremers, Stef and Droomers, Mariel and Brug, Johannes, "A systematic review of associations between environmental factors, energy and fat intakes among adults: is there evidence for environments that encourage obesogenic dietary intakes?" *Journal of Public health nutrition*, Vol.10, No.10, pp. 1005-1017, 2007.

[3] Haines, P. S., Hama, M. Y., Guilkey, D. K., and Popkin, B. M. "Weekend eating in the United States is linked with greater energy, fat, and alcohol intake", *Obesity*, 11(8), 945-949, 2013.

[4] Shahar, D. R., Yerushalmi, N., Lubin, F., Froom, P., Shahar, A., and Kristal-Boneh, E. "Seasonal variations in dietary intake affect the consistency of dietary assessment" *European journal of epidemiology*, 17(2), 129-133, 2001.

[5] 山田寛康, 工藤拓, and 松本裕治. "Support Vector Machine を用いた日本語固有表現抽出" *情報処理学会論文誌*, Vol.43, No.1, pp.44-53, 2002.

[6] Sasada, Tetsuro and Mori, Shinsuke and Kawahara, Tatsuya and Yamakata, Yoko, "Named entity recognizer trainable from partially annotated data," *International Conference of the Pacific Association for Computational Linguistics*, (Springer), pp. 148-160, 2015.

[7] Mori, Shinsuke and Maeta, Hirokuni and Yamakata, Yoko and Sasada, Tetsuro, "Flow Graph Corpus from Recipe Texts," *International Conference of the Pacific Association for Computational Linguistics*, (Springer), pp. 148-160, 2015.

[8] 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.

[9] 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.

[10] Mikolov, Tomas and Chen, Kai and Corrado, Greg and Dean, Jeffrey, "Efficient estimation of word representations in vector space," *CoRR*, <http://arxiv.org/abs/1301.3781> (2013).

[11] Mikolov, Tomas and Sutskever, Ilya and Chen, Kai and Corrado, Greg S and Dean, Jeff, "Distributed Representations of Words and Phrases and their Compositionality," *Advances in Neural Information Processing Systems 26*, NIPS 2013, pp.3111-3119, 2013.

[12] Hu, Bing and Liu, Bin and Gong, Neil Zhenqiang and Kong, Deguang and Jin, Hongxia, "Protecting Your Children from Inappropriate Content in Mobile Apps: An Automatic Maturity Rating Framework," *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, pp.1111-1120, 2015.

[13] MeCab, <http://taku910.github.io/mecab/>

[14] food2vec - Augmented cooking with machine intelligence, <https://jaan.io/food2vec-augmented-cooking-machine-intelligence/>

[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] Barkan, Oren, and Noam Koenigstein, "Item2vec: neural item embedding for collaborative filtering." *IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2016.

[17] 佐藤敏紀, 橋本泰一, 奥村学, "単語分かち書き用辞書生成システム NEologd の運用 文書分類を例にして '情報処理学会自然言語処理研究会研究報告, NL-229-15, 2016.