

レビューサイトの差異に基づくラベル伝播を用いた ユーザレビュー分類手法

浅野 祥汰[†] 北山 大輔[†]

[†] 工学院大学 情報学部 〒163-8677 東京都新宿区西新宿1丁目24番地2号
E-mail: tj113005@ns.kogakuin.ac.jp kitayama@cc.kogakuin.ac.jp

あらまし 近年、ユーザは飲食店を決定する際に、飲食店検索サイトを用いる機会が増えている。そこで得られる情報の中でも、レビューは非常に重要な役割を持っている。しかし、ユーザが求めている情報をレビューから見つけ出すのは容易ではない。本研究では、Doc2Vecを用いて、レビューをメニュー単品、メニュー全体、店舗の雰囲気、の3つの軸で分析し、ユーザに提示することで飲食店の選択を支援するシステムの提案をする。

キーワード ユーザレビュー, 飲食店検索サイト, Doc2Vec

1. はじめに

近年、ユーザは飲食店を決める際に飲食店検索サイトを用いる機会が多い。株式会社ジャストシステムの調査結果[1]によると、飲食店口コミサイトを使ったことがあるユーザは95.1%と、非常に多くのユーザが飲食店検索サイトの利用経験がある。また、39.4%のユーザが「口コミの質を重視している」と答えている。また、NTTコムリサーチの調査結果[2]によると、インターネットサイト内にあるレビューを参考にして飲食店を選んだことのあるユーザは91.2%もいる。またレビューの中で重視している情報としては、味が75.4%、店の雰囲気が73.6%であり、レビューから味や店の雰囲気が重視されていることがわかる。一方で、インターネットサイト内のレビューを参考にしていないユーザもいる。それらのユーザが参考にしていない理由の中で、一番多い理由として「情報を探すことが手間」があげられた。

これらの調査結果から、レビューは飲食店を決める際に重要視されていると考えられる。飲食店レビューサイトの中にはぐるなびのように、メニューに対して直接レビューを書くことができるサイトもある。このようなサイトでは、味に対するレビューを効率よく取得可能であると考えられる。しかしながら、レビューサイト最大手の食べログに代表される一般的なレビューサイトでは店舗に対してレビューが書かれるため、メニューや雰囲気など様々な対象へのレビューが混在している。また、レビュー数も新宿エリアを例にとると、ぐるなびが約8千件、食べログが約41万件と、混在したレビューの方が非常に多くユーザの求めているレビューを見つけるのは困難である。

これらのことより我々は、ぐるなびや食べログのようにレビュー対象が異なるレビューサイトに着目する。すなわち、ぐるなびのようにレビュー対象が明確なレビューを手がかりにすることで、ユーザの求めているレビューを提示することが可能になると考えた。上記より、レビューでは「味」と「店の雰囲気」が特に重視されていることから、本研究ではレビューをメニュー単品、メニュー全体、雰囲気の3つに分類する手法を提

案する。このことより、ユーザが目的のレビューを取得しやすくなる。

2. 関連研究

村野[3]は、文型パターンを用いてレビュー中から評価文の抽出を行った。しかし、文章内で省略されている単語も多く、評価文が正しく抽出されないことがあった。本研究では、類似度を用いて文の分類を行うため、単語の省略により抽出精度が落ちるといったことはない。藤村[4]は、機械学習C4.5を用いて、評判を抽出するルールを発見し、決定木を分析することで意思決定支援へ応用する手法を提案した。この藤村らの研究では、C4.5で属性数を増やしすぎると過学習によって精度が落ちる[5]という課題もある。立澤ら[6]は、頻度情報を抑えたTF-IDFを用いて評価文の抽出手法の提案を行った。また、矢野ら[7]は、Web上の飲食店の店舗情報を取得し、その店舗情報を形態素解析することで評価文を抽出する手法の提案を行った。しかし、どちらの手法も雰囲気に関する情報はノイズとして扱われ、メニューに対する評価文のみの抽出である。本研究では、メニューのみでなく雰囲気情報も抽出可能となっている。

3. レビューサイトの差異

本研究では、食べログとぐるなびのレビューを用いる。そこで、両サイトに登録されているサイト数、レビュー数、レビューに含まれている特徴を表1に示す。なお、ここで示すデータはすべて新宿エリアのデータである。

表1からわかる通り、食べログの登録飲食店数はぐるなびの約3倍、登録レビュー数はぐるなびの約20倍もある。また、レビューに含まれる情報も、食べログは3種類、ぐるなびは1種類と食べログのレビューは非常に煩雑である。ここで、メニューの感想は、「ポテトサラダがおいしかった」のように特定のメニュー単品への感想もあれば、「全体的に味のバランスがよかった」のように店で食べた食事全体への感想もある。そのため、メニューの感想を「単品」と「全体」の2種類に分けている。

表 1 食ベログとぐるなびの特徴

	食ベログ	ぐるなび
登録飲食店数	5418	1811
登録レビュー数	151594	8250
レビュー文字数平均値	535.2	60.6
レビュー文字数中央値	406	39
レビューに含まれる情報	メニュー単品の感想 メニュー全体の感想 飲食店の雰囲気	メニュー単品の感想



図 1 ぐるなびレビューの例

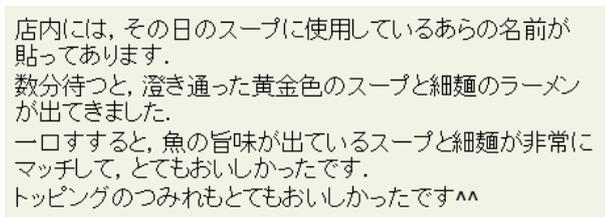


図 2 食ベログレビューの例

図 1 はぐるなびのレビューであるが、単品メニューに対してレビューが投稿されていることがわかる。図 2 は食ベログのレビューであるが、店の雰囲気や提供されたメニューなど様々なことが書いてあることがわかる。

4. ラベル伝播を用いたレビュー分類手法

4.1 概要

前章では、食ベログのレビューにはメニュー単品の感想、メニュー全体の感想、飲食店の雰囲気の 3 種類の情報が、ぐるなびのレビューにはメニュー単品の感想のみが書かれていることがわかった。ここで、ぐるなびのレビューには「メニュー単品」のラベルがついていると仮定する。この「メニュー単品」のラベルを文書類似度を使って、図 3 のようにして食ベログのレビューに伝播させる。これが強く伝播する（図 3 のレビュー間を実線でつないでいる）食ベログレビューには「メニュー単品」のラベルを、中間（図 3 のレビュー間を点線でつないでいる）なレビューには「メニュー全体」のラベルを、伝播しない（図 3 のレビュー間をつなぐ線がない）レビューには「雰囲気」のラベルをつける。各ラベルの定義は以下の通りである。

- メニュー単品：そのお店のあるメニューに対する感想であると特定できるレビュー
 - － コース料理等はメニュー単品に含まれる
 - － メニュー名が書かれていなくても、何について書かれて

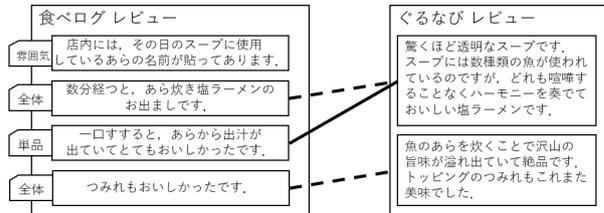


図 3 分類の概念

表 2 分かち書きと基本形の例

分かち書き	基本形
ラーメン	ラーメン
は	は
濃厚な	濃厚だ
スープ	スープ
と	と
太麺	太麺
が	が
絡み合っ	絡み合う
て	て
美味しっか	美味しい
た	た
です	です

いるか判断できる場合はメニュー単品に含まれる

- メニュー全体：そのお店のメニューに対する感想であるが、メニューの特定はできないレビュー
 - － メニュー全体は複数のメニューに対する感想も含む
- 雰囲気：内装や外装の情報、立地に関する情報、混雑情報などの飲食店の雰囲気に関する情報

ここで、食ベログのレビューは 1 つのレビューに 3 種類の情報が含まれているので、文単位に分割し、その文それぞれにラベル付けを行うものとする。この文単位に分割したレビューを「レビュー文」とする。また、ラベルがどれだけ強く伝播するかを示す数値を伝播度とする。例えば、伝播度が高い場合には、よりメニュー単品に近いレビュー文が書かれていて、逆に伝播度が低い場合には、より雰囲気に近いレビュー文が書かれていることになる。

4.2 レビューの分かち書き

本手法では、レビュー間の類似度を求めるために、レビュー文の分かち書きを行う。分かち書きは MeCab [8] を用いて行う。食ベログのレビューには、レビューを句点ごとに分割した文章に対して分かち書きを行い、ぐるなびのレビューには提供されたメニューにつけられたレビューそのものに対して分かち書きを行う。また、分かち書きによって出現する語句を可能な限り統一するために、分かち書きをした語句はすべて基本形に直す。分かち書きした語句を基本形に直す例を「ラーメンは濃厚なスープと太麺が絡み合っって美味しかったです」の文章を用いて、表 2 に示す。

4.3 Doc2Vec を用いたレビュー間のラベル伝播

レビュー間の類似度を求める。本手法では、Doc2Vec [9] を用いて類似度を求める。Doc2Vec とは、Word2Vec [10] という



図 4 類似度算出方法の例

単語にベクトルを割り当てて単語間の類似度を求める技術があるが、それを応用させて文書にベクトルを割り当てることで、文書間の類似度を求めることができる技術である。この技術をレビューに用いることで、レビュー同士で類似度を求めることができる。学習モデルには、新宿エリアの食ベログとぐるなびのレビューすべてを用いる。ここで、ある飲食店に投稿された食ベログのレビュー文 r_t とぐるなびのレビュー r_g の間のラベルの伝播度 *propagation* を以下の式で求める。 R_g とは、ある飲食店と同エリアかつ同ジャンルのぐるなびのレビューの集合である。

$$propagation(r_t) = \max_{r_g \in R_g} (sim(r_t, r_g)) \quad (1)$$

これを図 4 を使った例を用いて説明をする。ある飲食店を、「食ベログラーメン店 TA」とし、この飲食店が持つレビューが「レビュー文 TA1.1~TA1.3」である。この「食ベログラーメン店 TA」と同じエリアにあり、かつ同じジャンルの飲食店が「ぐるなびラーメン店 GA」「ぐるなびラーメン店 GB」であり、これらの飲食店が持つレビューが「レビュー GA1~GA2」「レビュー GB1~GB2」である。この中でまず、食ベログのレビュー「レビュー文 TA1.1」とぐるなびのレビュー「レビュー GA1~GB2」の間で類似度を求める。図 3 の「レビュー文 TA1.1」と線で結ばれているぐるなびのレビューに対して類似度を求める。ここで求められる類似度が表 4 の線上にある数値である。これらの数値の中では、食ベログのレビュー「レビュー文 TA1.1」とぐるなびのレビュー「レビュー GA1」の間の類似度である 0.8 が最も高い数値（類似度）であるので、この「0.8」を食ベログのレビュー「レビュー文 TA1.1」のラベル伝播として付与する。また、図 1 では、食ベログのレビュー「レビュー文 TA1.1」しか類似度を求めていないが、ユーザが決定した飲食店の食ベログのレビューすべて（図 4 であれば「レビュー文 TA1.1~TA1.3」）に、この手順で類似度を求める。

4.4 伝播度を用いたレビュー分類

4.3 節で求められる伝播度を用いてレビュー文の分類を行うが、「メニュー全体」のレビュー文に関して伝播度の範囲を決定する必要がある。決定のために行った予備実験について説明する。ラーメン、カフェ、居酒屋、カレーライス、お好み焼き、回転寿司の 6 ジャンル、各最大 250 レビュー文のラベルの伝播度を求める。このレビュー文は各店舗からランダムに選択し、最大 50 レビュー文を抽出した。レビュー文に対し、5 人の被験者が 4.1 節に示した各ラベルの定義にしたがって分類を行い、最も多くの被験者がつけたラベルを正解ラベルとした。被験者は定義した 3 つのラベル以外であると判断した場合は「その他」

表 3 各ラベルの伝播度の中央値

ジャンル名	单品	全体	雰囲気
ラーメン	0.420	0.384	0.312
カフェ	0.410	0.386	0.330
居酒屋	0.414	0.388	0.338
カレーライス	0.385	0.350	0.296
お好み焼き	0.334	0.306	0.275
回転寿司	0.333	0.346	0.244

表 4 しきい値の設定

	しきい値上限	しきい値下限
ラーメン	0.402	0.348
カフェ	0.398	0.358
居酒屋	0.401	0.363
カレーライス	0.367	0.323
お好み焼き	0.320	0.290
回転寿司	0.340	0.296

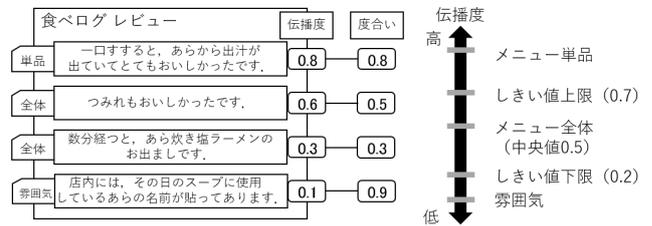


図 5 ランク付けの例

をつけている。ジャンル毎の分類したレビュー文の各ラベルの伝播度の中央値は表 3 の通りである。

次に「メニュー全体」のレビュー文を抽出する伝播度のしきい値をそれぞれ設定する。各ジャンルの「メニュー单品」と「メニュー全体」の中央値の平均値をしきい値の上限、「メニュー全体」と「雰囲気」の中央値の平均値をしきい値の下限とすると、表 4 のようにしきい値が設定できる。

次に、伝播度を用いて、各ラベルの度合いを求める。メニュー单品度合い、メニュー全体度合い、雰囲気度合いはそれぞれ以下の式で求める。

$$\text{单品度} = x \quad (2)$$

$$\text{全体度} = \begin{cases} \frac{\alpha - x}{\alpha - m} & (\alpha \geq x > m) \\ 1 - \frac{m - x}{m - \beta} & (m \geq x \geq \beta) \end{cases} \quad (3)$$

$$\text{雰囲気度} = 1 - x \quad (4)$$

x はそのレビューの伝播度、 α はしきい値の上限、 β はしきい値の下限、 m はメニュー全体の伝播度の中央値である。それぞれの度合いを求めることで、ラベルとしてのランク付けもできるため、このランクが高い順にレビュー文の表示を行う。また、ラベル自体はしきい値で決めるものとする。例を図 5 に示す。この場合では、伝播度が 0.6 のレビュー文の方が、伝播度が 0.3 のレビュー文より「メニュー全体」の度合いが高いため、伝播度が 0.6 のレビュー文が優先的に表示される。

表 5 分類精度 (提案手法 : 3 分類)

ジャンル	分類ラベル	適合率	再現率	F 値
ラーメン	メニュー単品	0.667	0.467	0.549
	メニュー全体	0.121	0.226	0.157
	雰囲気	0.449	0.716	0.552
カフェ	メニュー単品	0.691	0.507	0.585
	メニュー全体	0.125	0.304	0.177
	雰囲気	0.432	0.635	0.514
居酒屋	メニュー単品	0.571	0.431	0.491
	メニュー全体	0.196	0.278	0.230
	雰囲気	0.514	0.789	0.623
カレーライス	メニュー単品	0.536	0.481	0.507
	メニュー全体	0.244	0.204	0.222
	雰囲気	0.284	0.673	0.400
お好み焼き	メニュー単品	0.417	0.625	0.500
	メニュー全体	0.135	0.128	0.132
	雰囲気	0.412	0.600	0.488
回転寿司	メニュー単品	0.313	0.645	0.421
	メニュー全体	0.269	0.135	0.174
	雰囲気	0.405	0.773	0.531
6 ジャンル平均	メニュー単品	0.532	0.526	0.509
	メニュー全体	0.182	0.212	0.183
	雰囲気	0.416	0.698	0.518

表 6 分類精度 (ランダムフォレスト : 3 分類)

ジャンル	分類ラベル	適合率	再現率	F 値
ラーメン	メニュー単品	0.476	0.556	0.513
	メニュー全体	0.273	0.194	0.226
	雰囲気	0.485	0.731	0.583
カフェ	メニュー単品	0.440	0.440	0.440
	メニュー全体	0.116	0.348	0.174
	雰囲気	0.620	0.671	0.644
居酒屋	メニュー単品	0.347	0.585	0.481
	メニュー全体	0.435	0.785	0.339
	雰囲気	0.647	0.489	0.557
カレーライス	メニュー単品	0.427	0.961	0.591
	メニュー全体	0.692	0.164	0.265
	雰囲気	0.543	0.510	0.481
お好み焼き	メニュー単品	0.392	0.797	0.526
	メニュー全体	0.333	0.077	0.125
	雰囲気	0.552	0.757	0.639
回転寿司	メニュー単品	0.303	0.323	0.313
	メニュー全体	0.465	0.635	0.537
	雰囲気	0.375	0.636	0.472
6 ジャンル平均	メニュー単品	0.398	0.643	0.477
	メニュー全体	0.386	0.282	0.278
	雰囲気	0.537	0.632	0.570

5. 評価実験

提案した分類手法の精度を評価するため、性能評価実験を行った。実験では、実際のレビューデータから構築したデータセットを用いて精度の評価を行った。実験に使用したデータは、新宿の飲食店 (表 1) の通りである。

5.1 実験方法

実験は、レビュー数の多い順に飲食店 5 店舗を選択する。ただし、予備実験で正しい値を求めるために使用した店舗は、評価実験では使用していない。各店舗から最大 50 レビュー文をランダムに選択し、被験者がレビュー文にラベル付けをする。これを正解ラベルとする。またレビュー文の伝播度を計算・分類をし、被験者が付けた正解ラベルと本手法での分類から適合率、再現率、F 値を求める。レビュー文のラベル付けには、各レビュー文に 5 人の被験者がラベル付けをし、一番多く付けられたラベルをそのレビュー文のラベルとする。今回は、予備実験でも使用した、ラーメン、カフェ、居酒屋、カレーライス、お好み焼き、回転寿司の 6 ジャンルで評価実験を行う。また、比較手法としてランダムフォレストで分類を行った。ランダムフォレストは、予備実験で用いた店舗のレビュー文の形態素に分割したもので学習を行い、提案手法と同じデータを分類した。

5.2 実験結果

提案手法での分類結果を表 5、ランダムフォレストでの分類結果を表 6 にまとめる。

ジャンルによって分類精度に誤差があるものの、6 ジャンル平均としては、F 値が提案手法とランダムフォレストがほぼ同等の数値となった。そのため、提案手法は人手による正解ラベルを必要とせず、ぐるなびデータをを用いるだけで教師あり学習

表 7 適切に分類ができたレビュー文の例

ジャンル	表示ラベル	伝播度	レビュー
ラーメン	メニュー単品	0.530	三種類の生醤油をブレンドした店主拘りのかえしは鶏の清湯スープと見事に融合
居酒屋	メニュー全体	0.383	ほとんどのメニューは 300 円、500 円に設定されているので、ツマミ 2~3 品にビールをつけて千円ちょっと
カレーライス	雰囲気	0.288	店内はシックで小洒落た雰囲気
カフェ	雰囲気	0.293	でも、店内の音楽もいいし、本もたくさんあるので、ゆっくりできるときにお茶にきたいかな、と思った

による分類と同程度に分類可能であることを確認した。また、表 7 に、分類が適切に行えたレビュー文の例を示す。

図 6 は、ラーメンの伝播度のヒストグラムである。ヒストグラムからわかるように、「メニュー単品」と「メニュー全体」の伝播度の重なりは大きい。実際、ラーメンでは、「メニュー全体」と分類されたが、「メニュー単品」の正解ラベルが付けられたレビュー文の割合が 0.535 であった。このように、「メニュー単品」と「メニュー全体」の分類は難しいと考えた。そこで、改めて表 8 と表 9 に「メニュー単品」と「メニュー全体」をあわせた「メニュー」と「雰囲気」に分類した精度を示す。

「メニュー単品」と「メニュー全体」をあわせた「メニュー」

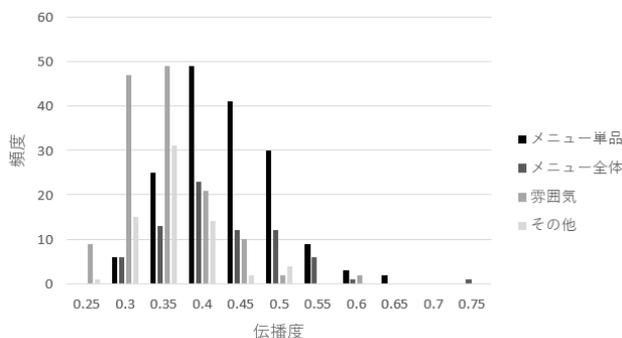


図 6 ジャンル：ラーメンの伝播度ヒストグラム

表 8 分類精度 (提案手法：2 分類)

ジャンル	分類ラベル	適合率	再現率	F 値
ラーメン	メニュー	0.760	0.760	0.760
	雰囲気	0.449	0.716	0.552
カフェ	メニュー	0.658	0.745	0.699
	雰囲気	0.432	0.635	0.514
居酒屋	メニュー	0.680	0.673	0.677
	雰囲気	0.514	0.789	0.623
カレーライス	メニュー	0.673	0.664	0.710
	雰囲気	0.284	0.673	0.400
お好み焼き	メニュー	0.602	0.777	0.678
	雰囲気	0.412	0.600	0.488
回転寿司	メニュー	0.656	0.711	0.682
	雰囲気	0.405	0.773	0.531
6 ジャンル平均	メニュー	0.686	0.722	0.701
	雰囲気	0.416	0.698	0.518

表 9 分類精度 (ランダムフォレスト：2 分類)

ジャンル	分類ラベル	適合率	再現率	F 値
ラーメン	メニュー	0.693	0.727	0.710
	雰囲気	0.485	0.731	0.583
カフェ	メニュー	0.569	0.837	0.678
	雰囲気	0.620	0.671	0.644
居酒屋	メニュー	0.518	0.871	0.649
	雰囲気	0.647	0.489	0.557
カレーライス	メニュー	0.685	0.962	0.800
	雰囲気	0.543	0.510	0.481
お好み焼き	メニュー	0.576	0.777	0.661
	雰囲気	0.552	0.757	0.639
回転寿司	メニュー	0.606	0.759	0.674
	雰囲気	0.375	0.636	0.472
6 ジャンル平均	メニュー	0.608	0.822	0.695
	雰囲気	0.537	0.632	0.570

と「雰囲気」の 2 つに分類をした結果、「メニュー」の分類精度は「メニュー単品」と「メニュー全体」の分類精度と比べ、精度の向上が見られた。

また、雰囲気について書かれている様々なレビュー文の分類ができた。例えば、内装・外装以外の雰囲気情報として、例えば「通りから脇道に少し入った所にある...」等の立地に関する情報、「5分も待たず席に座れた」等の混雑情報、「客層は 10代・

20代がメイン...」等の客層情報、「お店の方々も...」等の飲食店で働く店員の情報など、様々な種類の情報が抽出できていた。

6. まとめ

本論文では、食べログのレビュー文とぐるなびのレビューの伝播度を求めることで、「メニュー単品の感想」「メニュー全体の感想」「飲食店の雰囲気」を分類する手法の提案をした。具体的には、食べログのレビューとぐるなびのレビューの構成から、食べログレビュー文の伝播度が高ければメニュー単品の感想に、伝播度が低ければ雰囲気に、中間であればメニュー全体の感想に分類する。評価実験より、「メニュー単品」と「雰囲気」では、提案手法はランダムフォレストとほぼ同等の精度で分類ができることを確認した。

今後の分類精度を高める課題としては、食べログレビュー文内で伝播度をノードの初期値として文書の類似度をエッジの重みとした LexRank を計算するなど、再度類似度に基づく伝播度を計算することで、レビュー文の各ラベルの分離ができ、より精度の高い分類ができるか調査する。分類手順の課題としては、「メニュー」と「雰囲気」に分類をしたあとに、それぞれを詳細に分類するように手順を改善する。分類手法の拡張の課題としては、ぐるなびのレビューにはメニュー名が書かれているので、食べログのレビュー文からメニュー名が含まれているものを特定して分類を行う手法への拡張をする。

謝 辞

本研究の一部は、平成 28 年度科研費若手研究 (B) (課題番号：15K16091) によるものです。ここに記して謝意を表すものとします。

文 献

- [1] 株式会社ジャストシステム. スマートフォンを利用した飲食店検索に関するアンケート調査, <https://marketing-rc.com/report/report-shopsearch-20150529.html>. 2015.
- [2] NTT コムリサーチ. クチコミによる飲食店選びの調査, <http://research.nttcoms.com/database/data/000738/>. 2008.
- [3] 村野誠治. 文型パターンを用いた主観的評価文の自動抽出. 言語処理学会第 11 回年次大会発表論文集, 2003, pp. 67-70, 2003.
- [4] 藤村滋, 松村真宏, 岡崎直観, 石塚満. 電子掲示板上の評判情報に基づく意思決定支援. 人工知能学会全国大会論文集, Vol. JSAI03, pp. 98-98, 2003.
- [5] 博順平, 雅彦春野. Support vector machine によるテキスト分類における属性選択. 情報処理学会論文誌, Vol. 41, No. 4, pp. 1113-1123, 2000.
- [6] 亜弓川又, 裕樹立澤, 誠岡, 博彦森. 飲食店口コミサイトを対象にした評価文抽出. 電子情報通信学会技術研究報告, Vol. 115, No. 9, pp. 47-51, 2015.
- [7] 宏実矢野, 和也目良, 輝昭相沢. 嗜好を考慮した評判情報検索手法. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, Vol. 104, No. 418, pp. 13-18, 2004.
- [8] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. In Dekang Lin and Dekai Wu, editors, *Proceedings of EMNLP 2004*, pp. 230-237, 2004.
- [9] Quoc V Le and Tomas Mikolov. Distributed representations of sentences and documents. *arXiv preprint arXiv:1405.4053*, 2014.

- [10] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pp. 3111–3119. 2013.