

機械学習モデルの解釈手法による競合サービスと比較したレビュー分析

小野川 稜之, 折原 良平, 清 雄一, 田原 康之, 大須賀 昭彦

今日, 商品やサービスのレビューを対象とした研究が多くなされている。様々な観点から研究が行われているが, 消費者がよく遭遇する, ”具体的な競合商品及びサービスと比較する”といったシチュエーションを想定した研究は少ない。一方, 近年では機械学習モデルの出力について解釈を行い, 人間がその出力の理由を理解できるようにするといった研究が行われており, LIME 及び SP-LIME などが代表例として挙げられる。そこで, 本研究では商品に関するレビューを分類する機械学習モデルについて LIME 及び SP-LIME によって分析し, ユーザーから見た商品の比較の特徴を抽出する試みを行う。

A lot of research related to products or services reviews is doing today, and is being conducted from various viewpoints. However, there are few research that assume situations where consumers often encounter, such as comparing with specific competitive products or services. On the other hand, in recent years, research has been conducted to interpret the output of machine learning models so that humans can understand the reason for the output. In particular, LIME and SP-LIME can be mentioned as representative examples. Therefore, in this research, we analyze machine learning models that classify service reviews by LIME and SP-LIME, and try to extract features which compared with competitive service viewed from users. In this paper, taking business hotels in Japan as an example, we tried to extract characteristic words from each of the hotel reviews.

1 はじめに

消費者が商品またはサービスを選択する時, 様々な情報を収集し, 購入時点で消費者自身に適合すると思われるものを選択する。その際, 消費者はカタログなどの情報はもちろん, 他の消費者が投稿した様々なレビューを参考にしながら商品・サービスを購入する。特に最近では Amazon^{†1} や楽天市場^{†2} など, 主に

商品を扱う EC サイトや, 楽天トラベル^{†3}, ホットペッパービューティー^{†4} などのサービス予約サイトも多く存在し, それぞれについて膨大な商品・サービスのレビューが蓄積されている。そのため, 消費者がレビューを目にする機会が多く, 商品・サービス選択の際の重要性は増しているといえる。

三菱 UFJ リサーチ&コンサルティングが 2018 年 9 月に行った調査” 口コミサイト・インフルエンサーマーケティングの動向整理”^{†5} によると, 商品・サービスの購入時に重視する情報源として, 46.5%の人々が 口コミサイトを挙げている。これはテレビについて 2 番目に多い数字である。このように, 消費者がレビューをもとにして商品・サービスの比較検討を行

Review analysis compared to competing services by interpreting method of machine learning model.

This is an unrefereed paper. Copyrights belong to the Author(s).

Takayuki Onogawa, Ryohei Orihara, Yuichi Sei, Yasuyuki Tahara, Akihiko Ohsuga, 電気通信大学大学院情報理工学研究科, Dept. of Infomatics, Graduate School of Informatics and Engineering, The University of Electro-Communications.

†1 <https://www.amazon.co.jp>

†2 <https://www.rakuten.co.jp>

†3 <https://travel.rakuten.co.jp>

†4 <https://beauty.hotpepper.jp>

†5 https://www.caa.go.jp/policies/policy/consumer_policy/policy_coordination/internet_committee/pdf/internet_committee_180927_0002.pdf

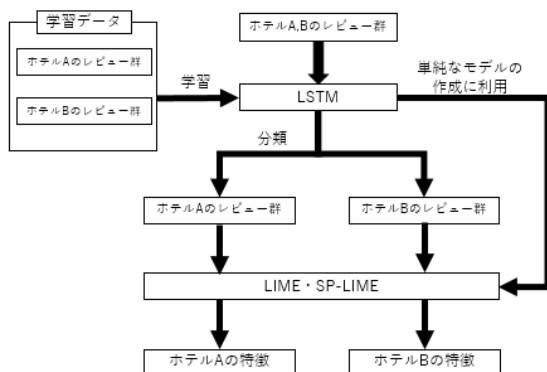


図1 本研究の全体の流れ

うということは珍しくない。

このようにECサイトのレビュー等を用いて、消費者目線で商品・サービスの比較検討を行うことが容易になった。しかし裏を返せばそれだけ多くのレビューがインターネット上にあふれていることになる。その結果、商品・サービスによってはレビューの数が膨大になってしまい、一つ一つのレビューに目を通すことは難しい状態になっている。そこで、レビュー要約の研究が多数行われている。しかし、これらの研究は一つの商品についてそのレビューの特徴を抜き出すといったものが主で、実際の消費者が陥るであろう、“複数の具体的な商品の中から相対的に良い点もしくは悪い点を抜き出す”といったことを考慮しているものは少ない。

そこで、本研究では最近注目を集めている機械学習の解釈手法を用いて商品・サービスのレビューから競合商品・サービスとの比較された特徴を抽出する試みを行う。機械学習の解釈手法は様々なものが提案されているが、なかでも代表的なものとしてLIME [?] 及びSP-LIME [?] が挙げられる。LIMEとは、ある特定インスタンスにおいて近似した有効な単純モデルを作成するものであり、SP-LIMEとはLIMEによって得た各インスタンスにおける単純モデルを複数集めることでモデル全体の挙動を理解するものである。今回はサービスの代表例としてホテルについてのレビューを得て、競合ホテルと比較した特徴を得ることを目的とした。以下は本研究の全体像を示した図である。

本稿は第2章、第3章で関連研究と関連手法に触れ

た後、第4章、第5章にて実験について述べている。その後、第6章、第7章、第8章にて考察、まとめ、今後の展望について記載している。

2 関連研究

レビュー分析は数多く行われており、消費者が商品やサービスのどのような点を評価したかを見ることで商品の特徴を得ているものもある。松尾ら [?] は、レビューの中から商品属性について記述している文章を抽出し、それらを用いて消費者が商品のどのような属性をどのように評価したかを抽出を試みた。またLIME及びSP-LIMEによる商品特徴抽出について小野川らが調査を行っている。 [?] ここではカメラについてのレビューを分類するモデルを作成し、それを解釈することによって商品特徴抽出を試みている。

3 関連手法

この項目では、本研究で用いた手法であるLIME及びSP-LIMEについて述べる。

3.1 LIME

LIMEとは、Ribeiroらによって提案された機械学習の解釈手法である。LIMEの特徴として、機械学習のモデルに依存しない解釈手法であることが挙げられる。様々なモデルの個々の分類結果について、人間にも理解しやすいシンプルなモデルによって解釈することによって、複雑な機械学習モデルの理解に貢献している。例えば、ある二値分類モデル $f: \mathbb{R}^d \rightarrow \mathbb{R}$ と、あるインスタンス $x \in \mathbb{R}^d$ を考える。この時、解釈可能なモデル群 G から、 f の働きに近いような、解釈可能なモデル g を見つけることが、LIMEの働きである。LIMEは以下のような最適化問題として表されている。

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \mathcal{L}(f, g, \pi_x) + \Omega(g) \quad (1)$$

$$\mathcal{L}(f, g, \pi_x) = \sum_{z, z' \in \mathcal{Z}} \pi_x(z) (f(z) - g(z'))^2 \quad (2)$$

上記の式において、 $x' \in \{0, 1\}^d$ はインスタンス x を二値変数ベクトルに変換したものであり、

$z' \in \{0, 1\}^{d'}$ は x' の非ゼロ変数の一部を含んだものである。そして、 $z \in \mathbb{R}^d$ は $f(z)$ を得るために z' から変換したものである。さらに $\pi_x(z)$ は、インスタンス x の近傍にある点 z との距離に応じて $(f(z) - g(z'))^2$ 、つまり $f(z)$ と $g(z')$ との誤差に重みづけを行うものである。この時、 x と z の距離が近ければ近いほど重みが大きくなる。つまりあるインスタンス x の近傍においていくつかデータを取り、元のデータと解釈可能なモデルとの誤差を計算し、 x に近いデータにおいてその誤差をより重視するようになっている。 $\Omega(g)$ は解釈可能なモデル g の複雑さを表す。これと $\mathcal{L}(f, g, \pi_x)$ を足したものを最小化することによって、インスタンス x 付近について元のモデルとの誤差がより小さく、より単純なモデルを作ることを目的としている。

3.2 SP-LIME

先ほど述べた LIME による説明を複数集めて、モデル全体の理解を目的にした手法である。SP-LIME では、様々な特徴をカバーできるように、上限 B 個の中で、1 つでも多くの特徴量を含むようなインスタンス群を見つける。あるインスタンス群 $X(|X|=n)$ について、各インスタンスについての LIME による出力を用いて $n \times d'$ サイズの行列 W を作成する。この時、行列 W の各列の構成要素について、説明空間における全体的な重要性を $I_j (j \leq d')$ とおく。例えば各インスタンスがテキストについてのデータであれば、 $I_j = \sqrt{\sum_{i=1}^n W_{ij}}$ と計算される。上限 B 個のインスタンス群を選ぶ上で、できるだけ多くのインスタンスの説明に関与し、重要度も高い特徴量が多く含まれているインスタンスを順に集めるということが SP-LIME の行うことである。これは、以下の式によって表される。

$$c(V, W, I) = \sum_{j=1}^{d'} \mathbb{1}_{[\exists i \in V: W_{ij} > 0]} I_j \quad (3)$$

$$\text{Pick}(W, I) = \underset{V, |V| \leq B}{\text{argmax}} c(V, W, I) \quad (4)$$

この時、 $c(V, W, I)$ はインスタンス群 V の中で、少なくとも一つのインスタンスに含まれる全ての特徴

量の I_j の和を計算する。この特徴量の和が最大となるような V を設定することによって、より重要な特徴量をより多く含むようなインスタンス群を得ることができる。

4 実験アプローチ

4.1 実験データ

本稿では、商品・サービスの例としてビジネスホテルを取り上げた。競合するホテルとして、横浜に存在するアパホテル<横浜関内>^{†6} と、コンフォートホテル横浜関内^{†7} を取り上げた。この二つのホテルはどちらも日本のビジネスホテルチェーンのうちの一つであり、道路を挟んで立地している。立地やビジネスホテルという形態が同じであることから今回競合として選択した。

本稿では楽天データセット [?] を国立情報学研究所情報学研究データリポジトリより入手し、その内 1997 年 1 月から 2015 年 11 月までに楽天トラベルへ蓄積されたレビューデータ [?] を用いた。ここからアパホテル<横浜関内>について記述されたものと、コンフォートホテル横浜関内について記述されたものを抽出し、実験に用いた。レーティングに関係なく 2 つのホテルについて記載された全てのレビューを抽出したものと、総合評価が 5 段階中 4 以上を付けたもの限定して抽出したものの 2 つを用意し、それぞれについて実験を行った。

4.2 レビューデータの前処理

先述によって得られたレビューデータについて、分類モデルに投入するための前処理を行った。英語・記号を全て半角及び小文字にし、MeCab^{†8} を用いて分かち書きを行った。この際、辞書として NEologd [?] を用いた。分かち書きした文章から助詞、助動詞、記号と判定されたものを除去したうえで全ての単語を原型にすることで単語の統一を図った。その後、日本

^{†6} https://www.apahotel.com/hotel/shutoken/08_yokohamakannai/

^{†7} <https://www.choice-hotels.jp/hotel/kannai/>

^{†8} <https://taku910.github.io/mecab/>

語ストップワード辞書の一種である slothlib^{†9} を用いた単語の除去やデータ中 1 度のみ使用された単語の除去を行った。

4.3 LSTM モデルによる学習

前処理を行ったレビューデータについて、それぞれのホテルについてのレビューからランダムに同じ数抽出し、それぞれトークナイズとパディングを施したのち、LSTM [?] モデルへ投入した。この時学習データ、バリデーションデータ、テストデータの割合をそれぞれ 67.5%、7.5%、25.0%とおいた。LSTM モデルは各単語を 64 次元による表現を入力し、64 次元の出力を得たのち、sigmoid 関数によってどのホテルについてのレビューであるかを確率値によって出力する。今回は dropout, recurrent_dropout の値をそれぞれ 0.5 に設定し、最適化には Adam [?] を用いた。

4.4 LIME 及び SP-LIME の適用とカイ二乗値との比較

テストデータについて、LIME 及び SP-LIME を適用したのち、SP-LIME による各インスタンスの単語ごとの重みの和と、テストデータや学習データ、バリデーションデータの全てのレビューデータから得た各単語のカイ二乗値について比較を行い、どのような単語が出力されているかの比較を行った。ここでは、カイ二乗値はそれぞれのホテルのレビューにおいての単語出現頻度の差を表したものであり、値が大きければ大きいほどどちらかのホテルに偏って単語が出現していることを表す。また、SP-LIME によるインスタンス群の抽出と、ランダムによる抽出でどの程度インスタンス群においての出現単語の数に差が出るかの確認も行った。LIME では各インスタンスについて説明に利用される単語を上限 10 個とした。

5 実験結果

本稿では、2 つのホテルについての全てのレビューを用いた場合を実験 1、総合評価が 5 段階中 4 以上で

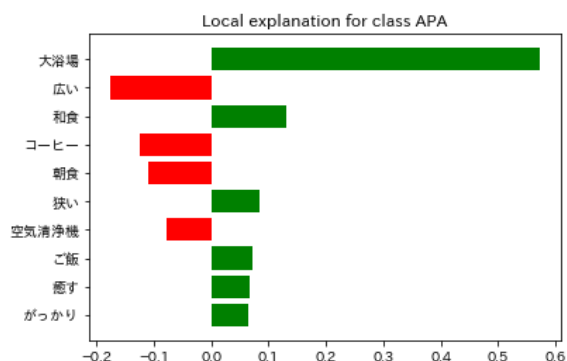


図 2 実験 1 における SP-LIME による出力例

あるレビューを用いた場合を実験 2 とし、それぞれについて実験を行った。

5.1 実験 1

実験 1 では各ホテルそれぞれ 5234 件、合計 10468 件のレビューを用いた。LSTM モデルによって学習では 10 エポック学習し、学習データにおいて 93.57%、テストデータについて 84.10%の accuracy を得た。このモデルについて、LIME 及び SP-LIME を用いて解釈を行った。図 2 は SP-LIME によって一番最初に収集されたインスタンスにおける LIME による出力を示している。図 3 は抽出されるインスタンス群の数と、説明に利用される単語数について表したグラフであり、SP-LIME による抽出とランダムによる抽出を比較している。また、カイ二乗検定によって有意となった単語数とも比較している。有意水準は 0.05 とした。表 1 は、SP-LIME によって得られたインスタンス群での単語の重みの和と、カイ二乗値について絶対値が大きいものについて順に 30 単語を示したものである。ここではインスタンス群のサイズを 132 と設定した。これは、カイ二乗値が有意である単語数と、インスタンス群を 132 としたときに SP-LIME で説明に寄与する単語数がほぼ同じだからである。

^{†9} <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt>

^{†1} SP-LIME の出力において、正の値を示す単語はコンフォートホテル横浜関内に寄与しているもの、負の値を示す単語はアパホテル<横浜関内>に寄与しているものである。

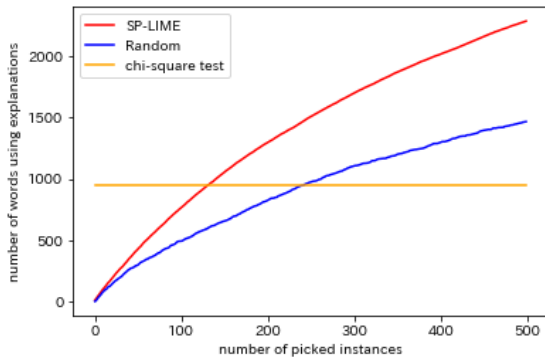


図3 実験1におけるインスタンス群のサイズと説明に利用された単語数の比較

5.2 実験2

実験1では各ホテルそれぞれ1636件、合計3272件のレビューを用いた。LSTMモデルによって学習では6エポック学習し、テストデータにおいて89.86%、テストデータについて78.97%のaccuracyを得た。このモデルについて、LIME及びSP-LIMEを用いて解釈を行った。図4は図2同様、SP-LIMEによって一番最初に収集されたインスタンスにおいてのLIMEによる出力を示している。図5は図3同様、抽出されるインスタンス群の数と、説明に利用される単語数について表したグラフである。表2は表1同様、SP-LIMEによって得られたインスタンス群での単語の重みの和と、カイ二乗値について絶対値が大きいものについて順に30単語を示したものである。インスタンス群のサイズは実験1と同様の理由で50とした。

6 考察

実験1, 2の結果について考察を行う。

SP-LIMEがどの程度の幅広く説明に利用する単語を集めているかを確認するためにランダムによるインスタンス収集と比較して、カイ二乗検定において有意な単語数にどれほどのインスタンス数で到達するかを見た。図3・図5を見ると、実験1, 2において説明に利用された単語の数はランダムによる収集と比較して大きく上回っていることがわかる。これにより、SP-LIMEが幅広い単語を用いて説明を行っていることが確認できる。

表1 実験1におけるSP-LIMEによる重みとカイ二乗値 †1

SP-LIME		カイ二乗値	
朝食	7.884809	朝食	2286.803
大浴場	-3.701995	大浴場	2211.660
狭い	-2.479558	無料	790.0334
無料	2.048154	コーヒー	745.1935
利用	1.128136	狭い	715.4207
コーヒー	1.104071	おにぎり	529.5876
ホテル	1.014865	サービス	513.0529
おにぎり	0.956077	種類	456.6773
COMFORT	0.949595	風呂	341.9325
ツイン	0.885914	パン	339.8572
朝食会	0.838011	サウナ	339.2225
コンフォートホテル	0.752975	枕	287.0810
食事	0.744619	露天風呂	267.6736
面す	-0.732293	豊富	255.9728
風呂	-0.693815	また	251.7871
休む	0.668611	美味しい	238.5051
サービス	0.662576	充実	235.4759
種類	0.654569	アパホテル	234.9571
アパホテル	-0.635659	サラダ	206.3131
眠れる	0.622635	満足	181.4549
手管	0.571073	COMFORT	178.9773
バイキング	-0.547223	広い	171.4963
ん	0.544555	食べる	166.4457
溜まる	-0.534637	おいしい	162.7459
浴衣	-0.533109	フルーツ	161.0951
最低	-0.529818	快適	157.5804
枕	0.503011	コンフォートホテル	151.9697
サウナ	-0.477166	うるかむ!	145.8846
空気清浄機	0.472982	利用	144.9079
入浴剤	0.472021	お部屋	138.5463

次に、SP-LIMEによる重みの合計とカイ二乗値の上位の単語について比較を行った。

実験1においては”朝食”，”大浴場”などのカイ二乗値が抜きんで大きいものについてはSP-LIMEによる出力でも同じように大きい重みが観測された。しかし、それ以外の単語についてはカイ二乗値とは異なる出力がされた単語も見られた。例えば、SP-LIMEにおいて10番目に大きい重みをもつ”ツイン”という単語は、カイ二乗値において上位30個の中に入っていない。それぞれのホテルの部屋の内訳を確認してみると、アパホテル<横浜関内>は全451室中ツインルームは2室しか存在しない。対してコンフォートホテル横浜関内では37室あり、ファミリールームなどのエキストラベッドを備えたものを含めると47室

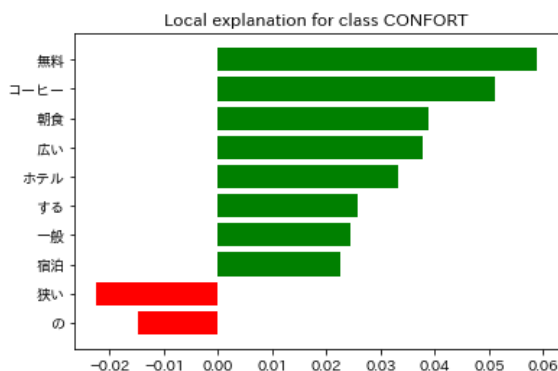


図 4 実験 2 における SP-LIME による出力例

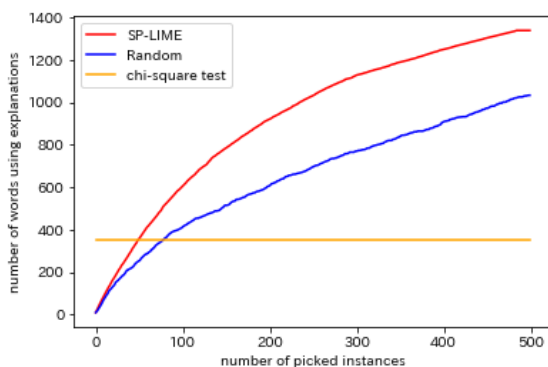


図 5 実験 2 におけるインスタンス群のサイズと説明に利用された単語数の比較

ある。SP-LIME によって出力された ” ツイン ” という単語はコンフォートホテルに寄与しており、カイ二乗値では抽出できなかったアパホテルと比較しての特徴をとらえている。かつ、カイ二乗値においては高い値を示しておらず、本稿の手法によって抽出できた単語であるといえる。

実験 2 における出力でも ” 朝食 ”, ” 大浴場 ” などの単語の重みが大きく、それぞれのホテルのレビューについて出現頻度の差が激しいものを上位に出力している。また、カイ二乗値においては上位に出現しない ” 空気清浄機 ” という単語を SP-LIME は出力している。これについて確認してみると、コンフォートホテル横浜関内は全客室において加湿空気清浄機を設置している旨の記載が公式ホームページにおいてなされているが、アパホテル<横浜関内>についてはその

表 2 実験 2 における SP-LIME による重みとカイ二乗値 †1

SP-LIME		カイ二乗値	
朝食	2.604478	大浴場	1085.637712
する	1.537726	朝食	739.741483
大浴場	-1.512860	コーヒー	287.214928
宿泊	0.726262	無料	258.070253
枕	0.647186	サウナ	173.410474
ホテル	0.630677	風呂	169.632145
の	-0.620051	サービス	157.358902
狭い	-0.552285	おにぎり	145.763921
楽しめる	0.483439	露天風呂	128.963524
快適	0.472759	種類	127.155511
コンフォートホテル	0.453169	狭い	124.564731
もう少し	-0.450689	美味しい	113.142959
コーヒー	0.431212	アパホテル	106.742112
アパホテル	-0.405381	バン	105.226362
美味しい	0.401400	枕	94.268708
食事	0.401051	また	67.800695
風呂	-0.393582	広い	65.776422
新しい	0.386190	思う	63.197737
中華街	0.385386	COMFORT	63.117810
十分	0.383040	おいしい	62.861362
空気清浄機	0.376613	アパ	62.301055
すぐ	-0.360058	スープ	61.025647
サービス	0.341661	コンフォートホテル	60.491922
エレベーター	0.327680	最上階	59.808286
便利	0.317349	充実	58.680654
ベッド	0.315901	入れる	56.436886
コストパフォーマンス	0.310012	豊富	54.427202
助かる	0.295816	ホテル	52.946369
清掃	0.286976	する	52.174216
入れる	-0.281533	疲れ	50.025764

ような記載がなされていない。ここでも、SP-LIME がアパホテルと比較したコンフォートホテルの特徴を抽出できており、カイ二乗値によって得られた単語とは異なる単語が得られている。

最後に、SP-LIME によって最初にインスタンス群へピックアップされたインスタンスの出力と、SP-LIME 全体の出力とを比較する。

実験 1 では、” 大浴場 ”, ” 狭い ”, ” コーヒー ”, などの SP-LIME 及びカイ二乗値の出力において上位に位置している単語が含まれたレビューが抽出されている。これにより、SP-LIME が多くのインスタンスの説明に関与し、かつ重要な単語を含むインスタンスを抽出することでモデルを代表するようなインスタンスを確認することができる。実験 2 においても ” 無料 ”, ” コーヒー ”, ” 朝食 ” などの単語が出現しているレビューを抽出できている。

7 まとめ

消費者はレビューなどを見て競合商品・サービスと比較しながら選択をする。この際、多くのレビューを全て見ることは困難であり、レビューから商品・サービスの特徴を抽出するニーズがある。そこで本稿ではホテルを例にとってレビュー分析を行い、実際の消費者行動に近いような”競合との比較した特徴”を抽出することを目的とした。その結果、他方のホテルと比較した特徴を抽出することができた。また、SP-LIME を利用することでホテル比較に役立つような代表的なレビューについてモデルの挙動を確認することができた。

8 今後の展望

本稿では、2つのホテルについて他方と比較した特徴を抽出した。しかし、実際の消費者行動においてはもう少し多くの商品・サービスから選ぶことも少なくない。そこで、3つ以上の商品・サービスについてお互いを競合商品・サービスとおき、他の競合とと比較した特徴を抽出することを行っていききたい。また本稿では、他方のホテルと異なる点を単語にて抽出したが、より消費者にわかりやすいような文章としての出力を行うことを目指している。文章による出力によっ

て、単語だけではわかりにくい特徴について消費者に提示したいと考えている。

謝辞

本研究は JSPS 科研費 JP17H04705, JP18H03229, JP18H03340, 18K19835, JP19H04113, JP19K12107 の助成を受けたものです。

本研究を遂行するにあたり、研究の機会と議論・研鑽の場を提供して頂き、御指導頂いた早稲田大学 本位田 真一 教授をはじめ、活発な議論と貴重な御意見を頂いた研究グループの皆様にご感謝致します。

また本研究では、楽天株式会社が提供し、国立情報学研究所が配布している「楽天データセット」を利用した。関係者の皆様にご大変感謝致します。

参考文献

- [1] Lanin, V. and Shasha, D.: A Symmetric Concurrent B-Tree Algorithm, *Proc. 1986 Fall Joint Computer Conference*, IEEE, 1986, pp. 380–389.
- [2] Shapiro, E.: The Family of Concurrent Logic Programming Languages, *ACM Computing Surveys*, Vol. 21, No. 3(1989), pp. 413–510.
- [3] Sleator, D. D. and Tarjan, R. E.: Self-Adjusting Binary Search Trees, *J. ACM*, Vol. 32, No. 3(1985), pp. 652–686.
- [4] Tarjan, R. E.: Amortized Computational Complexity, *SIAM J. Alg. Disc. Math.*, Vol. 6, No. 2(1990), pp. 306–318.
- [5] 和田久美子: スプレイ木の並列データ探索, *Proc. KLI Programming Workshop '90*, Tokyo, ICOT, 1990, pp. 42–49.