

単語分散表現と文法的な表現に着目した カスタマーレビューの観点ごとの評価値推定

大塚 達也[†] 青野 雅樹^{††}

[†] 豊橋技術科学大学情報・知能工学専攻 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

^{††} 豊橋技術科学大学情報・知能工学系 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

E-mail: [†]otsukat@kde.cs.tut.ac.jp, ^{††}aono@tut.jp

あらまし インターネット上には、ユーザがレビューを書き込みできるサイトが多く、そのレビューの評価が他のユーザの購買活動に大きな影響を与える。本研究では、宿泊施設のレビューサイト、ならびに薄型テレビのレビューサイトのカスタマーレビューに焦点をあて、観点ごとの評価値予測を試みる。宿泊施設のレビューでは、立地、部屋、食事、風呂などの観点、薄型テレビのレビューでは、デザイン、操作性、画質などの観点を持つ。評価値の予測では、観点ごとに回帰モデルの構築を試みる。このため、異なる観点で有効となる素性の提案を行う。カスタマーレビューの特徴は、極性表現が多いこと、また、観点到じた文法的な表現や特殊な用語が多く観察されることである。そこで、それぞれのレビューに関連するドメインで別途入手した比較的大規模なコーパスで分散表現モデルを構築した。その後、極性表現や観点ごとの専門用語が多いといった特徴を反映できる素性を分散表現から抽出し、回帰モデルを訓練し、未知なレビューでの観点ごとの評価値予測を行った。実証実験では、BoWによる素性から訓練された回帰モデルをベースラインとし、提案手法との比較実験を行った。その結果、提案素性の有効性を確認することができた。

キーワード 分散表現, カスタマーレビュー, 回帰

1. はじめに

近年、商品やサービスに対する評価や意見が、Web上の様々な人によって発信されている。特に、インターネットやスマートフォンが普及し、個人が気軽に商品やサービスの情報を発信できるようになった。情報が発信される場所の一つとして、商品やサービスのWebページのカスタマーレビューがある。カスタマーレビューを読むことで、まだ商品を購入していない人や、サービスを利用していない人は購入や利用を検討することができる。また、商品やサービスを提供する企業は、カスタマーレビューに書かれている要求や苦情を分析して、次世代の商品やサービスの開発に役立てることができる。

しかし、商品やサービスへの評価がカスタマーレビューから判明しても、それだけでは不十分な可能性がある。商品やサービスを利用する際は、レビュー対象のどの観点を重視するかが顧客ごとに異なることがある。例えば、レビュー対象が薄型テレビの場合、リビングにテレビを置こうとしている人はデザインを重視し、テレビゲームをしようとしている人は応答性能を重視するといったことが挙げられる。また、サービスを運営している企業も、顧客がレビュー対象のどの観点について評価しているかを知ることが有益である。例えば、自社の宿泊施設に泊まった宿泊客がカスタマーレビューに評価を書いたとき、それが立地についての評価なのか、食事についての評価なのか、風呂についての評価なのかなどを正確に知ることで、宿泊施設のセールスポイントを伸ばしたり、クレーム処理を効率よく行えるといったことが挙げられる。このように、商品やサービスの全体の評価ではなく、ある1観点到絞った上での評価を判定

することは、顧客と企業双方にとって重要であると考えられる。

本研究では、宿泊施設と薄型テレビのカスタマーレビューを取り扱う。利用する宿泊施設のカスタマーレビューには、カスタマーレビューの本文とともに、評価観点として「立地」「部屋」「食事」「風呂」「サービス」「設備・アメニティ」「総合評価」の7項目が設けられている。薄型テレビのカスタマーレビューには、同様に「デザイン」「操作性」「画質」「音質」「応答性能」「機能性」「サイズ」「満足度」の8項目が設けられている。

カスタマーレビューにはこれらの観点の評価が複数に混じり合っており、文章そのままではレビュー対象がどのように評価されているのかを判断することが難しい。そのため、文章を分析して、観点ごとに評価値を推定することにより、レビュー対象に対して行われた評価を観点ごとにまとめるシステムが求められている。そこで、本研究ではカスタマーレビューから素性を抽出し、観点ごとの評価値をより正確に推定することを目的とする。

2. 関連研究

Poriaら[1]は、5種類のルールに基づく言語パターンと、畳み込みニューラルネットワークを使用して、文章から観点表現を抽出する手法を提案した。ニュースサイトで訓練された単語分散表現と、Amazonで訓練した単語分散表現の2種類を利用して比較実験を行った。

Santosら[2]は、単語単位ではなく、文字単位での特徴も利用した畳み込みニューラルネットワークの利用を提案した。映画レビューとTwitterのデータセットにおいて、既存のRNNを用いる手法よりも性能を改善した。

Takamura ら [3] は、英語と日本語の文章に含まれる各単語を電子とみなし、各単語が持つ感情極性（ネガティブ、ポジティブ）を電子のスピンの向きと対応させ、統計物理分野の電子のスピンモデルを用いた単語の感情極性抽出を行った。

3. 提案手法

カスタマーレビューから観点ごとの評価値を推定する方法について述べる。カスタマーレビューから観点ごとの評価値を推定するシステムの全体図を図 1 に示す。本研究では、まず、カスタマーレビューから素性の抽出を行う。また、他のカスタマーレビューを用いて、単語分散表現モデルを構築する。構築した単語分散表現モデルからも素性の抽出を行う。さらに、単語分散表現モデルから単語感情極性辞書を作成し、辞書からも素性の抽出を行う。そして、抽出した素性を用い、観点ごとの評価値を正解ラベルとして回帰モデルを訓練する。テスト時は、テストデータ、単語分散表現モデル、単語感情極性辞書、回帰モデルを用いて、観点ごとの評価値を推定する。

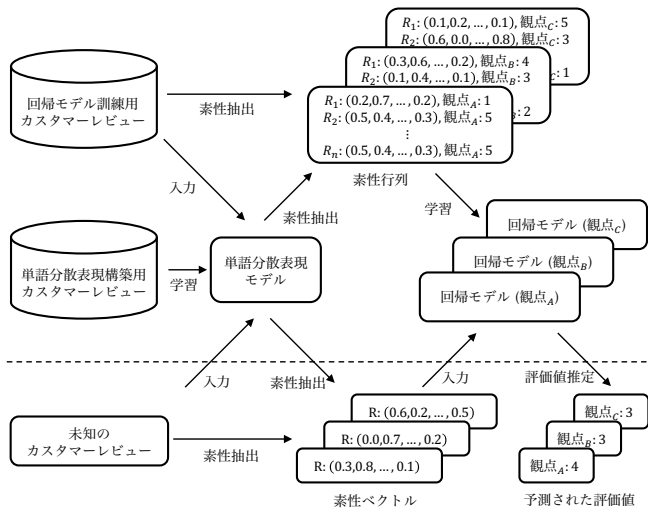


図 1 カスタマーレビューから観点ごとの評価値を推定するシステムの全体図

3.1 形態素解析を用いた前処理

カスタマーレビューを単語に分割し、それぞれの単語の品詞を識別するために形態素解析を用いた前処理を行う。本研究では、物が詰まっていることを表す「つまり」という単語と、「すなわち」と同じ意味の「つまり」などのような単語を区別して処理を行うために、また、後述する素性で品詞の情報を利用するために、単語に品詞を付与してわかち書きを行う。付与された品詞の中でも、動詞、形容詞、助動詞の単語は原型に変換する。そのため、「比較的安く泊まりました」という文は「比較的-副詞-一般」「安い-形容詞-自立」「泊まれる-動詞-自立」「ます-助動詞-」た-助動詞-」という単語に分割される。

3.2 単語分散表現の学習

word2vec [4], [5] や GloVe [6] のような、単語分散表現 (distributed word representation) と呼ばれる、単語を多次元空間のベクトルとして表現する手法が研究されている。本研究で使用している訓練用のデータセットは、少ないもので 23,482 件、多いも

ので 43,983 件のカスタマーレビューしかないので、十分に単語分散表現の学習ができないという問題がある。この問題は、Google News で訓練された word2vec のモデル¹や Wikipedia で訓練された GloVe のモデル²などの、事前に訓練された単語分散表現を使用することで解決できる。

しかし、ニュースや Wikipedia で訓練した単語分散表現を、カスタマーレビューから評価値を推定するために利用することは問題があると考えられる。なぜなら、同じ単語であっても、ニュースや Wikipedia で使われる場合の意味とカスタマーレビューで使われる場合の意味が異なることがあるからである。例えば、ニュースサイトにおいて「高い-形容詞-自立」という単語は、「高い科学技術水準」のように良い意味で使われることが多い。一方で、カスタマーレビューでは「価格が高い」のような悪い意味で使われることが多い。そこで、本研究では、同じカテゴリや似たカテゴリのカスタマーレビューを別途入手し、そのデータを用いて単語分散表現を構築する。

3.3 単語感情極性辞書の作成

本研究では、独自の単語感情極性辞書を作成する。単語感情極性辞書とは、ある単語がポジティブか、ネガティブかを示した辞書である。例えば、「素晴らしい-形容詞-自立」という単語はポジティブであり、「狭苦しい-形容詞-自立」という単語はネガティブである。

単語感情極性辞書は、Bing Liu's Opinion Lexicon [7], SentiStrength [8], Takamura らの単語感情極性対応表 [3] など様々なものが存在している。これらの単語感情極性辞書はあくまで汎用的な領域の辞書であり、特定のドメイン固有の辞書ではない。

そこで、本研究では、第 3.2 節の方法で構築した単語分散表現を用いて辞書を作成する。まず、単語分散表現の語彙に含まれる単語 w を用いて、単語感情極性値 $s(w)$ を以下の式で計算する。

$$s(w) = \text{sim} \left(\mathbf{v}(w), \frac{\sum_{i=1}^m \text{df}(u_i^{(p)}) \mathbf{v}(u_i^{(p)})}{\sum_{i=1}^m \text{df}(u_i^{(p)})} \right) - \text{sim} \left(\mathbf{v}(w), \frac{\sum_{i=1}^m \text{df}(u_i^{(n)}) \mathbf{v}(u_i^{(n)})}{\sum_{i=1}^m \text{df}(u_i^{(n)})} \right) \quad (1)$$

ここで、 $\text{sim}(\mathbf{a}, \mathbf{b})$ はベクトル \mathbf{a} とベクトル \mathbf{b} のコサイン類似度、 $\mathbf{v}(w)$ は単語 w の単語ベクトル、 $\text{df}(u)$ は単語 u が出現する訓練データセット内のカスタマーレビューの件数である。また、 $u_i^{(p)}$ はポジティブな m 種類の単語、 $u_i^{(n)}$ はその対義語のネガティブな m 種類の単語をそれぞれ手動で与える。

例として、宿泊施設データセット用に単語感情極性辞書を作成するとき $m = 3$ として与えた単語と、各単語が出現するカスタマーレビューの件数を表 1 に示す。

次に、単語分散表現の語彙 w を、式 (1) の単語感情極性値 $s(w)$ を用いて並び替え、値が大きい k 件の単語をポジティブ単語、値が小さい k 件の単語をネガティブ単語として、単語感情極性辞書を作成する。

(注 1) : <https://drive.google.com/file/d/0B7XkCwpI5KDYN1NUTTLSS21pQmM>

(注 2) : <http://nlp.stanford.edu/data/glove.6B.zip>

表1 単語感情極性辞書の作成に使う単語の例

| 単語 | 出現するレビュー数 |
|---------------------------|-------------------------|
| $u_1^{(p)}$ = “良い-形容詞-自立” | $df(u_1^{(p)}) = 13133$ |
| $u_2^{(p)}$ = “広い-形容詞-自立” | $df(u_2^{(p)}) = 6349$ |
| $u_3^{(p)}$ = “近い-形容詞-自立” | $df(u_3^{(p)}) = 3443$ |
| $u_4^{(n)}$ = “悪い-形容詞-自立” | $df(u_4^{(n)}) = 1994$ |
| $u_5^{(n)}$ = “狭い-形容詞-自立” | $df(u_5^{(n)}) = 2294$ |
| $u_6^{(n)}$ = “遠い-形容詞-自立” | $df(u_6^{(n)}) = 955$ |

表2 単語感情極性辞書の例 (名詞, 形容詞, 動詞のみ)

| 単語 w | 極性値 $s(w)$ | ポジ/ネガ |
|--------------|------------|-------|
| 良い-形容詞-自立 | 0.6571 | ポジティブ |
| 良い-形容詞-非自立 | 0.5812 | ポジティブ |
| よい-形容詞-自立 | 0.5313 | ポジティブ |
| 最高-名詞-一般 | 0.4879 | ポジティブ |
| よい-形容詞-非自立 | 0.4730 | ポジティブ |
| 好い-形容詞-自立 | 0.4173 | ポジティブ |
| いい-形容詞-自立 | 0.3960 | ポジティブ |
| ⋮ | ⋮ | ⋮ |
| 悪い-形容詞-自立 | -0.3625 | ネガティブ |
| 以外-名詞-非自立 | -0.3636 | ネガティブ |
| それ-名詞-代名詞 | -0.3925 | ネガティブ |
| 手狭-名詞-形容動詞語幹 | -0.3931 | ネガティブ |
| 窮屈-名詞-形容動詞語幹 | -0.3960 | ネガティブ |
| せまい-形容詞-自立 | -0.4714 | ネガティブ |
| 狭い-形容詞-自立 | -0.5549 | ネガティブ |

例として、宿泊施設データセット用に辞書を作成した結果を表2に示す。

3.4 素性抽出

本研究で用いる素性とその抽出方法について順番に述べる。

3.4.1 Bag-of-Words 素性

Bag-of-Words (BoW) 素性は自然言語処理問題で用いられる素性抽出方法である。BoW 素性は文章中に単語が含まれているかだけを考え、単語の位置は考慮しないモデルである。

本研究では、訓練データセット内で出現回数が15回以上の単語のうち、図2に示す品詞を用いる。

- 名詞 (“の”を除く)
- 動詞 (“する”, “ある”, “いる”を除く)
- 形容詞
- 副詞
- 接続詞
- 接頭詞
- 感動詞
- 記号 (“!”と“?”のみ)
- 連体詞
- 助動詞 (“ない”と“ん”を除く)

図2 使用する品詞

BoW 素性の単語の重みには Binary, tf-idf などがよく用いら

れる。出現回数が15回以上現れる単語の数はデータセットによって異なるので、この素性の次元数はデータセットに依存する。

3.4.2 単語感情極性辞書素性

第3.2節で作成した単語感情極性辞書を使用した単語感情極性辞書素性を提案する。この素性は、カスタマーレビュー中の単語の感情極性値を使用する。作成した単語感情極性辞書には、単語 w と感情極性値 $s(w)$ が含まれている。実数の感情極性値 $s(w)$ が割り当てられている。この辞書を用いて、ある形態素解析されたカスタマーレビュー R に対して以下の10種類の素性を定義する。

(1) ポジティブ単語数

1種類目の値はカスタマーレビュー中のポジティブ単語数 p_1 である。 p_1 の値は次のようになる。

$$p_1 = \sum_{w \in R} \chi_P(w) \quad (2)$$

$$\chi_P(w) = \begin{cases} 1 & \text{if } w \in P \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

ここで、 P は単語感情極性辞書に登録されているポジティブ単語の集合である。

(2) ネガティブ単語数

2種類目の値はカスタマーレビュー中のポジティブ単語数 n_1 である。 n_1 の値は次のようになる。

$$n_1 = \sum_{w \in R} \chi_N(w) \quad (4)$$

$$\chi_N(w) = \begin{cases} 1 & \text{if } w \in N \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

ここで、 N は単語感情極性辞書に登録されているネガティブ単語の集合である。

(3) ポジティブ単語数とネガティブ単語数の差

3種類目の値は、(1)と(2)で定義した p_1 と n_1 を用いて、ポジティブ単語数 p_1 とネガティブ単語数 n_1 の差を以下の式 $diff_1$ で与える。

$$diff_1 = p_1 - n_1 \quad (6)$$

(4) ポジティブ単語数とネガティブ単語数の均衡

4種類目の値は、ポジティブ単語数 p_1 とネガティブ単語数 n_1 の均衡を以下の式 $balance_1$ で与える。

$$balance_1 = \begin{cases} 0 & \text{if } p_1 + n_1 = 0 \\ \exp\left(-\frac{(p_1 - n_1)^2}{p_1 + n_1}\right) & \text{otherwise} \end{cases} \quad (7)$$

(5) ポジティブ単語数とネガティブ単語数の調和平均

5種類目の値は、ポジティブ単語数 p_1 とネガティブ単語数 n_1 の調和平均を以下の式 $mean_1$ で与える。

$$mean_1 = \begin{cases} 0 & \text{if } p_1 = 0 \text{ or } n_1 \\ \frac{2p_1 n_1}{p_1 + n_1} & \text{otherwise} \end{cases} \quad (8)$$

(6) ポジティブ感情極性値

6種類目の値はカスタマーレビューのポジティブ感情極性値 p_2 である。 p_2 の値は次のようになる。

$$p_2 = \sum_{w \in \mathbf{R}} \psi_{\mathbf{P}}(w) \quad (9)$$

$$\psi_{\mathbf{P}}(w) = \begin{cases} s(w) & \text{if } w \in \mathbf{P} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

ここで、 \mathbf{P} は単語感情極性辞書に登録されているポジティブ単語の集合、 $s(w)$ は単語 w の単語感情極性値である。

(7) ネガティブ感情極性値

7種類目の値はカスタマーレビューのネガティブ感情極性値 n_2 である。 n_2 の値は次のようになる。

$$n_2 = \sum_{w \in \mathbf{R}} \psi_{\mathbf{N}}(w) \quad (11)$$

$$\psi_{\mathbf{N}}(w) = \begin{cases} -s(w) & \text{if } w \in \mathbf{N} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

ここで、 \mathbf{N} は単語感情極性辞書に登録されているネガティブ単語の集合、 $s(w)$ は単語 w の単語感情極性値である。

(8) ポジティブ感情極性値とネガティブ感情極性値の差

8種類目の値は、(6)と(7)で定義した p_2 と n_2 を用いて、ポジティブ感情極性値 p_2 とネガティブ感情極性値 n_2 の差を以下の式 $diff_2$ で与える。

$$diff_2 = p_2 - n_2 \quad (13)$$

(9) ポジティブ感情極性値とネガティブ感情極性値の均衡

9種類目の値は、ポジティブ感情極性値 p_2 とネガティブ感情極性値 n_2 の均衡を以下の式 $balance_2$ で与える。

$$balance_2 = \begin{cases} 0 & \text{if } p_2 + n_2 = 0 \\ \exp\left(-\frac{(p_2 - n_2)^2}{p_2 + n_2}\right) & \text{otherwise} \end{cases} \quad (14)$$

(10) ポジティブ感情極性値とネガティブ感情極性値の調和平均

10種類目の値は、ポジティブ感情極性値 p_2 とネガティブ感情極性値 n_2 の調和平均を以下の式 $mean_2$ で与える。

$$mean_2 = \begin{cases} 0 & \text{if } p_2 = 0 \text{ or } n_2 = 0 \\ \frac{2p_2n_2}{p_2+n_2} & \text{otherwise} \end{cases} \quad (15)$$

3.4.3 単語分散表現素性

第3.2節で構築した単語分散表現を使用する素性について述べる。

あるカスタマーレビューが、「おいしい-形容詞-自立」という単語とどの程度類似しているかを考える。例えば、「朝食のスープにコクがありました」というカスタマーレビューは宿泊施設の食事が美味しいということを述べているので、「おいしい-形容詞-自立」と類似していると考えられる。一方で、「駅からかなり歩くのが難点」というカスタマーレビューは食事のこと

を述べておらず、かつネガティブな内容を述べているので「おいしい-形容詞-自立」とは類似していないと考えられる。一方で、「遠い-形容詞-自立」という単語は、「駅からかなり歩くのが難点」という文に類似していると考えられる。

そこで、ある単語 w と、ある形態素解析されたカスタマーレビュー \mathbf{R} の類似度を以下の式 $wrsim(w, \mathbf{R})$ で定義する。

$$wrsim(w, \mathbf{R}) = \text{sim}\left(v(w), \frac{\sum_{u \in \mathbf{C} \cap \mathbf{R}} v(u)}{|\mathbf{C} \cap \mathbf{R}|}\right) \quad (16)$$

ここで、 $\text{sim}(\mathbf{a}, \mathbf{b})$ はベクトル \mathbf{a} とベクトル \mathbf{b} のコサイン類似度、 $v(w)$ は単語 w の単語ベクトル、 \mathbf{C} は図2に示す品詞の単語の集合である。

本素性では、複数個の単語 w と形態素解析したカスタマーレビュー \mathbf{R} との類似度 $wrsim(w, \mathbf{R})$ を計算し、それらの値を素性とする。BoW素性では、使用する単語数を5,000や10,000のような大きい数にしても、ほとんどの単語がカスタマーレビュー中に現れないので素性ベクトルは疎ベクトルとなる。したがって、学習に用いる素性ファイルを効率良くファイルに保存したり、少ない計算量で学習を行うことができる。

しかし、本素性の $wrsim(w, \mathbf{R})$ はほとんどの場合で0でない値になるため、 w に使用する単語数を1,000や10,000のような大きい数にすると、素性ベクトルが密となり、素性ベクトルのファイルサイズや学習の計算量が大きくなってしまう。

そこで、本素性の次元数を小さくするために、少ない種類の単語を選択する手法を、次の項で述べる。

(a) 単語感情極性値を利用した単語分散表現素性

まず、単語感情極性辞書の中の最もポジティブな ($s(w)$ の値が最も大きい) 名詞を40単語、形容詞を40単語、動詞を40単語ずつ選択して $wrsim(w, \mathbf{R})$ の値を計算して素性とする。また、同様に最もネガティブな ($s(w)$ の値が最も小さい) 名詞、形容詞、動詞も40単語ずつ選択して $wrsim(w, \mathbf{R})$ の値を計算して素性とする。合計で240種類の単語を選択するので、この素性の次元数は240になる。

(b) 単語重要度を利用した単語分散表現素性

第3.4.3節で示した通り、BoW素性の素性ベクトルが疎であるため、学習が効率よく行える。そこで、BoW素性で事前学習を行い、単語の重要度を計算し、その結果を用いて単語 w を選択することを考える。素性の重要度が算出できる学習器にはRandom ForestやGradient Boostingをはじめ、様々なものがある。

まず、BoW素性を第3.4.1節で示した方法で抽出し、学習器でカスタマーレビューにおける「総合評価」または「満足度」という観点の評価値の推定を行う。「総合評価」または「満足度」という観点はある商品やサービスを総合的に判断した評価値がユーザによって入力されている。次に、BoW素性での学習時に最も重要度が高かった単語の中から名詞を80単語、形容詞を80単語、動詞を80単語ずつ選択する。選んだ単語 w の $wrsim(w, \mathbf{R})$ の値をそれぞれ計算してその値を素性とする。合計で240種類の単語を選択するので、この素性の次元数は240になる。

(c) 観点ごとの単語重要度を利用した単語分散表現素性

第3.4.3節で使用した「総合評価」または「満足度」という観点の評価値は、ある商品やサービスを総合的に判断した評価値であるので、個々の観点に固有の重要な単語が選択されないことがある。例えば、「遠い-形容詞-自立」という単語は「立地」という観点では重要だが、「総合評価」という観点では重要度が低くなってしまふことがある。

そこで、観点ごとに、BoW素性で事前学習するときに推定する評価値の観点を変更する。具体的には、「立地」という観点の単語の重要度を計算するときは同じ「立地」という観点の評価値を推測するための事前学習を行う。

次に、BoW素性での学習時に最も重要度が高かった単語の中から名詞を80単語、形容詞を80単語、動詞を80単語ずつ選択する。選んだ単語 w の $wrsim(w, \mathbf{R})$ の値をそれぞれ計算してその値を素性とする。合計で240種類の単語を選択するので、この素性の次元数は240になる。

3.4.4 文法的な表現を利用した素性

カスタマーレビューの中には、ある商品やサービスが良い・悪いという評価極性だけでなく、様々な情報が述べられている。例えば、「浴衣が2セットあると嬉しいですね」のような改善案や、「昔18インチのテレビを使っていたことがあります」のような個人の事情は、商品やサービスの評価極性を直接述べているわけではないが、それを補助的に説明している。

そこで、評価極性を補助的に説明している部分に着目し、文法的な表現を用いた素性を提案する。条件を表す表現や時間を表す表現など、カスタマーレビューによく使われる文法項目を、日本語能力試験出題基準の3級、4級から選択する。それよりも上の級の表現は、論理的に複雑な文章や抽象度が高い文章を理解するための言語知識が含まれており、カスタマーレビューでは使われない表現が多いため、本研究では3級、4級に絞って表現を選択する。

それぞれの文法的な表現が現れる回数を数えて素性とする。例として、「条件を表す表現(と)」は、以下の3つのいずれかのパターンの単語列がカスタマーレビュー中に現れたときに回数を数える。

- 基本形動詞+接続助詞「と」(例:「浴衣が2セットあると嬉しいですね」)
- 基本形形容詞+接続助詞「と」(例:「欲をいえば部屋はもう少し広いとよかったです」)
- 未然形の単語+助動詞「ない」+接続助詞「と」(例:「予約をしないとこのような最悪な結果になります」)

この表現も含め、次の15種類の表現がカスタマーレビュー中に現れる数を素性として用いる。15種類の表現の現れる回数を素性とするので、この素性の次元数は15となる。

(1) 変化を表す表現(なる, する)

変化を表す典型的な動詞の「~なる」, 「~する」が現れる回数を素性とする。

(2) 変化を表す表現(ようになる, ようにする)

動詞の肯定形で表される、状態の変化を表す「~ようにする」と、ある出来事が起こるように働きかける表現の「~ようにす

る」が現れる回数を素性とする。

(3) 条件を表す表現(と)

反復的、恒常的に成り立つ依存関係を表す際によく用いられる「~と」が現れる回数を素性とする。

(4) 条件を表す表現(ば)

「~と」と同じように、恒常的に成り立つ依存関係を表す際によく用いられる「~ば」が現れる回数を素性とする。「~ば」は「~と」とは異なり、仮定条件によく用いられる。

(5) 条件を表す表現(たら)

特定の、一回的な依存関係を表す際によく用いられる「~たら」が現れる回数を素性とする。

(6) 条件を表す表現(なら)

仮定したことを述べる際によく用いられる「~なら」が現れる回数を素性とする。

(7) 願望を表す表現(ほしい)

希望や願望を表す際によく用いられる「~てほしい」「~ほしい」が現れる回数を素性とする。

(8) 願望を表す表現(たい)

希望や願望を表す際によく用いられる「~たい」が現れる回数を素性とする。「~たい」は「~てほしい」とは異なり、書き手自身の願望を表すためによく用いられる。

(9) 逆接を表す表現(ても)

仮定的な逆接を表す「~ても」が現れる回数を素性とする。

(10) 逆接を表す表現(のに)

事実的な逆接を表す「~のに」が現れる回数を素性とする。

(11) 逆接を表す表現(けれども, けれど, けど, が)

事実的な逆接を表す「~けれども」「~けれど」「~けど」「~が」が現れる回数を素性とする。この4つの表現は書き言葉であるかどうかだけが異なり、意味の違いは無いので同じ素性として回数を数える。

(12) 時間を表す表現(てしまう)

完了を表す「~てしまう」が現れる回数を素性とする。「~てしまう」は書き手の後悔を表すためにも使われることが多いが、今回は完了と後悔の区別はせずに回数を数える。

(13) 時間を表す表現(たことがある)

経験の有無や経歴を述べる表現の「~たことがある」が現れる回数を素性とする。

(14) 時間を表す表現(たことがない)

経験の有無や経歴を述べる表現の「~たことがない」が現れる回数を素性とする。

(15) 時間を表す表現(ことがある)

動作や出来事が行われる場合があるということを表す「~ことがある」が現れる回数を素性とする。

3.4.5 その他の素性

その他の、経験に基づく単純な素性を用いる。この素性の次元数は11となる。

- 文章長
- 単語数
- 名詞-形容動詞語幹の数
- 名詞-サ変接続の数

- その他の名詞の数
- 形容詞の数
- 動詞の数
- 「良い」「よい」「いい」「良かった」のなど回数（「良くなく」などは含まない）
- 「悪い」「わるい」「わるく」などの回数（「悪くなく」などは含まない）
- 「良くない」「よくない」「よくなく」などの回数
- 「悪くない」「わるくない」「悪くなく」などの回数

4. 評価実験

提案手法の有効性を検証するために評価実験を行った。以下、データセット、評価方法、実装、実験結果について順番に述べる。

4.1 評価用データセット

カスタマーレビューの評価値を観点ごとに推定する実験には、2種類のデータセットを使用する。

4.1.1 宿泊施設データセット

1種類目のデータセットとして、楽天株式会社と国立情報学研究所（NII）が協力して提供している「楽天データ公開」³において公開されている、楽天トラベルの「ユーザ評価データ」「レビューデータ」を用いる。

「ユーザ評価データ」には、「立地」「部屋」「食事」「風呂」「サービス」「設備・アメニティ」「総合評価」の7観点について、それぞれ「0」から「5」の評価値がユーザによって投稿されている。「0」という評価値については、「1～5」より低い評価であるという意味ではなく、無評価であるという意味である。評価値の推定において、「0」という値は「1～5」という値と別の性質を持っており、同じ手法で推定することが難しいため、本研究においては、すべての観点について「1～5」が付いているカスタマーレビューに限定する。

実験用のデータセットは次の方法で作成する。まず、カスタマーレビューの単語数が助詞を除き10単語を超え、7観点すべての評価値が「1～5」（「0」を含まない）のものを選択する。次に、それらのカスタマーレビューのレビュアーの重複を除去したうえで、90%のレビュアーを訓練レビュアーとし、残りの10%をテストレビュアーとする。訓練レビュアーが投稿したカスタマーレビュー43,983件を訓練データとし、テストレビュアーが投稿したカスタマーレビュー5,739件をテストデータとする。

各観点の評価値の数を表3と表4に示す。

4.1.2 薄型テレビデータセット

2種類目のデータセットとして、価格.comから収集した薄型テレビデータセットを用いる。このデータセットは、価格.comにおける「薄型テレビ」カテゴリの3,149製品のカスタマーレビュー31,299件を収集したものである。

カスタマーレビューには、カスタマーレビュー本文とともに、「デザイン」「操作性」「画質」「音質」「応答性能」「機能性」「サ

表3 各観点の評価値の数（宿泊施設訓練データ）

| 観点 | 1 | 2 | 3 | 4 | 5 |
|----------|-----|------|-------|-------|-------|
| 立地 | 160 | 1096 | 6087 | 16358 | 20282 |
| 部屋 | 649 | 1971 | 7970 | 16259 | 17134 |
| 食事 | 891 | 2303 | 7969 | 14975 | 17845 |
| 風呂 | 743 | 2443 | 11449 | 14656 | 14692 |
| サービス | 696 | 1304 | 8402 | 15447 | 18134 |
| 設備・アメニティ | 664 | 2278 | 11502 | 16398 | 13141 |
| 総合評価 | 528 | 1218 | 4522 | 19337 | 18378 |

表4 各観点の評価値の数（宿泊施設テストデータ）

| 観点 | 1 | 2 | 3 | 4 | 5 |
|----------|-----|-----|------|------|------|
| 立地 | 24 | 149 | 787 | 2165 | 2614 |
| 部屋 | 90 | 257 | 1074 | 2091 | 2227 |
| 食事 | 128 | 307 | 1062 | 1878 | 2364 |
| 風呂 | 116 | 323 | 1537 | 1911 | 1852 |
| サービス | 77 | 167 | 1105 | 1993 | 2397 |
| 設備・アメニティ | 79 | 284 | 1502 | 2173 | 1701 |
| 総合評価 | 71 | 157 | 603 | 2495 | 2413 |

イズ」「満足度」の8観点について、それぞれ「1」～「5」の評価値がユーザによって投稿されている。なお、評価しない項目については「無評価」という評価値が与えられている。

実験用のデータセットは次のように作成する。まず、カスタマーレビューの単語数が助詞を除き10単語を超え、8観点すべての評価値が「1～5」（「無評価」を含まない）のものを選択する。次に、それらのカスタマーレビューのレビュアーの重複を除去したうえで、90%のレビュアーを訓練レビュアーとし、残りの10%をテストレビュアーとする。訓練レビュアーが投稿したカスタマーレビュー23,482件を訓練データとし、テストレビュアーが投稿したカスタマーレビュー2,593件をテストデータとする。

各観点の評価値の数を表5と表6に示す。

表5 各観点の評価値の数（薄型テレビ訓練データ）

| 観点 | 1 | 2 | 3 | 4 | 5 |
|------|-----|------|------|------|-------|
| デザイン | 121 | 299 | 2303 | 7148 | 13575 |
| 操作性 | 380 | 1006 | 3730 | 8832 | 9498 |
| 画質 | 339 | 616 | 2203 | 6795 | 13493 |
| 音質 | 842 | 1800 | 6345 | 8434 | 6025 |
| 応答性能 | 415 | 1040 | 4212 | 8792 | 8987 |
| 機能性 | 329 | 707 | 3235 | 7296 | 11879 |
| サイズ | 96 | 249 | 2016 | 5506 | 15579 |
| 満足度 | 504 | 614 | 1485 | 6711 | 14132 |

4.2 単語分散表現構築用のデータ

提案素性の抽出や単語感情極性辞書を作成するための単語分散表現の構築には、2種類のデータを使用する。

4.2.1 楽天トラベルデータ

1種類目のデータとして、第4.1.1節で述べた楽天トラベルの「レビューデータ」を再び用いる。宿泊施設データセットの、

(注3) : <http://rit.rakuten.co.jp/opendataj.html>

表6 各観点の評価値の数 (薄型テレビテストデータ)

| 観点 | 1 | 2 | 3 | 4 | 5 |
|------|----|-----|-----|-----|------|
| デザイン | 14 | 43 | 246 | 773 | 1542 |
| 操作性 | 39 | 107 | 399 | 982 | 1091 |
| 画質 | 43 | 80 | 254 | 711 | 1530 |
| 音質 | 98 | 207 | 700 | 964 | 649 |
| 応答性能 | 56 | 118 | 458 | 984 | 1002 |
| 機能性 | 40 | 67 | 348 | 770 | 1393 |
| サイズ | 15 | 41 | 198 | 616 | 1748 |
| 満足度 | 64 | 62 | 167 | 716 | 1609 |

表7 実験に用いるパラメータ

| パラメータ | 宿泊施設 | 薄型テレビ |
|--------------|-------------|--------------|
| m | 3 | 3 |
| $u_1^{(p)}$ | “良い-形容詞-自立” | “良い-形容詞-自立” |
| $u_2^{(p)}$ | “広い-形容詞-自立” | “安い-形容詞-自立” |
| $u_3^{(p)}$ | “近い-形容詞-自立” | “明るい-形容詞-自立” |
| $u_1^{(n)}$ | “悪い-形容詞-自立” | “悪い-形容詞-自立” |
| $u_2^{(n)}$ | “狭い-形容詞-自立” | “高い-形容詞-自立” |
| $u_3^{(n)}$ | “遠い-形容詞-自立” | “暗い-形容詞-自立” |
| 単語分散表現の学習データ | 楽天トラベル | 楽天市場 |

訓練データやテストデータとして使わなかった残り 4,953,422 件のカスタマーレビューを改行区切りで連結したものを使用する。

4.2.2 楽天市場データ

2種類目のデータとして「楽天データ公開」において公開されている楽天市場の「みんなのレビュー・口コミ情報」を用いる。本研究では、商品ジャンルが「TV・オーディオ・カメラ」の商品に関する、2,360,346 件のカスタマーレビューを改行区切りで連結したものを使用する。

4.3 評価方法

提案した素性を用いて、カスタマーレビューの評価値を観点ごとに推定し評価を行う。また、単語分散表現素性、単語感情極性辞書素性、文法的な表現を利用した素性、その他の素性の4つを結合した素性を用いて、カスタマーレビューの評価値を観点ごとに推定し、同様に評価を行う。本研究では、BoW 素性をベースラインとする。

提案素性の評価を行う評価尺度として、MAE (Mean Absolute Error) と MSE (Mean Squared Error) を用いる。それぞれの評価尺度は以下のように定義される。

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \quad (17)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2 \quad (18)$$

ここで、 n はテストデータに含まれるカスタマーレビューの件数、 f_i は予測された評価値、 y_i は真の評価値である。

4.4 実験のパラメータ及び実装

形態素解析器には IPA 辞書と MeCab を用いる。単語分散表現モデルの学習には Stratos ら [9] の Singular を用いる。単語分散表現の導出には Stratos らの知見から、平方根変換と正準相関分析 (CCA) を用いる。その他の Singular のパラメータは、事前に実験を行い MSE が総合的に最も良い結果となった、共起行列のウィンドウ幅 21、単語ベクトルの次元数 300 を使用する。単語感情極性辞書で選択するポジティブ単語、ネガティブ単語の単語数は $k = 1000$ とする。BoW 素性の重みには、事前に実験を行い MSE が総合的に最も良い結果となった Binary を使用する。

各データセットで用いる、式 (1) のパラメータ m , $u_i^{(p)}$, $u_i^{(n)}$ と、単語分散表現の学習に用いるデータを表7に示す。

訓練と予測については、epsilon-SVR を使用する。カーネルは RBF カーネルを使用する。BoW 素性の重要度を求めるときは、Random Forest を推定器として用いる。Random Forest の木の数は 240 本に固定して実験する。素性の重要度の算出には、Gini 係数を使用する。

4.5 評価値の推定結果

それぞれのデータセットでの実験結果を示す。各実験結果の表において、太字で示した値が各評価尺度において最も良かった値である。また、素性の横の数値はその素性の次元数である。

表8と表9に宿泊施設データセットでの実験結果を示す。表8と表9から、宿泊施設データセットにおいては、すべての観点で比較手法よりも提案手法が優れた結果を得られたことがわかる。

表8 宿泊施設データセットの MAE

| 素性 | 立地 | 部屋 | 食事 | 風呂 | サービス | 設備 | 総合評価 |
|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| BoW «6758» | 0.614 | 0.661 | 0.690 | 0.712 | 0.618 | 0.677 | 0.538 |
| 感情極性辞書 «10» | 0.633 | 0.695 | 0.753 | 0.752 | 0.668 | 0.699 | 0.578 |
| 分散表現 (a) «240» | 0.600 | 0.628 | 0.656 | 0.676 | 0.582 | 0.648 | 0.502 |
| 分散表現 (b) «240» | 0.601 | 0.625 | 0.649 | 0.673 | 0.573 | 0.645 | 0.497 |
| 分散表現 (c) «240» | 0.602 | 0.627 | 0.648 | 0.673 | 0.573 | 0.646 | 0.497 |
| 文法的な表現 «15» | 0.658 | 0.721 | 0.772 | 0.782 | 0.698 | 0.722 | 0.601 |
| その他 «11» | 0.660 | 0.729 | 0.780 | 0.773 | 0.703 | 0.706 | 0.622 |
| 結合 (a) «276» | 0.597 | 0.625 | 0.655 | 0.674 | 0.580 | 0.647 | 0.500 |
| 結合 (b) «276» | 0.598 | 0.623 | 0.649 | 0.671 | 0.572 | 0.644 | 0.495 |
| 結合 (c) «276» | 0.598 | 0.625 | 0.648 | 0.671 | 0.572 | 0.644 | 0.495 |

表9 宿泊施設データセットの MSE

| 素性 | 立地 | 部屋 | 食事 | 風呂 | サービス | 設備 | 総合評価 |
|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| BoW «6758» | 0.597 | 0.722 | 0.812 | 0.813 | 0.629 | 0.724 | 0.513 |
| 感情極性辞書 «10» | 0.674 | 0.810 | 0.957 | 0.894 | 0.741 | 0.795 | 0.605 |
| 分散表現 (a) «240» | 0.591 | 0.638 | 0.736 | 0.738 | 0.550 | 0.651 | 0.431 |
| 分散表現 (b) «240» | 0.595 | 0.630 | 0.719 | 0.737 | 0.533 | 0.645 | 0.417 |
| 分散表現 (c) «240» | 0.595 | 0.633 | 0.719 | 0.736 | 0.534 | 0.648 | 0.417 |
| 文法的な表現 «15» | 0.691 | 0.865 | 1.006 | 0.968 | 0.824 | 0.847 | 0.692 |
| その他 «11» | 0.664 | 0.879 | 0.994 | 0.976 | 0.804 | 0.843 | 0.674 |
| 結合 (a) «276» | 0.586 | 0.635 | 0.734 | 0.735 | 0.544 | 0.650 | 0.427 |
| 結合 (b) «276» | 0.590 | 0.627 | 0.719 | 0.734 | 0.529 | 0.646 | 0.414 |
| 結合 (c) «276» | 0.591 | 0.630 | 0.718 | 0.732 | 0.530 | 0.648 | 0.414 |

表10と表11に薄型テレビデータセットでの実験結果を示す。

表 10 と表 11 から、薄型テレビデータセットでは、すべての観点で比較手法よりも提案手法が優れた結果を得られたことがわかる。

表 10 薄型テレビデータセットの MAE

| 素性 | デザイン | 操作性 | 画質 | 音質 | 応答性能 | 機能性 | サイズ | 満足度 |
|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| BoW «6986» | 0.553 | 0.675 | 0.610 | 0.773 | 0.705 | 0.656 | 0.486 | 0.567 |
| 感情極性辞書 «10» | 0.566 | 0.689 | 0.620 | 0.789 | 0.722 | 0.668 | 0.489 | 0.571 |
| 分散表現 (a) «240» | 0.546 | 0.653 | 0.584 | 0.758 | 0.689 | 0.621 | 0.471 | 0.532 |
| 分散表現 (b) «240» | 0.544 | 0.651 | 0.581 | 0.758 | 0.688 | 0.620 | 0.469 | 0.526 |
| 分散表現 (c) «240» | 0.546 | 0.653 | 0.581 | 0.758 | 0.689 | 0.618 | 0.469 | 0.526 |
| 文法的な表現 «15» | 0.572 | 0.711 | 0.640 | 0.809 | 0.736 | 0.728 | 0.490 | 0.593 |
| その他 «11» | 0.575 | 0.706 | 0.630 | 0.795 | 0.729 | 0.716 | 0.490 | 0.588 |
| 結合 (a) «276» | 0.547 | 0.650 | 0.578 | 0.755 | 0.687 | 0.620 | 0.471 | 0.529 |
| 結合 (b) «276» | 0.544 | 0.649 | 0.576 | 0.755 | 0.686 | 0.621 | 0.468 | 0.523 |
| 結合 (c) «276» | 0.546 | 0.650 | 0.576 | 0.756 | 0.688 | 0.619 | 0.469 | 0.523 |

表 11 薄型テレビデータセットの MSE

| 素性 | デザイン | 操作性 | 画質 | 音質 | 応答性能 | 機能性 | サイズ | 満足度 |
|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| BoW «6986» | 0.668 | 0.742 | 0.833 | 0.999 | 0.842 | 0.717 | 0.672 | 0.812 |
| 感情極性辞書 «10» | 0.739 | 0.777 | 0.884 | 1.083 | 0.888 | 0.842 | 0.674 | 0.830 |
| 分散表現 (a) «240» | 0.569 | 0.681 | 0.666 | 0.904 | 0.762 | 0.663 | 0.598 | 0.597 |
| 分散表現 (b) «240» | 0.568 | 0.673 | 0.657 | 0.896 | 0.755 | 0.660 | 0.594 | 0.563 |
| 分散表現 (c) «240» | 0.569 | 0.677 | 0.660 | 0.901 | 0.758 | 0.659 | 0.592 | 0.563 |
| 文法的な表現 «15» | 0.803 | 0.847 | 1.082 | 1.120 | 0.932 | 0.866 | 0.689 | 1.026 |
| その他 «11» | 0.768 | 0.835 | 0.985 | 1.060 | 0.913 | 0.855 | 0.689 | 0.939 |
| 結合 (a) «276» | 0.573 | 0.674 | 0.660 | 0.898 | 0.757 | 0.663 | 0.593 | 0.596 |
| 結合 (b) «276» | 0.570 | 0.666 | 0.652 | 0.892 | 0.751 | 0.663 | 0.588 | 0.560 |
| 結合 (c) «276» | 0.572 | 0.670 | 0.653 | 0.898 | 0.755 | 0.661 | 0.587 | 0.560 |

5. おわりに

本研究では、宿泊施設、薄型テレビのカスタマーレビューを用いて、観点ごとの評価値の推定に取り組んだ。各データセットごとに単語分散表現モデルを学習し単語感情極性辞書を構築し、辞書に含まれる単語と単語感情極性値を用いる素性を提案した。また、単語とカスタマーレビューの類似度を定義し、それを用いる、単語分散表現素性を提案した。単語を選択する方法として、単語感情極性辞書を用いる方法、事前学習した際の単語の重要度を用いる方法、事前学習した際の単語の重要度を個々の観点ごとに用いる方法を提案した。さらに、条件を表す表現や時間を表す表現などの、文法的な表現を用いる素性を提案した。これらの素性を用いて、SVR を使用して、カスタマーレビューからの評価値の推定を行った。

評価実験の結果、提案素性を用いることで、推定結果が比較手法より正確になることを確認した。宿泊施設データセットにおいては、すべての観点で、推定結果がより正確になることを確認した。薄型テレビデータセットにおいても、すべての観点で、推定結果がより正確になることを確認した。また、すべての素性を結合することで、より推定結果が正確になることを確認した。

今後の課題としては、推定結果をより正確にするために、大

別して評価値が「4」以上の場合の推定の改良と、評価値が「3」以下の場合の推定の改良が挙げられる。評価値が「4」以上の場合の推定の改良のためには、単語分散表現素性において単語を選択する方法をさらに工夫することが挙げられる。評価値が「3」以下の場合の推定の改良のためには、文単位やフレーズ単位で評価値の推定を行ってから、それらを統合する手法を考案することが課題となる。

謝 辞

本研究の一部は、科研費基盤 (B) (課題番号 26280038) の支援を受けて遂行した。

文 献

- [1] Soujanya Poria, Erik Cambria, and Alexander F. Gelbukh. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowl.-Based Syst.*, Vol. 108, pp. 42–49, 2016.
- [2] Cícero Nogueira dos Santos and Maira Gatti. Deep convolutional neural networks for sentiment analysis of short texts. In Jan Hajic and Junichi Tsujii, editors, *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*, pp. 69–78. ACL, 2014.
- [3] Hiroya Takamura, Takashi Inui, and Manabu Okumura. Extracting semantic orientations of words using spin model. In Kevin Knight, Hwee Tou Ng, and Kemal Oflazer, editors, *ACL 2005, 43rd Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, 25-30 June 2005, University of Michigan, USA*. The Association for Computer Linguistics, 2005.
- [4] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *CoRR*, Vol. abs/1301.3781, , 2013.
- [5] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pp. 3111–3119, 2013.
- [6] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pp. 1532–1543. ACL, 2014.
- [7] Bing Liu, Mingqing Hu, and Junsheng Cheng. Opinion observer: analyzing and comparing opinions on the web. In Allan Ellis and Tatsuya Hagino, editors, *Proceedings of the 14th international conference on World Wide Web, WWW 2005, Chiba, Japan, May 10-14, 2005*, pp. 342–351. ACM, 2005.
- [8] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. Sentiment in short strength detection informal text. *JASIST*, Vol. 61, No. 12, pp. 2544–2558, 2010.
- [9] Karl Stratos, Michael Collins, and Daniel J. Hsu. Model-based word embeddings from decompositions of count matrices. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pp. 1282–1291. The Association for Computer Linguistics, 2015.