

# 多因子単語埋め込みを用いる複数分野感情極性の転移学習

田 然<sup>1</sup> 渡邊 研斗<sup>1</sup> 乾 健太郎<sup>1,2</sup>

<sup>1</sup> 東北大学 <sup>2</sup> 理研 AIP

{tianran, kento.w, inui}@ecei.tohoku.ac.jp

## 1 はじめに

現在広く使われている単語埋め込みの多くは、周辺文脈から単語の分布を予測することでその単語の意味を表現する [6,8]。しかし、周辺文脈以外にも単語分布に影響する因子はたくさんあり、それらが複雑に絡み合っている。例えば人間が商品レビューを書く時、流暢な文にするため前後の文脈を考慮して単語を選ぶが、他にも商品のカテゴリや商品に対する感情極性（評価スコア）などによって違う単語を使う。本稿は、これらの因子を統合する**多因子単語埋め込み**の枠組みを提案し、いくつかのカテゴリの中の商品レビューと評価スコアから学習した単語埋め込みを用いて、複数の別カテゴリの中の商品レビューの評価スコアを高精度に予測することで、提案枠組みの有用性を示す。

以下我々のアイデアを定式化する。単語分布に影響する因子（周辺単語、商品カテゴリ、評価スコア、など）を  $F_1, \dots, F_n$  とおく。我々の仮定は、単語を生成する複数のプロセス  $\tau_1, \dots, \tau_m$  が存在し、全ての単語はその内の**どれか一つ**によって生成されているとする。そして、それぞれのプロセスは限られた因子だけにしか依存しない。そうして複数の因子が絡み合った構造を表現し、埋め込みモデルを構築する。

具体的に、商品レビューに書かれた単語の分布に影響する因子として我々は以下を考える：

$F_1$ : 周辺文脈にある他の単語  $\in \{glass, quality, \dots\}$

$F_2$ : 商品に対する評価スコア  $\in \{1, 2, 3, 4, 5\}$

$F_3$ : 商品のカテゴリ  $\in \{books, dvd, \dots\}$

$F_4$ : レビューの文書 ID  $\in \{Doc1, Doc2, \dots\}$

因子  $F_4$  にレビューの文書 ID を考えたのは、同じ文書に一貫した「トピック」が単語分布に影響するからである。そして、次の単語生成プロセスを考える：

$\tau_1$ : 周辺文脈にある他の単語（アスペクト）に対して感情極性を述べる

$\tau_2$ : 商品の評価に関連する文書トピックを述べる

$\tau_3$ : 商品のカテゴリに関連する文書トピックを述べる

我々は、商品レビューに使われた全ての単語がこの3つのプロセスのどれかによって説明できる、という強い仮定を置く。すると単語  $w$  の生成確率は

$$P(w) = P(w, \tau_1) + P(w, \tau_2) + P(w, \tau_3)$$

と分解でき、しかもプロセス  $\tau_1$  は  $F_1, F_2$  に、 $\tau_2$  は  $F_2, F_4$  に、 $\tau_3$  は  $F_3, F_4$  にしか依存しない。よって

$$\begin{aligned} P(w | F_1, F_2, F_3, F_4) \\ = P(w, \tau_1 | F_1, F_2) + P(w, \tau_2 | F_2, F_4) \\ + P(w, \tau_3 | F_3, F_4) \quad (1) \end{aligned}$$

は多因子下の単語生成確率を与える。単語や各因子をベクトルで表現し、評価スコア・カテゴリ付きの商品レビューコーパスから最尤推定を行えば、単語・因子の埋め込みが得られる。そして学習された単語埋め込みを引き継ぎ、評価スコアが未知かつカテゴリが見えない商品レビューに対して、未知スコアのベクトルと未見カテゴリのベクトルの最尤推定を行えば、単語埋め込みを介した転移学習となる。未知の評価スコアの予測は、推定されたベクトルと既知の評価スコアベクトルとの類似度計算で行える。手法の詳細は3節で述べる。

25 カテゴリの商品に対するレビューと評価スコアを集めた Multi-Domain Sentiment Dataset [1] を用いて、我々はその内 12 カテゴリを転移元、残り 13 カテゴリを転移先にし転移学習実験を行なった。結果、ポジネガの2値分類では F 値が 70 を超え、同じ問題に取り組んだ吉田らの手法 [11] を大きく上回った。定性分析から、我々のモデルは **(a)** プロセス  $\tau_1$  で感情極性に依存して変化する単語共起を捉えた、**(b)** プロセス  $\tau_2$  と  $\tau_3$  を分けることで単語の「感情的類似性」と「分野的類似性」を分けて捉えた、**(c)** 3つのプロセスは共同で極性推定に寄与したことを実例で示す。

## 2 関連研究

文書の感情極性を推定し、それを基に商品などに対する評判を集めることが実用上重要なタスクである。文書に出てくる単語は、感情極性推定のための有用な手がかりになるが、分野によってどの単語がどの極性を示唆するかは変わる [4]。一方、分野によって得られる訓練データの量も大きく変わるので、いくつかの分野で訓練したモデルを他分野に適用する転移学習は、多分野を跨ぐ感情極性分析における中心的チャレンジである [1,5,10]。この問題に取り組んだ吉田らの提案 [11] は、感情極性が分野ラベル（商品カテゴリ）に依存する単語としない単語を分けるベイジアンモデルである。しかしこのモデルは商品カテゴリ・感情極性・単語文脈など多因子の相互作用を完全精密に捉えたとは言いつれず、実際モデルの推定は単語を形容詞と副詞だけに限定して行わなければならなかった。対して我々の提案は人間の直感に沿ったプロセス  $\tau_1, \tau_2, \tau_3$  を基にしており、多因子の絡み合いにより一歩踏み込んだと言える。また、単語の埋め込みに周辺文脈以外の因子として文書 ID を取り入れた研究 [9] や、単純で高精度な文書極性分類モデル [3] などがあるが、紙面の関係でこれらとの比較・評価は将来の発表に譲る。最後に、本稿と似たアイデアで話題と言葉遣いの違いを潜在トピックモデルで捉えた、渡邊ら [12] の研究を挙げたい。

## 3 手法

式(1)から出発して我々の埋め込みモデルを記述する。全単語の語彙集合を  $V = \{dog, cat, \dots\}$  とおく。単語  $w \in V$  に対して、生成プロセス  $\tau_1, \tau_2, \tau_3$  のそれぞれに関わるベクトル  $\mathbf{v}_1(w), \mathbf{v}_2(w), \mathbf{v}_3(w)$  がある。そして、プロセス  $\tau_i$  に関わる因子  $F_j$  が値  $f$  をとる時、ベクトル  $\mathbf{u}_i^j(f)$  を割り当てる。例えば商品レビュー Doc1 が商品に対してスコア 5 を付けた時、プロセス  $\tau_2$  がベクトル  $\mathbf{u}_2^5(5)$  と  $\mathbf{u}_2^4(\text{Doc1})$  をもって単語を生成する。単語  $w$  や値  $f$  を強調しない時は  $\mathbf{v}_i, \mathbf{u}_i^j$  のように略す。

### 3.1 多因子単語埋め込み

式(1)の右辺の項を以下のようにモデル化する：

$$P(w, \tau_1 | F_1, F_2) = \sum_{c \in \text{文脈}} \exp(\mathbf{v}_1 \cdot (\mathbf{u}_1^1(c) * \mathbf{u}_1^2(c)))$$

$$P(w, \tau_2 | F_2, F_4) = \exp(\mathbf{v}_2 \cdot [\mathbf{u}_2^5 : \mathbf{u}_2^4])$$

$$P(w, \tau_3 | F_3, F_4) = \exp(\mathbf{v}_3 \cdot [\mathbf{u}_3^3 : \mathbf{u}_3^4])$$

ここで  $\mathbf{u}_1^1 * \mathbf{u}_1^2$  は循環畳み込み積、そして  $[\mathbf{u}_2^5 : \mathbf{u}_2^4]$  と  $[\mathbf{u}_3^3 : \mathbf{u}_3^4]$  はベクトルの連結を表す。循環畳み込み積は

次のように定義された演算である：

$$(\mathbf{a} * \mathbf{b})_i := \sum_{j+k \equiv i \pmod{D}} (\mathbf{a})_j (\mathbf{b})_k$$

ただし  $(\mathbf{a})_i$  はベクトル  $\mathbf{a}$  の  $i$ -番目の成分、 $D$  は  $\mathbf{a}$  の次元を表す。 $P(w, \tau_1 | F_1, F_2)$  をこのようにモデル化する動機は、同じ単語でも周辺のアスペクトによって極性が変わることがあるからである。例えば単語 *high* は *high prices* の時ネガティブから生成されやすいが、*high quality* の時ポジティブから生成されやすい。このように単語の生成確率はアスペクトにも感情極性にも同時に依存しているため、両因子の相互作用をモデル化できる（と思われる）循環畳み込み積を用いた。

一方  $P(w, \tau_2 | F_2, F_4)$  に関して、因子  $F_2$ （評価スコア）は因子  $F_4$ （レビュー文書）から一意に決まるので、実質  $F_4$  が定める 1 個のベクトルをもって単語が生成されると考えられる。ただし、同じ評価スコアを持つ文書同士はそのベクトルの一部を互いにシェアしていると考えるのが自然で、これは  $P(w, \tau_2 | F_2, F_4)$  のモデル化にベクトルの連結  $[\mathbf{u}_2^5 : \mathbf{u}_2^4]$  を用いた動機である。 $P(w, \tau_3 | F_3, F_4)$  についても同様に考えられる。

更に、各プロセスを内積の指数関数（の和）で表現することで、内積が少しでも大きいほうの指数関数は他の値を大きく凌駕する Softmax 的な効果が働く。これは、単語生成が**どれか一つ**のプロセスによって説明できるとする我々の仮定に合致する。

提案モデルの推定は Noise Contrastive Estimation [2] で行われる。これは [6,7] などに広く使われた手法で、文書中の各単語  $w$  に対して負例  $n_1, \dots, n_k$  を unigram 分布からサンプリングし、負例を混ぜたデータの尤度を

$$\begin{aligned} & P(w \text{ が正例}, n_1, \dots, n_k \text{ が負例} | F_1, F_2, F_3, F_4) \\ &= \frac{P(w | F_1, F_2, F_3, F_4)}{k + P(w | F_1, F_2, F_3, F_4)} \\ & \quad \cdot \prod_{i=1}^k \frac{k}{k + P(n_i | F_1, F_2, F_3, F_4)} \end{aligned}$$

で計算し、最大化する手法である。

### 3.2 転移学習

複数分野感情極性の転移学習において、転移先にあるのは、評価スコア未知、かつ訓練時に全く見たことのないカテゴリの商品に対するレビューである。転移元から学習された、感情極性に関する言語知識と、転移先の商品レビュー文をもって評価スコアを予測しなければならない。この転移学習を行うために、転移元で

| 転移元                    |         | 転移先                   |       |
|------------------------|---------|-----------------------|-------|
| books                  | 975172  | video                 | 36180 |
| music                  | 174179  | baby                  | 4256  |
| dvd                    | 124435  | magazines             | 4191  |
| electronics            | 23009   | beauty                | 2884  |
| kitchen & housewares   | 19856   | software              | 2390  |
| toys & games           | 13147   | jewelry & watches     | 1981  |
| apparel                | 9252    | outdoor living        | 1599  |
| camera & photo         | 7408    | gourmet food          | 1575  |
| health & personal care | 7225    | cell phones & service | 1023  |
| sports & outdoors      | 5727    | automotive            | 736   |
| computer & video games | 2771    | office products       | 431   |
| grocery                | 2632    | musical instruments   | 332   |
|                        |         | tools & hardware      | 112   |
| 合計                     | 1364813 | 合計                    | 57690 |

表1 転移元・転移先のカテゴリ・レビュー数

訓練した多因子単語埋め込みを引き継ぎ、転移先のデータを転移元に混ぜてもう一回モデル推定を行う。ただし転移先の評価スコアが未知なため、それを文書 ID で「この文書の評価スコア」という仮のスコアとして代用する。また、カテゴリも未見なので、転移先の因子  $F_2, F_3, F_4$  のベクトルは全てランダム初期化から推定する。語彙  $V$  と因子  $F_1$  (周辺単語) に割り当てられたベクトルだけが、転移元から引き継いで初期化する。

転移先で推定された、文書 ID の仮スコアベクトルと、転移元から引き継いだ、既知のスコアベクトルとの類似度を計算することで、文書の評価スコアを予測する。具体的に、文書 ID を  $x$  とした時の予測スコアを

$$\arg \max_{y \in \{1,2,3,4,5\}} \cos(\mathbf{u}_1^2(y), \mathbf{u}_1^2(x)) + \cos(\mathbf{u}_2^2(y), \mathbf{u}_2^2(x))$$

で計算する。ここで  $\cos$  はコサイン類似度を表す。

## 4 実験

Multi-Domain Sentiment Dataset<sup>\*1</sup>は 25 カテゴリの商品に対して 1.4M のレビュー文書と評価スコアを集めたコーパスである。各カテゴリにあるレビュー数とカテゴリ間の関連性を考慮して、我々は転移元・転移先を表1のように分けた。分ける際、転移元でレビュー数の多いかつ多様なカテゴリをカバーするよう配慮した。これは、学習データが豊富なカテゴリで訓練し、レビュー数の少ない別カテゴリに対しては他のリソース (SNS など) から得られる (評価スコア未知の) テキストデータで評判分析を行う際に現実的な設定である。

テキストは前処理で句読点を除き、数字を 0 に置き換え、全大文字以外の単語を小文字にした。また、否定

<sup>\*1</sup> [www.cs.jhu.edu/~mdredze/datasets/sentiment/](http://www.cs.jhu.edu/~mdredze/datasets/sentiment/)

|          | 2-Label      | 4-Label      |
|----------|--------------|--------------|
| マジョリティー  | 45.04        | 18.46        |
| ランダム     | 44.31        | 21.32        |
| 吉田ら [11] | 49.26        | —            |
| 提案モデル    | <b>70.26</b> | <b>41.96</b> |

表2 評価スコア予測の Macro-F1 値

語 *not* を右に一番近い内容語とくっ付く (例えば *not the book* を *the not\_book* にし、*not\_book* を 1 単語と見なすことで否定を扱う) 変換を行なった。

埋め込みの次元数は、 $\mathbf{u}_1^1, \mathbf{u}_1^2, \mathbf{u}_2^2, \mathbf{u}_2^4, \mathbf{u}_3^3, \mathbf{u}_3^4$  に対してそれぞれ 256, 64, 64, 64, 64, 64 にした。周辺文脈は平均長さ 3 でサンプリングした窓幅内の単語とし、また左側の文脈と右側の文脈を区別した (例えば単語 *beer* の文脈が “*glass beer*” なら *glass\_*、 “*beer glass*” なら *\_glass* と、左側にある *glass* と右側にある *glass* を違う単語として区別する)。訓練は負例数を  $k = 3$  にし、最適化は SGD を用いた。学習はまず転移元で 4 エポック行い、そして転移先を転移元に混ぜたデータで 16 エポック行なった。

### 4.1 評価スコア予測

実験では、評価スコアのラベル  $\{1, 2, 4, 5\}$  を予測する 4 値分類と、 $\{1, 2\}$  をネガティブ、 $\{4, 5\}$  をポジティブとする 2 値分類で性能を比較した。正解ラベルの分布が偏っている<sup>\*2</sup> ため、評価指標は Macro-F1 を用いた。比較したのはマジョリティーとランダム・ベースライン、そして吉田らのモデル [11] の再実装である。結果を表2に示す。提案手法はベースラインを大きく上回り、多因子単語埋め込みの強い汎化性能を示唆した。

### 4.2 定性分析

感情極性によって変化する共起単語の分布は、循環量み込み  $\mathbf{u}_1^1 * \mathbf{u}_1^2$  で捉えたかを調べるため、いくつかの文脈  $c$  に対して  $\mathbf{u}_1^1(c) * \mathbf{u}_1^2(1)$  と  $\mathbf{u}_1^1(c) * \mathbf{u}_1^2(5)$  を計算し、それらと  $\mathbf{v}_1(w)$  の内積が上位の  $w$  を表3で示した。感情極性がネガティブ (1) の時とポジティブ (5) の時を比較すると、例えば文脈 *\_weak* の場合、ネガティブの *terribly weak, pathetically weak* とポジティブの *n't too weak, albeit weak*; 及び *book\_* の場合ネガティブの *book stinks, book unreadable* とポジティブの *book inspires, book invaluable* と差がはっきり現れた。

次に、単語ベクトルがプロセス  $\tau_2$  に関わる  $\mathbf{v}_2$  と  $\tau_3$  に関わる  $\mathbf{v}_3$  に分かれたことに注目し、この二つのベクトルが単語意味の違う側面を捉えたかを調べるため、

<sup>\*2</sup> 転移先でラベル 5 は 59%、 $\{4, 5\}$  で 82% の割合を占める。

| _weak (1) _weak (5) |             | _quality (1) _quality (5) |             | book_ (1) book_ (5) |                | fresh   |               | error       |             |
|---------------------|-------------|---------------------------|-------------|---------------------|----------------|---------|---------------|-------------|-------------|
|                     |             |                           |             |                     |                | 感情的 分野的 |               | 感情的 分野的     |             |
| weak                | weak        | redeeming                 | quantity    | rehashes            | signings       | 感情的     |               | 感情的         |             |
| awfully             | awfully     | grainy                    | sound       | misrepresents       | contains       | 分野的     |               | 感情的         |             |
| irritatingly        | notoriously | poor                      | grainy      | not_helpful         | addresses      |         | cherry        | not_play    | contact     |
| lame                | n't_too     | quantity                  | high        | stinks              | previews.com   |         | freshness     | not_sell    | tracking    |
| shallow             | albeit      | mediocre                  | transfer    | unreadable          | inspires       |         | ideal         | not_sweet   | spool       |
| terribly            | damned      | shoddy                    | audio/video | not_informative     | clarifies      |         | staple        | beautifully | not_nothing |
| pathetically        | not_single  | subpar                    | improved    | purports            | accomplishes   |         | appropriately | freshest    | amazonians  |
| laughably           | kinda       | high                      | pristine    | BORING              | re-enforces    |         | bewitched     | bec         | steriods    |
| frustratingly       | admittedly  | crummy                    | production  | touts               | covers         |         | lovely        | brie        | tacky       |
| appallingly         | seeming     | not_high                  | not_high    | panders             | invaluable     |         | describe      | n't_sweet   | not_display |
| ludicrously         | pretty      | inferior                  | audio       | not_worth           | summarizes     |         | determined    | tapioca     | fron        |
| horrendously        | losers      | sound                     | quality     | advertises          | teaches        |         | unlikely      | peaches     | NOT_same    |
| unbelievably        | not_overly  | reproduction              | highest     | not_useful          | cover-to-cover |         | splendidly    | topping     | flattens    |
| severely            | somewhat    | amateurish                | produces    | not_recommended     | reaffirms      |         | natural       | spicy       | topmost     |
|                     |             |                           |             |                     |                |         |               |             | problem     |

表3 評価スコアを考慮した共起単語の上位リスト

表4 感情的・分野的類似単語の上位リスト

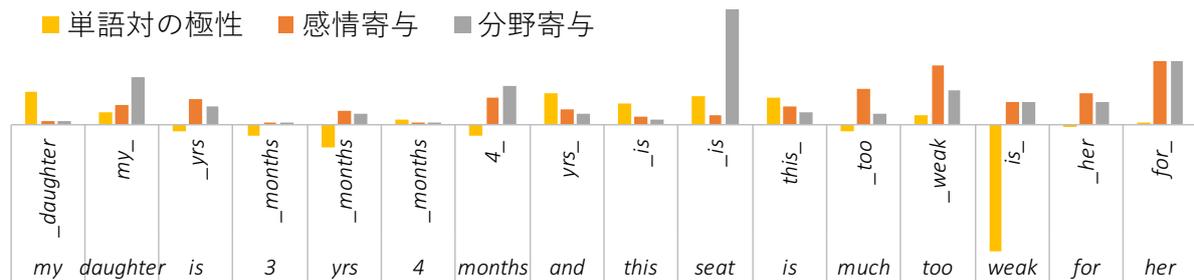


図1 baby カテゴリにあるレビューの一文に対する推定結果

$v_2$  の「感情的部分」(=評価スコアベクトル  $u_2^2$  と内積をとる部分) のコサイン類似度を使って得られる感情的類似単語リストと、 $v_3$  の「分野的部分」(=カテゴリベクトル  $u_3^3$  と内積をとる部分) を使った分野的類似単語リストを表4で示した。結果、*fresh* は同じくポジティブな感情を惹き起こす *bewitched*, *lovely* と感情的に似ているが、食品関連の *cherry*, *not\_sweet* と分野的に似ている傾向が見られた。同様に、*error* は *not\_play*, *not\_sell* など実際の故障と感情的に似ているが、*tracking*, *operating* など操作的用語と分野的に似ている。また、*fresh* も *error* も**転移先**のカテゴリ(それぞれ gourmet food と software) と一番共起が強いことに注意されたい。これは、感情別・分野別の傾向が転移学習された証拠の一つと思われる。

更に、我々のモデルで  $P(w, \tau_2 | F_2, F_4)$  (感情寄与の確率) と  $P(w, \tau_3 | F_3, F_4)$  (分野寄与の確率)、及び  $E[F_2 | F_1, w, \tau_1]$  (単語とそれに一番強い共起を持つ文脈のペアに対して評価スコアの期待値) を推定できる。図1にこれらの推定結果の一例を示した(評価スコアの期待値は -3 シフトして「単語対の極性」として表示する)。図には、*daughter*, *seat* は分野性が強い、*much*, *too* は感情性が強い、そして *is\_weak* が強いネガティブ性を持つという、人間の直感に近い推定が得られた。このレビューのスコアは 2 と正しく予測されている。

## 5 おわりに

単語分布に影響する複数の要因を埋め込みに取り組み手法として、我々は多因子単語埋め込みを提案した。感情極性の転移学習において楽観的結果が得られたが、これからの更なる実験・検証が求められる。

**謝辞** 本研究は JST CREST(課題番号: JPMJCR1301) の支援を受けて行った。

## 参考文献

- [1] J. Blitzer, M. Dredze, and F. Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, 2007.
- [2] M. Gutmann and A. Hyvärinen. Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. *J. Mach. Learn. Res.*, 13(1), 2012.
- [3] M. Iyyer, V. Manjunatha, J. Boyd-Graber, and H. Daumé. Deep unordered composition rivals syntactic methods for text classification. In *ACL*, 2015.
- [4] B. Liu. *Sentiment Analysis: mining sentiments, opinions, and emotions*. Cambridge University Press, 2015.
- [5] Y. Mansour, M. Mohri, and A. Rostamizadeh. Domain adaptation with multiple sources. In *NIPS*, 2009.
- [6] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *NIPS*, 2013.
- [7] A. Mnih and K. Kavukcuoglu. Learning word embeddings efficiently with noise-contrastive estimation. In *NIPS*, 2013.
- [8] J. Pennington, R. Socher, and C. Manning. Glove: Global vectors for word representation. In *EMNLP*, 2014.
- [9] F. Sun, J. Guo, Y. Lan, J. Xu, and X. Cheng. Learning word representations by jointly modeling syntagmatic and paradigmatic relations. In *ACL*, 2015.
- [10] F. Wu and Y. Huang. Sentiment domain adaptation with multiple sources. In *ACL*, 2016.
- [11] Y. Yoshida, T. Hirao, T. Iwata, M. Nagata, and Y. Matsumoto. Transfer learning for multiple-domain sentiment analysis — identifying domain dependent/independent word polarity. In *AAAI*, 2011.
- [12] 渡邊研斗, 赤間怜奈, 横井祥, 田然, and 乾健太郎. 話題と言葉遣いの違いを考慮した潜在トピックモデル. In 言語処理学会年次大会, 2018.