



図 3 : TE-CNN のモデル

4. 実験

本研究では、TE-CNN で属性語の抽出を行い、DE-CNN との比較を行う。

4.1 実験データ

ミステリー小説 1 作品のレビューに対し、評価を述べている文章 1000 文を手で選んだ。このうち 800 文を学習データ、200 文をテストデータとして用いた。DE-CNN に使う general embedding は Wikipedia のコーパスから獲得し、domain embedding は学習データ 800 件から獲得した。また、各文章は属性語がないものや複数出現するものがあり、分布は表 1 で表す通り、属性語が出現しない文が 447 文、1 つ出現するものが 420 文、複数出現する文が 133 文存在する。

N	0	1	2	3	4	6
属性語が N 個出現する文の数	447	420	105	23	4	1

表 1 : 属性語の分布

4.2 実験結果

抽出した属性語を Precision, Recall および F 値で評価した。実験の結果、DE-CNN および TE-CNN の各値は表 2 に示す。

モデル	Precision	Recall	Score-F
DE-CNN	51.3	65.8	57.6
TE-CNN	59.5	69.1	64.0

表 2 : DE-CNN と TE-CNN の実験結果

品詞情報を含んだ TE-CNN は品詞情報を含まない DE-CNN と比べて Precision が 8.2、Recall が 3.3 向上している。合わせて F 値も 6.4 向上しており、総合的に品詞情報を用いたモデルが好結果を示している。

5. 考察

レビューから DE-CNN と TE-CNN が抽出した結果の比較を図 4 および図 5 に示す。下線の単語が正解の属性語、括弧内がそれぞれのモデルが抽出した語である。図 4 において DE-CNN が抽出できなかった「内容」という語を TE-CNN では抽出できている。すなわち、Recall における向上例である。図 5 においては、DE-CNN が誤って抽出していた「読ん」という語を TE-CNN は抽出することなく、Precision を向上させている。品詞情報は、Recall と Precision の双方に良い影響があると考えられる。

DE-CNN : 読み やすく <u>内容</u> にも大満足です、読み終わった直後に即、再読しました。
TE-CNN : 読み やすく (内容) にも大満足です、読み終わった直後に即、再読しました。

図 4 : Recall に寄与した抽出例

DE-CNN : (読ん) でいるうちに自然と考えさせられる (文章) でした
TE-CNN : 読ん でいるうちに自然と考えさせられる (文章) でした

図 5 : Precision に寄与した抽出例

6. おわりに

本稿では、レビュー俯瞰のための属性語抽出法である DE-CNN に対し、品詞情報の埋め込みを追加した Triple Embeddings CNN の提案を行った。実験の結果、従来手法から precision を 8.2、recall を 3.3 向上させ、品詞情報を付加した手法の有効性を示した。獲得された品詞の出現パターンについては未検証であり、今後の課題となる。

参考文献

- [1] 中野裕介, 湯本高行, 新居学, 上浦尚武, “機械学習による商品レビューの属性-意見ペアの抽出”, 研究報告書データベース(DBS) Vol.2015-DBS-162, No.14, pp.1-8(2015)
- [2] 山西良典, 古田周史, 福本淳一, 西原陽子, “出現頻度と構文特徴を用いたレビュー構造の俯瞰のための評価視点の抽出”, 知能と情報(日本知能ファジィ学会誌) Vol.27, No.1, pp.501-511(2015)
- [3] Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu, “Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction”, In ACL (2018)
- [4] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, “意見抽出のための評価表現の収集”, 自然言語処理, Vol.12, No.3, pp.203-222(2005)
- [5] 駒田康孝, 山名早人, “商品評価ツイートからの属性語自動抽出手法の提案”, DEIM Forum 2014, B5-6, (2014)
- [6] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov, “Enriching Word Vectors with Subword Information” arXiv:1607.04606(2016)

† 琉球大学 University of The Ryukyus