Human-level Semantic Score Prediction from dialogue using BERT

20225168 JaeYeon Bae
20225103 Chaeri Kim

Contents

Introduction

- Sentiment analysis is one of the most famous topic in natural language processing (NLP), used to determine whether data is positive, negative or neutral.
- Sentiment analysis researches are actively conducted using product reviews or movie reviews and so on.
- However, most of them are based on a single sentence.

Star Level	General Meaning
★	I hate it.
★★	I don't like it.
★★★	It's okay.
★★★★	I like it.
★★★★★	I love it.

Figure 1. Rating system for Amazon.com

Table 1. Top 10 sentiment phrases based on occurrence

Phrase	Type	Occurrence
not worth	NOA	26329
not go wrong	NOA	15446
not bad	NOA	15122
not be happier	NOA	14892
not good	NOA	12919
don't like	NOV	42525
didn't work	NOV	38287
didn't like	NOV	21806
don't work	NOV	10671
don't recommend	NOV	9670

Introduction

- There are few researches about sentiment analysis based on dialogue.
- Unlike prior researches that analyze emotions in a sentence, we would like to analyze emotions between people through sentiment analysis based on conversation.
- Our project focuses on human-level sentiment analysis.

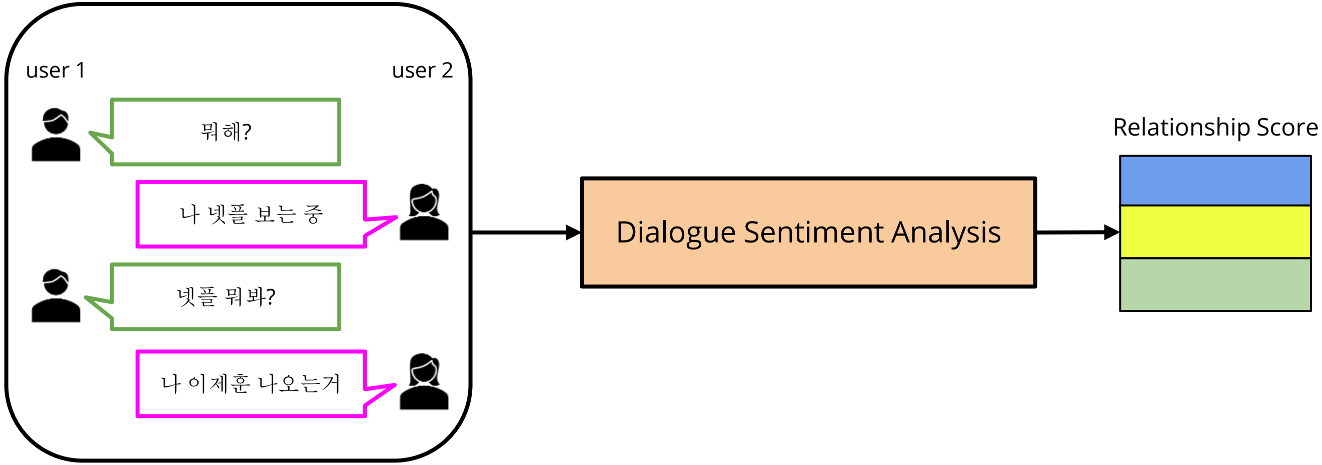


Figure 2. Brief model of our human-level sentiment analysis approach

Related Work

Bert : Pre-training of deep Bidirectional Transformers for Language Understanding

- Pre-trained NLP model which uses several transformers to deal with # of downstream tasks
- Outperforms in most NLP tasks (Sentence pair classification, Single sentence classification, QA tasks, Single sentence tagging task)

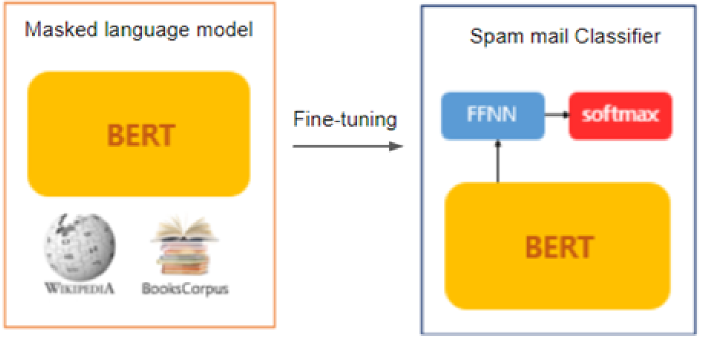


Figure 3. How to use BERT

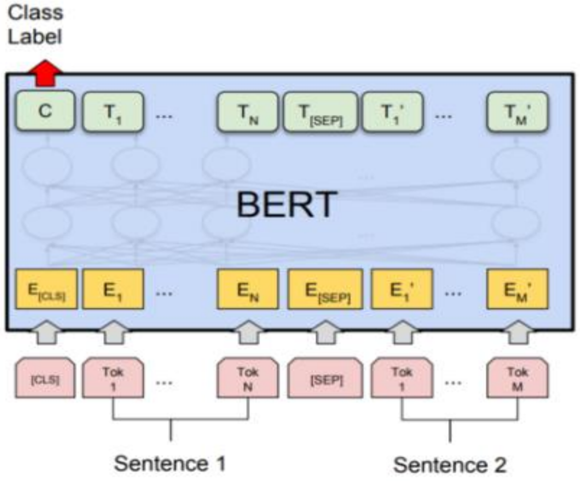


Figure 4. Classification model in BERT

Related Work

Simple BERT Models for Relation Extraction and Semantic Role Labeling

- BERT-based models for relation extraction and semantic role labeling.
- A simple BERT-based model can achieve state-of-the-art performance without using any external features (encode syntactic tree information etc..)
- It was the first time to successfully apply BERT in this manner

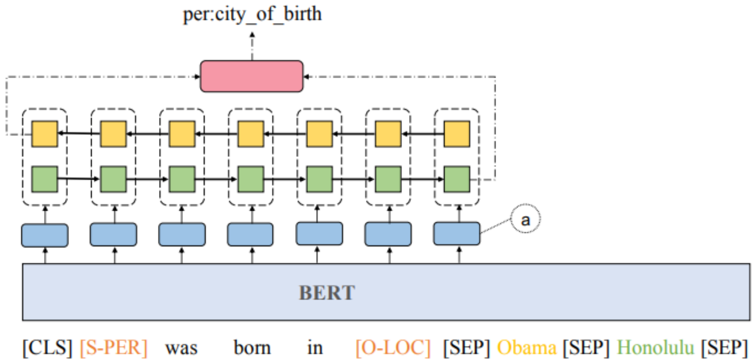


Figure 5. Architecture of relation extraction model

Proposed Model

Data Collection & Preprocess - Chatbot Data

- Chatbot data is synthetic data and there are 11,876 pairs of questions and answers.
- It was produced by referring to the stories frequently found in Daum Cafe's "Better than Love" in some breakup-related questions.
- In original data, daily conversations are labeled 0, break-up conversations are labeled 1, and love conversations are labeled 2.
- We re-labeled the break-up dialogue as -1, the daily dialogue as 0, and the love dialogue as 1.

	Q	A	label
0	12시 땡!	하루가 또 가네요.	0
1	1지망 학교 떨어졌어	위로해 드립니다.	0
2	3박4일 놀러가고 싶다	여행은 언제나 좋죠.	0
3	3박4일 정도 놀러가고 싶다	여행은 언제나 좋죠.	0
4	PPL 심하네	눈살이 찌푸려지죠.	0

	Q	A	label
11818	훔쳐보는 것도 눈치 보임.	티가 나니까 눈치가 보이는 거죠!	1
11819	훔쳐보는 것도 눈치 보임.	훔쳐보는 거 티나나봐요.	1
11820	흑기사 해주는 짝남.	설렜겠어요.	1
11821	힘든 연애 좋은 연애라는게 무슨 차이일까?	잘 헤어질 수 있는 사이 여부인 거 같아요.	1
11822	힘들어서 결혼할까봐	도피성 결혼은 하지 않길 바라요.	1

Figure 7. Example of chatbot data

Proposed Model

Data Collection & Preprocess - Chatbot Data

- In the case of the BERT classification model, each sentence is recognized by adding [CLS] to the front, and the end of the sentence is recognized by adding [SEP].
- By recognizing [CLS], it is possible to know that it is the beginning of a sentence, and by recognizing [SEP], it is possible to know the end of the sentence.
- Since we will train the entire conversation at once, not just a single sentence, we put [CLS] in front of the conversation and [SEP] not only in the end of the sentence but also speaker changes.

```
['[CLS] 12시 땡! [SEP] 하루가 또 가네요. [SEP]',
 '[CLS] 1지망 학교 떨어졌어 [SEP] 위로해 드립니다. [SEP]',
 '[CLS] 3박4일 놀러가고 싶다 [SEP] 여행은 언제나 좋죠. [SEP]',
 '[CLS] 3박4일 정도 놀러가고 싶다 [SEP] 여행은 언제나 좋죠. [SEP]',
 '[CLS] PPL 심하네 [SEP] 눈살이 찌푸려지죠. [SEP]',
```

Figure 8. Example of chatbot data after preprocessing

Proposed Model

Data Collection & Preprocess - Everyone's corpus:Messenger

- Everyone's corpus is data released by the National Institute of Korean Language.
- There are 47,421 training examples.
- This data is not labeled, so we labeled 4,000 dialogues personally.


```
"document": [
  {
    "id": "MDRW2100000010.1",
    "metadata": {
      "title": "온라인 대화",
      "author": "개인 대화 참여자",
      "publisher": "카카오톡",
      "date": "20210513",
      "topic": "미용과 건강",
```

```
"speaker": [
  {
    "id": "1",
    "age": "30대",
    "occupation": "가정 주부",
    "sex": "여성",
    "birthplace": "강원",
    "pricipal_residence": "충북",
    "current_residence": "경기",
    "device": "스마트폰",
    "keyboard": "나랏글"
  },
```

```
{
  "id": "2",
  "age": "30대",
  "occupation": "기타",
  "sex": "남성",
  "birthplace": "제주",
  "pricipal_residence": "경기",
  "current_residence": "경기",
```


Figure 9. Example of Everyone's corpus:Messenger data

Proposed Model

Data Collection & Preprocess - Everyone’s corpus:Messenger

- In addition to the each speaker’s utterance, this data includes various information such as the speaker's gender and relationship. Based on this information, we could label some data in rule-based.
- Like the previous chatbot data, [CLS] and [SEP] are added equally.
- [SEP] token also added when the speaker changes.

	data	label	relation	speaker1_gender	speaker2_gender
0	[CLS] 전 이제 집 도착해서 저녁먹었는데 저녁 드셨나요?ㅎㅎ [SEP] 네 게 ...	0	낯선 사람	여성	여성
1	[CLS] 안녕 [SEP] 나 옷 삼 [SEP] 뭔 옷 [SEP] 에이블리에서 후드...	0	가족>형제자매	여성	여성
2	[CLS] [SEP] 저는 홍차! [SEP] 아아vs뜨아? [SEP] 아아요! 차가...	none	낯선 사람	여성	남성
3	[CLS] name2 안녕 [SEP] 안녕 name1 [SEP] 날씨 좋다\r\n일...	none	낯선 사람	여성	남성
4	[CLS] [SEP] 요즘 날씨가 많이 선선해졌죠?!?! 아침엔 꽤나 쌀쌀하네요 오...	0	낯선 사람	여성	여성

Figure 10. Example of Everyone’s corpus:Messenger data after preprocessing

Proposed Model

Model Architecture - Word Embeddings

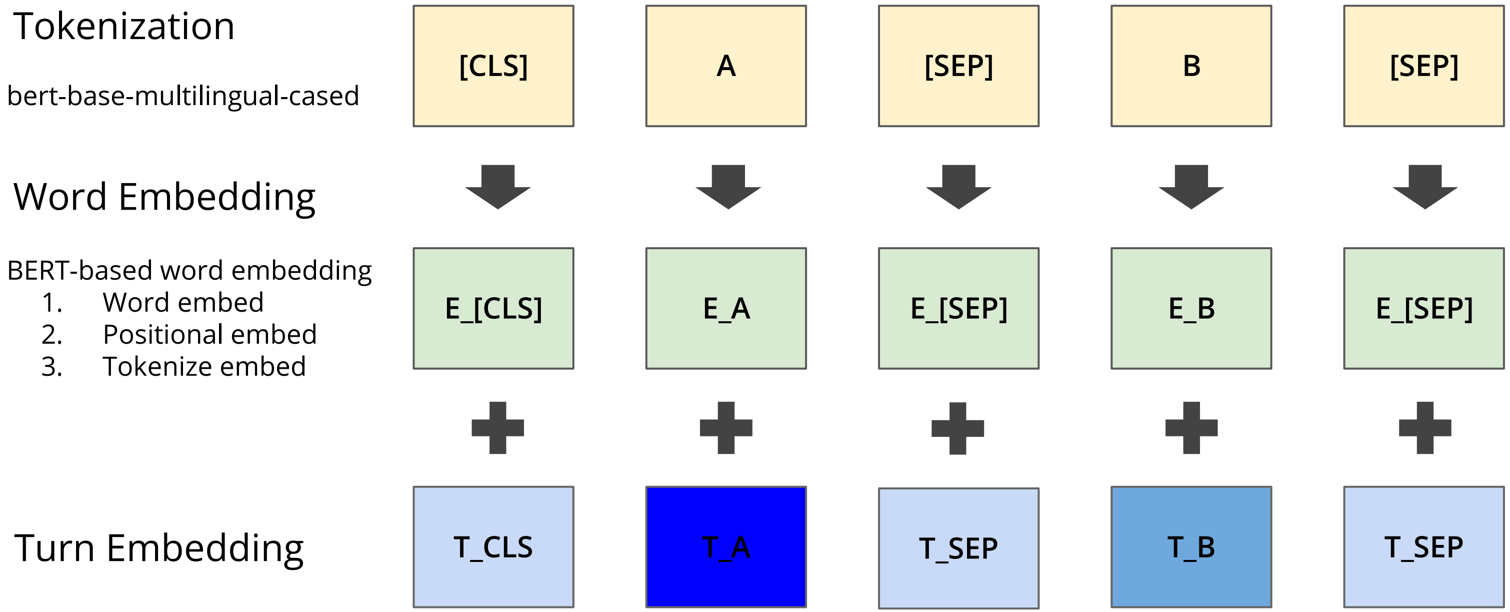
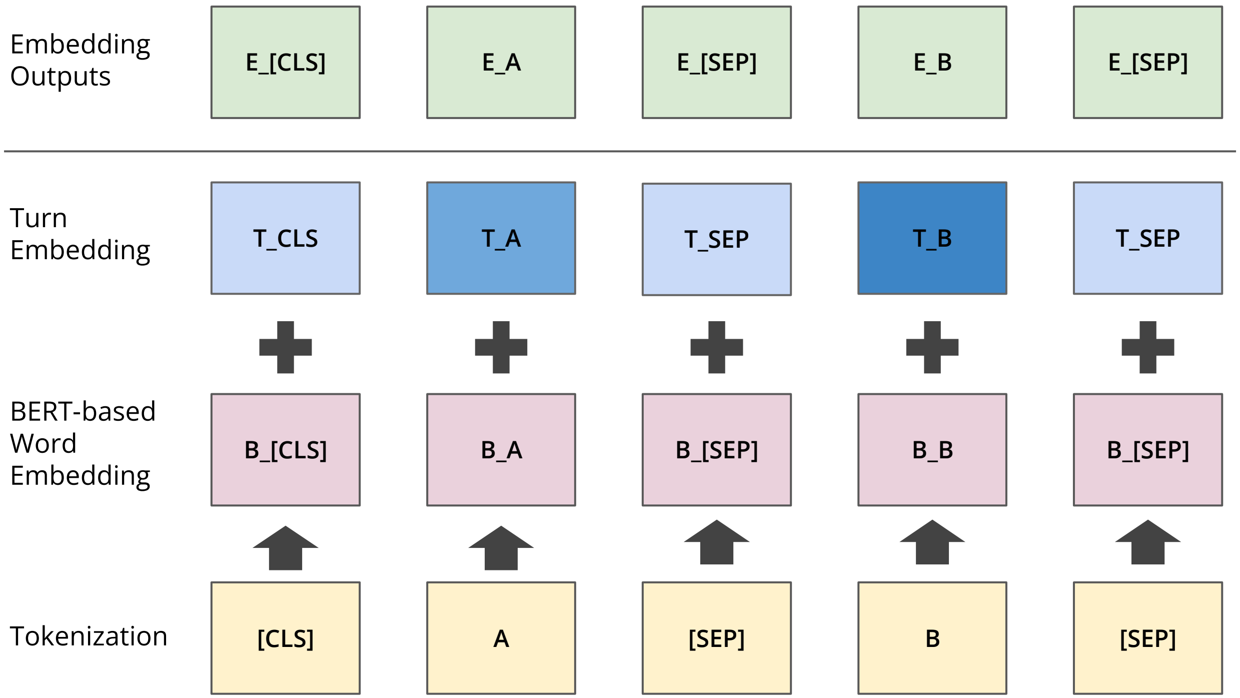


Figure 11. Process of Word Embeddings

Proposed Model

Model Architecture - Word Embeddings



Proposed Model

Model Architecture - Main Model

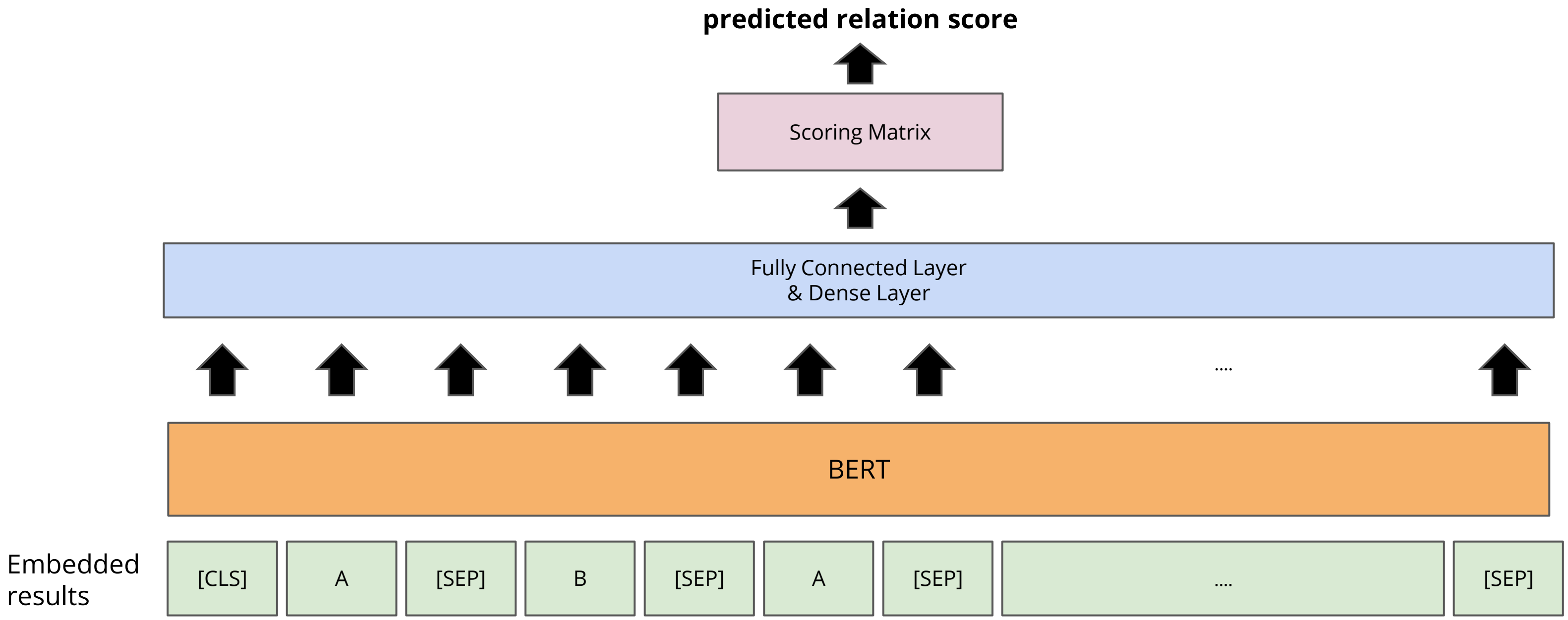


Figure 12. Semantic Score Prediction model(SSP)

Experiments

Evaluation Metrics

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

- **Accuracy** = TP + TN / TP + TN + FP + FN
- **Precision** = TP / TP + FP
- **Recall** = TP / TP + FN
- **F1 Score** on*Recall / (Precision + Recall)
- **RMSE** = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$

Experiments

Experiment Results

	Accuracy	Precision	Recall	F1_Score	RMSE
Baseline	79.38%	57.56%	59.61%%	58.26%	0.45
SSP	83.51%	57.52%	67.86%	58.99%	0.41

Limitation & Future Work

- **Quality of Data**
  - Because of the ambiguity in natural language, labeling is also not reliable.
  - Because of the limitation from tokenizing dialogue, word embedding is not efficient.
- **Data Imbalance**
  - Most of dialog from “Everyone’s corpus” are labeled as 0 (means, daily conversation).
  - This imbalance might interrupt the proper training of the model.
- **No Prior Research**
  - In our knowledge, It is the first time to try to predict the semantics in human-level using dialog so that there is no evaluation metric for this manner and hard to compare with other models.
- **Future Works**
  - We could try to use multimodal model so that the model not only learns the language but also emotion and accent. It will predict much better.
  - We can also try to use pre-trained knowledge based model for this manner.

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. (arXiv preprint arXiv:1810.04805, 2018.)

[2] Peng Shi and Jimmy Lin. 2019. Simple BERT models for relation extraction and semantic role labeling. (arXiv, 2019)

[3] Wang, Jiancheng, et al. "Sentiment classification in customer service dialogue with topic-aware multi-task learning." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. No. 05. 2020. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6897213

Thank you!