Pre-diagnostic Consultant: Bert that mental issue

Yuchen Dai, Angyu Lyu, Yanjie Su, Mengdi Yan, Bolin Zhu
@NOT NULL

USC Viterbi School of Engineering
Ming Hsieh Department of Electrical Engineering

Introduction

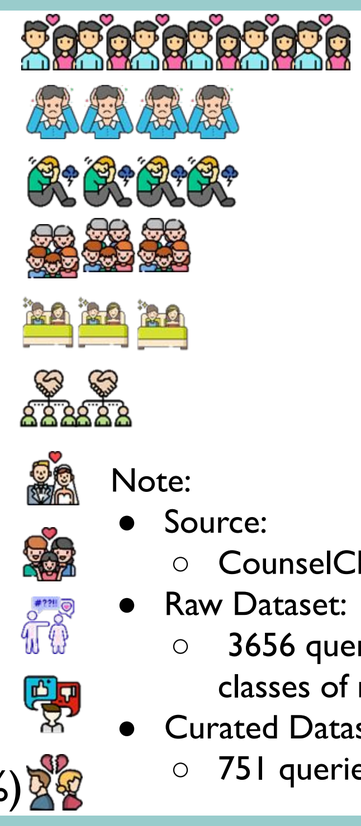
- Seeking for a mental health consultant is in a huge demand
- Identifying the mental issue beforehand will save the time and money

OUR TASK:
Mental Issues Multi-label Classification

- Given a description provided by consultees, identify the categories from the most common mental issues.

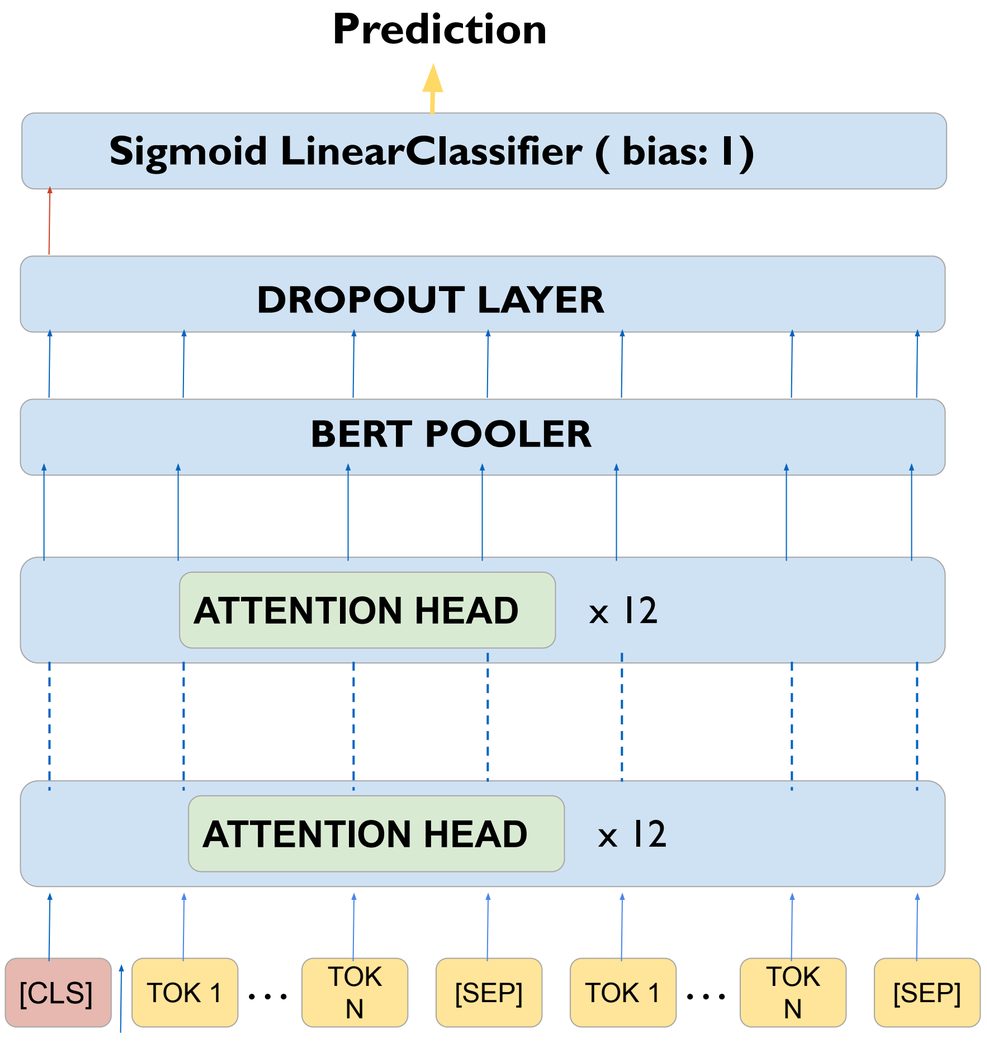
Dataset

Relationships(21.05%)
Anxiety(15.34%)
Depression(13.46%)
Family Conflict(10.05%)
Intimacy(9.7%)
Social Relationships(6.26%)
Marriage(5.63%)
Parenting(5.59%)
Human Sexuality(4.66%)
Behavioral Change(4.42%)
Relationship Dissolution(3.83%)



Note:
- Source:
  - CounselChat website
- Raw Dataset:
  - 3656 queries in 33 different classes of mental problems
- Curated Dataset:
  - 751 queries on 11 classes

Methodology



Output

Classifier (11, 768)

Pooler (768, 768)

Encoder 11 (768 , 768)

Encoder 0 (768, 768)

Baseline

Model Selection
- Naive Bayes
- SVM (OneVSRestClassfie)

Feature Embedding
- Bag-of-Words
- TF-IDF

Fine-tuned BERT for Classification

Pre-trained Model: BERT Base
- Masked Language Modeling (MLM) & Next Sentence Prediction (NSP) with BookCorpus and English Wikipedia.

Input
- Concatenation of title and consultation description

Output
- Binary One Dimensional of size 11

Experiments

Evaluation Metric: Micro-F1

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}$$

- Labels with larger quantities contribute more in this method, which f**its our imbalanced dataset.**

Method	Micro-F1
NB (BoW)	**30.0%**
NB (TF-IDF)	**23.3%**
SVM (Bow)	**26.0%**
SVM (TF-IDF)	**34.6%**
BERT-Base (uncased)	**66.16%**
BERT-Base(cased)	**65.99%**

Conclusion

Achievement
- Built a real-life problem-solver for more efficient online counseling
- Multi-label classification modeling provides a comprehensive understanding of user input
- Fine-tuned the State-Of-The-Art model, BERT

Limitation
- Dataset:
  - Limited
  - Unbalanced

Possible Improvement
- Data augmentation for training set diversity and increase of size
- Try more advanced bert models