



UNIVERSITY OF
ARKANSAS

Movie Reviews Sentiment Analysis Using BERT

Gibson Nkhata, Usman Anjum, Justin Zhan



UNIVERSITY OF
ARKANSAS



*The Fifteenth International Conference on
Information, Process, and Knowledge Management
eKNOW 2023*

*Presenter: Gibson Nkhata
University of Arkansas, USA
gnkhata@uark.edu*

Presenter Biography

- **Gibson Nkhata** received a Master's degree in Computer Science from the University of Arkansas, USA, in 2022. He is currently a doctoral student in Computer Science at the same university. He is also working as a Graduate Research Assistant in the Computer Science and Computer Engineering (CSCE) department under the DART project.
- His research interest lies in application of Deep Learning techniques in Natural Language Processing (NLP) tasks like Sentiment Analysis, Stance Detection and Polarization Quantification on social media.





Outline

- Motivation and objectives
- Related work
- Our approach
- Experiments and Results
- Conclusion and contributions
- Future Work

Motivation

➤ **Sentiment Analysis**

- Sentiment Analysis (SA) aims to determine the polarity of:
 - emotions like happiness, sorrow, grief, hatred, anger, and affection
 - opinions from text, reviews, and posts, which are available in many media platform
- Sentiment analysis helps in tracking people's viewpoints
- It is an important factor when it comes to social media monitoring, product and brand recognition, customer satisfaction, customer loyalty, advertising and promotion's success, and product acceptance, etc.

Motivation

➤ **Movie Reviews:**

- Movie reviews is an important approach to assess the performance of a particular movie.
- A collection of movie reviews gives a deeper qualitative insight on different aspects of the movie compared with stars or numerical rating.
- A huge number of movie reviews can be crawled from social media platforms and deep learning techniques can be applied to NLP tasks to mine knowledge from the reviews.
- Movie reviews have standard sentiment analysis benchmark datasets, e.g., IMDb dataset, where salient and qualitative works have been published on.

Motivation

➤ **BERT**

- BERT is a popular pre-trained language representation model and has proven to perform well on many NLP tasks like Named Entity Recognition, Question Answering and Text Classification.
- Therefore, BERT can just be finetuned with an additional component on any downstream task.
 - This speeds up training
- **Results of sentiment classification can be applied in computation of a predominant sentiment polarity.** This can be used in recommending a movie.

Objectives

- The work is divided into two main goals:
 - How to effectively fine-tune BERT to improve accuracy measure on Movie Reviews Sentiment Analysis.
 - How to compute an overall sentiment polarity of a collection of movie reviews sentiments predicted by the model.
- Contributions:
 - Fine-tuning BERT by coupling with Bidirectional LSTM (BiLSTM) for polarity classification on well known benchmark datasets and achieve accuracy that beats SOTA models.
 - Computing overall polarity of predicted reviews from BERT+LSTM-SA output vector.
 - Comparing our experimental outcomes with results obtained from other studies, including SOTA models, on benchmark datasets.



Our approach

- Sentiment Analysis conducted as polarity or binary classification.
- BERT is fine-tuned with BiLSTM.
- Computing overall or predominant sentiment polarity from the model output vector.

Fine-tuning BERT with BiLSTM

- BERT is not trained from scratch
- We add two input layers to BERT
 - Attention masks
 - Input ids
- The input embeddings (input ids, attention masks) are propagated through BERT afterwards.
- BiLSTM receives information from BERT and feeds it into its dense layer, which then predicts respective classes for the input features
 - BiLSTM has a fully connected layer, which predicts the sentiment polarity.
- During fine-tuning, first layers of BERT are frozen

Model overview

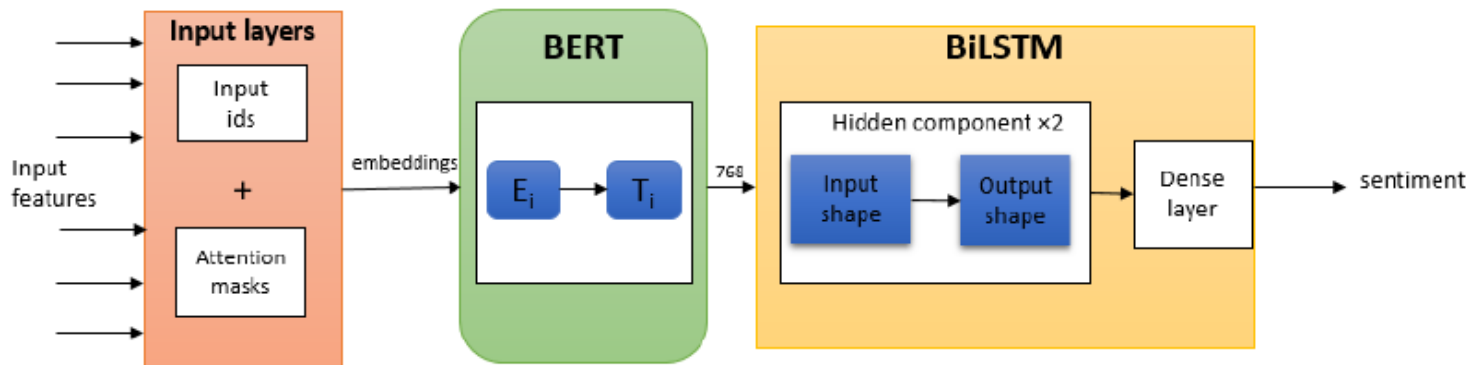


Fig. 2. Fine-tuning of the model



Fig. 3. Overview of our work

Classification and Overall polarity

➤ **Binary classification**

- Given a movie review R , classify it as carrying either a positive sentiment or a negative sentiment.

➤ **Overall polarity**

- Computed on the output of sentiment classification from BERT+BiLSTM
- The output vector is given to a heuristic algorithm which computes the predominant sentiment polarity.

Experiments

Datasets

Name	Size
IMDb Movie Reviews	50K
MR Movie Reviews	10.3K
SST	11.8K
Amazon	345.8K

Experimental settings

Batch size	Seq length	Learning rate	epochs	optimizer	Loss function
32	128	1e-4	5	Adam	Binary cross entropy loss



Results

TABLE II
ACCURACY (%) COMPARISONS OF MODELS ON BENCHMARK DATASETS
FOR BINARY CLASSIFICATION

Model name	Dataset			
	<i>IMDb-2</i>	<i>MR</i>	<i>SST-2</i>	<i>amazon-2</i>
RNN-Capsule [29]	84.12	83.80	82.77	82.68
coRNN [30]	87.4	87.11	88.97	89.32
TL-CNN [30]	87.70	81.5	87.70	88.12
Modified LMU [28]	93.20	93.15	93.10	93.67
DualCL [27]	-	94.31	94.91	94.98
L Mixed [31]	95.68	95.72	-	95.81
EFL [25]	96.10	96.90	96.90	96.91
NB-weighted-BON [14] +dv-cosine	97.40	-	96.55	97.55
SMART-RoBERTa [26] Large	96.34	97.5	96.61	-
Ours	97.67	97.88	97.62	98.76

TABLE III
RESULTS OF ABLATION STUDY

Model name	Dataset			
	<i>IMDb-2</i>	<i>MR</i>	<i>SST-2</i>	<i>amazon-2</i>
BiLSTM	90.42	90.5	91.12	92.18
BERT	93.81	94.29	93.55.97	94.78
BERT+BiLSTM-SA	97.67	97.88	97.62	98.76

TABLE IV
OVERALL POLARITY COMPUTATION ON ALL DATASETS

Dataset	Original overall polarity	Computed overall polarity
IMDb	Neutral	Neutral
MR reviews	Neutral	Neutral
SST-2	Neutral	Neutral
Amazon	Positive	Positive

- Our model outperforms all other models
- On ablation studies, BERT and BiLSTM separately give nonoptimal accuracy compared with BERT+BiLSTM.
- Computed overall polarity on BERT+BiLSTM predictions is always the same as original overall polarity before predictions.



Conclusion

- In this work, we extended the existing domain knowledge of sentiment analysis by showing
 - How to effectively fine-tune BERT to improve accuracy measure on Movie Reviews Sentiment Analysis.
 - How to compute an overall polarity of a collection of movies reviews sentiments predicted by a model.
- We coupled BERT with BiLSTM and applied on binary classification.
- Our model outperformed other models
- The computed overall polarity from predictions is the same as the computed overall polarity from original labels
- Ablation studies also show that BERT and BiLSTM separately provide non-optimal accuracy compared against BERT+BiLSTM-SA, implying coupling of the two tools is stronger for the model generalization.
- To the best of our knowledge, this is the first work to couple BERT with BiLSTM for Sentiment Analysis.



Future work

- Extend the model to three-class, four-class, or any fine-grained classification.
- How to effectively apply accuracy improvement techniques to transformed BERT features despite loss of semantic information in them
- Exploring other pre-trained language models for Sentiment Analysis.
- Exploring how different components of a sentence contribute to its sentiment prediction.