Transformer-based sentiment analysis and summarization on movie reviews

NICOLA TOSCAN   ANTONIOS MANTZARIS

`toscan|mantza @kth.se`

January 8, 2022

Contents

1   Project Description

Movies scores are not by themselves representatives of a movie. The majority of cinephiles are reading other people's reviews in order to extract information about a movie before they watch it and judge by themselves if that movie suits to their preferences or not. For that purpose we are going to develop a tool in which users can get multiple summaries of reviews expressing different points of views from other people (e.g.summary of all positive reviews on movie "Pulp Fiction").

Following this direction, we developed a classifier model which gets as input movie reviews and predicts the sentiment profile of the reviews which is either positive or negative. Afterwards, the user can select the movie they want to get the summary for. We then concatenate the reviews according to their sentiment analysis in order to provide clear, homogeneous summaries. Therefore, the final shown summaries could be up to two, one for positive, according to the previously conducted sentiment analysis, reviews and one for the negative ones.

2   Dataset

We used the open IMDB dataset [1] provided by [2] to train our final models and perform the tasks mentioned in 1. We made a split for training and testing the sentiment analysis models. Part of the dataset consists of unlabelled movies. These movies are used as request from the users for classification and eventually summarization. In our work we are used both regression and classification models, where the first one refers to the ranking of a movie while the second one to the positiveness or negativeness of it. Hence, we created two different versions of that split, one for each purpose. The only difference between these splits is on the target column, which is "rating" and "label" for regression and classification respectively.

The columns that form the used dataset are the following:

- `review:` the review of the movie. It is of String type and its size exceeds on average the 200 words.

- `rating:` the rating of the movie which ranges from 0 to 10. The values are integers and 5, 6 where missing. For the regression model rating were converted to float ranging from 0 to 1.

- `label:`positive or negative review. Represented by 0 and 1 respectively.

- `title:`title of the movie. Makes it easier for the user to browse the dataset.

- `url:` url of the movie in the IMDB website [3]. Helps the user visit the searched movie by themselves and validate the summary or make their own comment.

3   Methods

As mentioned in 1, the aim of this project is to facilitate cinephiles with their next movie research by providing to them a summary of reviews made for this movie grouped by their sentiment analysis. Eventually, the user faces a small of both positive and negative reviews separately which will help them distinguish the main positives and negatives of a movie in order to make the decision to watch it or not.

The project can be divided into two parts, sentiment analysis of the text and text summarization. Both were tackled from the transformers' point of view. Transformers are able to parallelize text input and perform contextual analysis with the help of their attention mechanisms [4]. They have proved that they outperform other models on several downstream tasks such as machine translation, text classification and question answering, so they were selected from exploration of their capabilities and usability.

3.1   Sentiment Analysis

We split the dataset into training and test datasets with a 70-30% ratio. We use the training dataset to train our model, and the test dataset for evaluating the produced models. The predictions and eventually summaries of the grouped reviews by their prediction is made on the unlabelled movies of the dataset. We explore several techniques, including classification, regression and transfer learning. For the regression parts the split was similar but the target column changed from "label" to "ratings" normalized by 10 in order to get values in the range of [0,1].

In total we have built 4 models. The first one performs a binary classification. Specifically, we are using the pre-trained small Bert [5]. We take the produced embeddings from it and then add another output layer on top which makes the binary classifications. This way we are taking advantage of the pre-trained Bert model and are fine-tuning it for our purpose which is now the binary classification. The second model is built on top of that. We make use of transfer learning and fine-tune a regressor on top of the first model. Both tasks are similar since the regressor is on the ratings normalised by 10 to bring it closer to the probabilistic classification task of model 1. The remaining 2 models are following a similar approach but the other way around. We first fine-tune a regressor on top of the pre-trained Bert which makes predictions for the normalised ratings and then we fine-tune a classification model on top of that for the binary classification task. Therefore, all four models are transformer based since they are all based on a pre-trained Bert variant and then the two out of four are applying additional transfer learning to produce similar to their base model predictions.

For the implementation of the aforementioned TensorFlow's Keras [6] was used. Diving a little more into detail, when pre-trained Bert is used, we have "trainable = True" so we unlock the weights of the pre-trained model. Moreover, we add a dropout layer to prevent the model from overfitting. As optimizer we use Adam which is a combination of RMSProp and Momentum and performs better than others in most cases. Concerning the loss and metrics, they are adjusted to the at the time task, classification or regression, and are specifically BinnaryCrossEntropy, BinaryAccuracy and MeanSquaredError respectively.

3.2   Review summarization

Text summarization is divided into two type, extractive and abstractive [7]. Extractive summarization extracts information that is identical to the original text. It rates each sentence against the others and selects the most informative. On the other hand, abstractive summarization produces a summary based on the full content focusing on the most significant points of the original text. Which type is better depends on the application and dataset. Most of the time abstractive summarizations perform better but also they are more difficult to train. In this project we are exploring both types with BertSum, GPT2 and XLNet as representatives of the extractive type and variants of T5 as representatives of the abstractive.

The models used in this section are pre-trained in a large corpus specifically for the summarization task. Developing a new model from scratch based on Bert,like we did for 3.1, would require way more time that we did not have. Moreover, additional training would be required only if we wanted to summarise a specific domain's text, containing a lot of terminologies that the pre-training dataset might had missing. This is not happening in our case so further fine-tuning would be redundant. Since further training was not performed, this tasks focuses on highlighting the capabilities of transformers in the task of text summarization and on the comparison of several models developed and are following one of the two mentioned summarization types.

4   Results

For all four models used to classify the reviews we got similar results on the test set accuracy:

- Classification: 86.5%

- Regression: 86.7%

- Classification fine-tuned on Regression: 86.4%

- Regression fine-tuned on Classification: 86.7%

We chose to use the label predicted by the Regression fine-tuned on Classification GPT model to use for the summarization, since it is the one with the highest accuracy and smallest loss, even if by a small margin.
As far as the summarization is concerned, some representative results can be seen in the following figures and in A. It is noticed that abstractive types are producing smaller summaries since they are not using whole sentences from the input text. Moreover, the T5ForConditionalGeneration has a limit on repeated n-grams 6 instead of a penalty and that facilitates the avoidance of repetitions as seen in 5

Figure 1: Summary of positive and negative reviews with BertSum

Figure 2: Summary of positive and negative reviews with GPT2

```
- POSITIVE:
Filmed in beautiful locations the director and scriptwriter should be awarded along with brilliant acting from Pierce, Woody an
d Salma. It makes a film well worth watching time and again - which I do. I read the comments about this movie before I saw it
and was afraid I might not like it. One comment said they were disappointed by the ending of the movie. It had just the right a
mount of action, suspense, and even some comedic moments. It is a perfect movie to watch if you just want to relax and be enter
tained. Information on Tanbo is easily found on the internet, and can be played at tanbo.com. Not being a big fan of Pierce Bro
snan, I was a bit apprehensive to even watching this movie. Perhaps one of the greatest attractions is the locations; filmed in
the beautiful Bahamas. Max Burdett (Pierce Bronsan) is one of the best jewel thief in the world, along with his partner/girlfri
end Lola (Salma Hayek). Only the disappointing performance is Cheadle, which his supporting role is very thin for this talented
actor. DVD has an interesting often funny commentary track by the filmmakers, deleted scenes with optional commentary by the fi
lmmakers, outtakes and more. It is a slightly suspenseful drama, with a few funny one-liners to make you laugh. Selma Hayek put
s on all her sexual charms for the role of Lola and would have probably done really well if she had actually had a really well
written role. But his role in the film is so small that he is unable to really make up for the bad jobs that some of the other
actors did. Don Cheadle doesn't have much to do, as his character looks as though he was an after thought.<br /><br />Just pret
end you are on vacation in the Bahamas. The action was more subdued and less far-fetched as a typical Brosnan movie. <br /><br
/>After the heist Harrelson follows the pair to an island where he believes they plan another big job.
```

```
- NEGATIVE:
A repeat of the Thomas Crown affair, which was only good because of the originality of the 1960's script. Like Mr Paul Zbyszews
ki's other production effort "The Weakest Link" what a talent! After the sunsets a great comedy and I think Pierce Brosnans gre
at. In After the sunset there is too much sex and nudity for the rating I mean eight year olds could go see that and Pierce Bro
snan has his shirt off for like half the film. The language is great though, not too much swearing but where there is swearing
its funny. But Agent Lloyd is equally out of his depth in a film that doesn't have any. The film also grotesquely fails as a st
udy of relationships and what we find attractive. Sometimes I had the impression that AFTER THE SUNSET should be a wannabe Jame
s Bond movie but it failed completely. Therefore the picture was very fine and clear but all that is not enough to get an avera
ge rating. jokes that play like the first draft of a "Will and Grace" episode. Don Cheadle's talents are wasted in a thankless
role as the villain--just when you've forgotten he's even IN the movie, he turns up again to bring the film to a halt. Into thi
s mix is a local gangster who wants Brosnan to steal the diamond for them.<br /><br />Overall, it was an OK movie. Fine, but t
hen give us something exciting, a story we can get involved with.<br /><br />Brosnan plays Max much like he played Bond - suave
and light. Harrelson mugs most of the time and Don Cheadle, the best of the bunch, is wasted in a role that had tremendous poss
ibilities. Clad in tight dresses, bikinis and showing off plunging necklines, she struts about the screen oozing charm and sexu
ality. At the beginning he steals a diamond, but the heist scene isn't very good at all, and neither is this movie. So the movi
e starts where most heist movies end, in paradise. As in all bad heist movies, the actual heist is all of 10 minutes. If so, wh
y wasn't there a calendar made for those who didn't make the final cut? Finally, am I to believe that Pierce could still move a
s quickly as Ratner defined? There was no excitement or intensity, just plain ⊠ "here it is ⊠ like it or not". I personally cou
ldn't digest this film with its poor acting and below-par story.
```

Figure 3: Summary of positive and negative reviews with XLNet

```
- POSITIVE:
<pad> 'after the sunset' is a new movie by director Brett Ratner. it stars Pierce Brosnan and Salma Hayek as lovers and retired
thieves living at the Bahamas. the movie doesn't focus much on the heist and it focuses on the great chemistry of its character
s.

- NEGATIVE:
<pad> a repeat of the Thomas Crown affair was only good because of the originality of the 1960's script. the acting is non-exis
tent as the performers seem only to be massaging their ego's. the story of a jewel thief who may or may not have retired has be
en done before.
```

Figure 4: Summary of positive and negative reviews with T5 base

```
- POSITIVE:
Watch After the Sunset.Watch The Seven Seas Navigator.Watch the movie's chemistry.Watch the movie's chemistry.Watch the movie's
chemistry.Watch the movie's chemistry.Watch the movie's chemistry.Watch the movie's chemistry.Watch the movie's chemistry.Watch
the movie's chemistry.

- NEGATIVE:
Know that this film is a good movie.Learn about the plot.Learn about the characters.Learn about the story.Learn about the role.
Learn about the character.
```

Figure 5: Summary of positive and negative reviews with T5 wikihow-t5-small

```
- POSITIVE:
"After the Sunset" is a very entertaining and entertaining film. it's a comedy about heist with nice elements of romance and th
riller. the film is based on a series of cat and mouse characters. if you have a jealous girlfriend, you will enjoy this movie.

- NEGATIVE:
"After the Sunset" is about Pierce Brosnan, a retired jewel thief. it's about a heist man who robs a diamond from a cruise shi
p. the film has a strong air of sexuality.
```

Figure 6: Summary of positive and negative reviews with T5ForConditionalGeneration

5   How to run the code

Everything is incorporated into a Google Colab notebook. Therefore, in order to run the code and reproduce the results or experiment with this or another dataset, the procedure is pretty straightforward and self-explanatory. The several pretrained models, the datasets and the jupyter notebook have been saved and are ready for use without training and can be found in the following shared Google Drive folder:
https://drive.google.com/drive/folders/1p6fW1tKAfFyyrHcIlJL1ONlhRRnmDFr$_u$sp = sharing
The shared folder "imdb" contains a folder called "toscan", inside that all the four pretrained model are present, while the dataset are contained in the "dataset" folder. The easiest way to emulate our results is to add shortcut to the shared folder "imdb" to the root of your own Google Drive account, and mount it the Google Colab environment as shown in the notebook. Following this instruction all the folders should already be in the right place, and all the code should run without errors.

References

[1] A. Maas, "Large movie review dataset." [Online]. Available: https://ai.stanford.edu/~amaas/data/sentiment/

[2] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, Oregon, USA: Association for Computational Linguistics, June 2011, pp. 142–150. [Online]. Available: http://www.aclweb.org/anthology/P11-1015

[3] "Ratings, reviews, and where to watch the best movies tv shows." [Online]. Available: https://www.imdb.com/

[4] S. Ravichandiran, *Getting Started with Google BERT Build and Train State-Of-the-art Natural Language Processing Models Using BERT*. Packt Publishing, 2021.

[5] I. Turc, M.-W. Chang, K. Lee, and K. Toutanova, "Well-read students learn better: On the importance of pre-training compact models," *arXiv preprint arXiv:1908.08962v2*, 2019.

[6] K. Team, "Simple. flexible. powerful." [Online]. Available: https://keras.io/

[7] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Systems with Applications*, vol. 165, p. 113679, 2021. doi: 10.1016/j.eswa.2020.113679

A   Appendix

Figure 7: Summary of positive and negative reviews with BertSum

Figure 8: Summary of positive and negative reviews with GPT2

```
- POSITIVE:
Even if the stars do a great job, the action is definitely the biggest star in this film, as in the others in the series, and t
herefore the story has it faults. Like how ** SPOILER WARNING** Anck-su-namun would leave her beloved because she is afraid of
a little raining stone!
- NEGATIVE:
This film is similar to Jurassic Park II, in that the computer graphics make the movie. I loved the first movie so I thought th
e second one would be even better. The first Mummy picture was quite good, in that it managed to fit in characters and plot amo
ngst the big budget special effects. Heck, you could be paying attention the whole movie and still be confused. There is some e
ntertainment to be found, but you may find after you finish this, you'll be in desperate need for something intelligent.<br /><
br />My rating: * 1/2 out of ****. Most of the characters do a lot of traveling and then have a showdown. Mixed in is some chum
ming of the good guys that leads no where. This adds to the idea that it was just a poorly written movie to cash in on the firs
t movie's success. If you can snag it cheap, you'll feel like you got your money's worth. It isn't one of those movies where yo
u just feel cheated for your time. Historical innaccuracies, zero characterization, cliche's galore...any one of those I could
have forgiven, but not all of them.<br /><br />This is a big, soulless, hollywood, summer movie. Why didn't these neato abiliti
es kick in in the first movie??? Sorry, suspension of disbelief only goes so far...jets...we didn't even HAVE JETS THEN!!!
```

Figure 9: Summary of positive and negative reviews with XLNet

```
- POSITIVE:
<pad> the action is definitely the biggest star in this film, as in the others in the series. the film could be really good for
what it is trying, cut the brakes and full gas. but at least they should have gotten the special effects right so the scorpion
king didn<unk>t look like a plastic doll.

- NEGATIVE:
<pad> "The Mummy Returns" is a campy film that doesn't take itself too seriously. the film is similar to Jurassic Park II in th
at the computer graphics make the movie. the worst part of the movie is all the animations. the movie is not better.
```

Figure 10: Summary of positive and negative reviews with T5 base

```
- POSITIVE:
Be aware that the action is a big part of the movie, and it is not just a film.Be aware that there are many things that can be
forgiven in this movie.Ask yourself why you want to go for a rollercoaster ride.
- NEGATIVE:
Know that this movie is a good movie, but it's not the first one.Know that this movie isn't an action movie.Learn about the cha
racters in the movie.Learn about the plot.Learn about the film.
```

Figure 11: Summary of positive and negative reviews with T5 wikihow-t5-small

```
- POSITIVE:
the action is definitely the biggest star in this film, as in the others in the series. but some things just plain bug me in th
is kind of movie. if you go for a rollercoaster ride, do you expect subtlety? I hope not, and with that in mind, a lot can be f
orgiven.

- NEGATIVE:
if you're an action movie lover like me, you'll be tempted to try this movie. the movie is a big, soulless, hollywood, summer m
ovie, but it doesn't mesh with the first one. it's not a movie that's a lot of fun, but a good movie, too.
```

Figure 12: Summary of positive and negative reviews with T5ForConditionalGeneration