Multi-review opinion summarization via jointly modelling semantic and helpfulness

ENGG 5501 Project Report

Xin Li
SEEM
The Chinese University of Hong Kong
1155083286

Dihan Chen
MAE
The Chinese University of Hong Kong
1155085918

Jialong Chen
MAE
The Chinese University of Hong Kong
1155100371

Jiangtao Liu
EE
The Chinese University of Hong Kong
1155116798

ABSTRACT

Nowadays, reviews and ratings are playing a more important role in the Internet social media. In this project, we plan to construct a system to analyse Chinese movie reviews as to realize multi-review opinion summarization via jointly modelling semantic and helpfulness. The whole project includes five parts: data scraping, word segmentation, semantic analysis, multi-purpose aggregation and visual presentation. Among them, more attention will be paid to the semantic analysis and multi-purpose aggregation. Several concepts, such as linear regression, machine learning and unsupervised learning, are combined to help realize several objectives of the opinion summarization, including the cleaning of useless comments, the refining of the key characteristics, the evaluation of review quality and other objectives. Further, since there are many scenes full of reviews and ratings upon different objects, such as foods, books, attractions and Weibo hot topics, our system could be easily transformed to analyse them, which might provide a new perspective on things and could perform much more efficiently upon data of large volume or high velocity.

KEYWORDS

Review semantic, Chinese Natural language processing, Movie rate prediction, Multi-review opinion summarization

1 Introduction

1.1 Motivation

Since the Internet has achieved exponential increasing in these decades, billions of people all over the world benefit from the diversity and immediacy of information brought by the Internet, which serves as an infinite platform for sharing opinions about everything, such as book reviews, video reviews, shop reviews, social hot comments, etc. Meanwhile, people are also getting accustomed to enjoying these reviews and ratings before or after appreciating the contents to help them preview or search for someone with similar interests. Compared with ratings, text reviews, especially long text comments, are usually more reliable and convinced due to the responsibility and interest of the authors, which might provide plenty of valuable information. However, due to the limited reading speed with respect to the massive data in the Internet, it becomes much harder to obtain useful information efficiently without any auxiliary tools. At the same time, a large amount of junk comments induced by commercial interests also influence the quality of the ratings and reviews. To solve the problem of the summarization of these text-based reviews, we plan to build such an online review analysis system via advanced big data algorithms to improve the process, starting from Chinese movie reviews. Multiple review purposes, such as the cleaning of useless comments, the refining of the key characteristics, the evaluation of review quality and other objectives will be implemented. Further hidden interesting clues are also expected to be dug out via applying the algorithms taught in this course.

1.2 Topics related

To achieve the final goal, many subtopics in this course are highly useful for data processing. Two of the most related course topics should be natural language processing and machine learning. Natural language processing is the preliminary step for the abstraction, and machine learning is employed to aggregate and summarize the conclusions from raw big data. Besides, locality sensitive hashing and dimensionality reduction might also have a chance to be applied to simplify data processing or enhance the final performance.

1.3 Related work & differences

Our project is to capture the semantic features and the helpfulness of the review and determine the movie opinions according to both of the review representations and the review helpfulness. First of all, this project is related to regression-oriented representation learning. [1-4]employ the convolutional neural networks to learn the sentence/document-level representations.[5] explicitly models the contributions of the sentences within the review via a hierarchical Gated Recurrent Unit (GRU). Different from the above-mentioned works, [6, 7]proposes to combine the temporally recurrent structure and convolutional feature extractor, achieving competitive classification results. Recently, the language model pre-training techniques [8, 9] have been introduced into the classification models and substantially advance the state-of-the-art for the document classification task.

This project is also related to review helpfulness prediction; whose aim is to automatically assess the helpfulness score of the reviews. [10]and [11]are the pioneering works exploiting hand-crafted features for the prediction of review helpfulness. [12] further improves the prediction performance based on argument-based features. [13, 14] conduct some initial attempts to perform the helpfulness prediction with the help of neural network-based models and establish the new state-of-the-art results.

The differences between these problems and the problem investigated in our project are as follows:

1. Our project does not solve these two problems separately but unifies them within a multi-task learning framework for predicting the movie-level rating.
2. The supervision signals required by the helpfulness prediction task are dispensable in our framework. Theoretically, in order to predict movie-level opinions more accurately, the model has the tendency to highlight the useful reviews and down-weight those useless.

2 Fetching data

The dataset comes from the famous movie website: movie.douban.com, which have millions of movie lovers from mainland China. The whole dataset contains 20 different types of movies that are classified by official Douban website. Each type includes around 1000 popular movies, where one movie has at least 2000 reviews. The reviews are divided into three categories following the scored stars when reviewers made these reviews. We assume 4-5 stars is positive attitude, 3 stars is middle attitude and 1-2 stars is negative emotion. Now we have approximately 500 reviews per class of emotions per movie to train our model and another 500 hottest reviews plus 100 newest reviews per movie for testing. We developed our unique crawler to obtain all the reviews from website and build the special dataset ourselves.

The dataset contains the corresponding reviews of various movies and each movie has its own five rows of reviews, including 500 positive reviews, 500 middle reviews, 500 negative reviews, 500 hottest reviews and 100 newest reviews. The total number of movies in our dataset is around 10000.

The example of basic statistics of the dataset is shown in Table 1, which takes The Shawshank Redemption (肖申克的救赎) for example. The dataset contains the corresponding reviews of various movies and each movie has its own five rows of reviews, including 500 positive reviews, 500 middle reviews, 500 negative reviews, 500 hottest reviews and 100 newest reviews. The total number of movies in our dataset is around 10000.

Movie Name	*1*	*2*	...
(e.g., 肖申克的救赎)			
POS (500 reviews)	超级喜欢，不看的话人生不圆满	经典的电影	...
MID (500 reviews)	好看但这么高的分有点假了吧	在我的心中，它一直都是最被高估的电影	...
NEG (500 reviews)	真的不喜欢，不好看，没感觉	大众经典我从不感冒，为什么？我欣赏水平不行?	...
HOT (500 reviews)	不需要女主角的好电影	关于希望最有力的诠释	...
NEW (100 reviews)	超级震撼	看完后劲很大的一部经典，强推!	...

Table 1: Basic statistics of the dataset

Result of crawled reviews data crawled around 15000 movies and finally got 10269 unique movies after processing. All of the movie information and the corresponding comments have been written into one csv file. We upload this file to kaggle.com and it is now publicly available at: https://www.kaggle.com/liujt14/dou-ban-movie-short-comments-10377movies. Each record in the dataset has 12 attributes, including movie name, overall score, stars associated with each comment and so on.

The following figure shows a heading for the whole dataset, and the most important part should be the reviews part.

```
ID,Movie_Name,Score,Review_People,Star_Distribution,Craw_Date,Username,Date,Star,Comment,Comment_Distribution,Like
0,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,尾黑,2018-06-23,3,成本低廉的PPT电影，用Nico生命中最后一年发生的事给Nico的歌配上情节，倒不算尴尬。女猪演得不错，在电影中非常少见的中老年女
1,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,大奇特(Grinch),2018-06-28,3,传记片能做到半真半假的真实，就成功一半了，然后就是走进人物的内心世界。这部影片做到了纪录片般的真实。,66%31%3%
2,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,西楼尘,2018-07-01,3,意大利房东记得他的嚣张，岔开大腿抽着烟说，我的计划是成为优雅的老女人。总念错名字的劳拉记得她的黯然，在病床前问起你有
3,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,张晓禾,2018-06-28,3,意大利电影周开幕片,66%31%3%,0
4,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,汪金卫,2018-06-29,3,【中国电影资料馆展映】歌星Nico的最后三年，青春不再，人气不再，衰老暴躁，在漂泊不定的巡回演唱、吸毒、与儿子修复关系中
5,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,seabisuit,2018-06-29,3,我曾到达巅峰也跌落过谷底发现都是空空如也，当摇滚少女变成毒瘾大妈，唯有那颗躁动的心能维持了，丹麦女星也是这样，表演
6,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,Lycidas,2018-07-16,3,很好奇，她在东欧旷野里因为戒断反应呕吐不止的时候，是不是也会回忆起年轻时那么多光芒四射名动天下的恋人们？还是真的如她
7,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,科克托马斯,2018-06-28,3,6.28 毕业看片，找工作苦涩,66%31%3%,1
8,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,美卡 mecca,2018-07-17,3,7/14@电博意大利影展 7。The struggling single mother musician on her last tour. 罗马尼亚光明节如谶一瞥。捷克快闪。
9,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,脱,2019-07-31,3,片名还是要叫妮可，就足以消解所有反思和深情。,66%31%3%,0
10,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,华盛顿樱桃树,2018-06-28,3,制作粗糙的学生毕业作品也就这样了吧？故事更是随意的东一下西一下。年份叙述毫无节奏感。我认为从头到尾的每一分钟都
11,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,LEのn个晴天,2018-06-28,3,"“我的内心是如此的空虚，但是我的歌曲里，满是对你的爱”
“我到过顶峰，也到过低谷，全都是空空如也。”
威尼斯节精粹展映，意大利影展开幕影片。",66%31%3%,0
12,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,Captain_C,2018-06-21,3,#siff21# 音乐响起时，才是最真实的Christa。地下演出的那一场，癫狂放纵，随性自由，是经历一生坎坷的呐喊，感染力十足。
13,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,LeonBible,2018-07-14,3,2018106 电博。原来是与菲利普加瑞尔合作《内心的伤痕》的Nico传记。整片以一种缓慢甚至于笨拙的姿态讲述Nico逝世前的生
14,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,不是异,2018-07-07,3,The morning is small, the evening is tall.尼伯龙根、空空如也、留在身后。,66%31%3%,1
15,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,大 螨 螨,2018-08-05,3,太用力了。。,66%31%3%,0
16,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,凤瞳,2019-09-04,3,第74届威尼斯电影节地平线单元最佳影片 被仰望的传奇 巅峰过后的自我,66%31%3%,0
17,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,五月和七月,2018-06-28,3,"其实演的很不错，就是还是会出戏，毕竟容颜不是那么相似……
不停闪回的片段倒是让人有些联接，布拉格的那出非法演唱会很赞，还有一直在身边的录音设备。",66%31%3%,0
18,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,本位,2018-07-14,3,我也不知道这算不算是音乐类传记片在编剧上比较容易出现的顺拐，不太喜欢这种把所有人物情绪变化都要糅入最近的一场演唱中的做
19,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,akasa|呆,2018-06-16,3,Young people are boring/But the world belongs to them/Do you really believe it/No/赶紧回去听solo Don't call me Nic
20,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,Jacqueline,2018-07-16,3,我到过顶峰也到过谷底，都空空如也。我一直在找一种声音。,66%31%3%,0
21,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,chushizhang,2018-08-17,3,cinematography多有出彩之处，但无论是角色还是演员的问题或是有意为之，被设定为经历过风浪的Nico行为像一只草原上的
22,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,INFITHELLs,2018-06-21,3,SIFF2018…7/10 看來是真樂迷了，視角沒能嚴守、敘事邏輯或缺都不成問題……橫移鏡頭下的明日派對就夠哭的了。I’l have it
23,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,王哈哈,2018-06-28,3,"其实这个导演是有感觉的，某些片段颇有味道，也确实对人物内心有不少挖掘。不过人物转变有些突兀，感觉是完全没铺垫够就被
随身携带的录音器是个很好的线索，应当有更多着墨和阐释的。
虽不成熟，但有感觉总比没感觉好，好好打磨剧本，熟练导演技巧，应该会有更好的作品。",66%31%3%,0
24,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,James Brown,2017-08-31,3,对Nico的印象之来自地下丝绒和切尔西女孩，不过这电影是Nico的传记吗？明明就是Christa。,66%31%3%,1
25,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,在胳肢窝跳舞,2018-06-29,3,从来没接触过地下丝绒和Nico，女主角的演绎很动人，影片比较普通吧。。。。,66%31%3%,0
26,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,宁二狗,2018-07-14,3,等我从伊维萨回来我就停止做音乐。1988妮可死于伊维萨。,66%31%3%,0
27,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,深邃,2018-07-14,3,刚现场听完John Cale就来看这片，做人呐到了巅峰以后要做的就是怎么摆脱这个巅峰的阴影了吧。,66%31%3%,0
28,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,Philia的青猫,2018-06-21,3,Siff第四场。每次都靠当年纪录片段和安迪沃霍尔的闪回回血。乐队巡演真的是一件很复杂的事情。当年的Nico做回自己，拿
29,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,李草木,2018-06-24,3,和女性小伙伴看完一致认为“贵圈真乱”，不到10人的男男女女排列组合真的会尴尬，唯一的好处是省些房费(歪果仁出差不拼房)。迟
30,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,宋人玫瑰,2018-08-10,3,上海电影节观影之一,66%31%3%,0
31,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,嗝,2018-07-07,3,我想成为一个优雅的老女人,66%31%3%,0
32,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,盛夏瞬噬,2018-06-28,3,反正就是无论经历过什么高潮低谷到头来都只是一片虚无吧……,66%31%3%,0
33,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,小R酱,2018-06-30,3,昏昏欲睡…只有音乐能听。,66%31%3%,0
34,"1988年的妮可 Nico, 1988",7.5,565,15.2%48.2%32.3%3.4%0.8%,2019-10-05,小青龙,2018-06-29,3,女主长相好像我的一个老师啊……,66%31%3%,0
```

Figure 1: The public available dataset on kaggle.com

3 Algorithm

3.1 Model

As shown in Fig 2, the core components of the proposed framework are two transformer-based encoders, namely, word encoder and sentence encoder. Both of these two encoders perform the representation learning with the help of self-attention networks [15, 16] multi-head attention mechanism[16], and residual connection [17]. Apart from these features, the word/sentence encoder further introduces a point-wise feed-forward networks to filter some irrelevant information in the transformer output.

Given the i-th review $x = \{x_{i1}, \ldots, x_{iN}\}$ of length N, the embedding component firstly maps the input words to the low-dimensional word embeddings. Then, the obtained word embeddings are fed to word-level encoder for calculating sentence representation s_i of the i-th review. Specifically, the word encoder is responsible for capturing word-word semantic dependency and word representation learning. The former one is achieved via self-attention and the transformation process is below:

$$
\begin{gathered}
\widehat{\mathbf{H}}^{l-1} = \mathrm{SLF-ATT}(\mathbf{Q}^{l-1}, \mathbf{K}^{l-1}, \mathbf{V}^{l-1}) \\
\mathbf{Q}^{l-1} = \mathbf{H}^{l-1}\mathbf{W}^{Q} \\
\mathbf{K}^{l-1}, \mathbf{V}^{l-1} = \mathbf{H}^{l-1}\mathbf{W}^{K}, \mathbf{H}^{l-1}\mathbf{W}^{V}
\end{gathered} \tag{1}
$$

$$
\mathrm{SLF-ATT}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{T}}{\sqrt{d_k}}\right)\mathbf{V}
$$

where $\mathrm{SLF-ATT}$ denotes multi-head self-attention and $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ refer to query, key and value respectively. d_k is the dimension of hidden representations and $\widehat{\mathbf{H}}^{l-1}$ are the transformed word representations encoding the global context. In order to filter the irrelevant information in the transformed word representations, a feed-forward networks FFN together with residual connection are applied:

$$
\begin{gathered}
\mathbf{H}^{l} = \mathrm{FFN}(\widehat{\mathbf{H}}^{l-1} + \mathbf{H}^{l-1}) \\
\mathrm{FFN}(x) = \mathrm{relu}(x\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2
\end{gathered} \tag{2}
$$

Where relu refers to relu activation function. The residual connection enables feeding the original word-level features to the calculation of the following layer and can reduce the training difficulty when the neural network is very deep. Note that we pack the word embeddings as $\mathbf{H}^0$ here and set the word encoder as a two-layer transformer. As with the aggregation of word-level features, we simply regard the first word representation as sentence representation:

$$
\mathbf{s}_i = \mathbf{H}_i^2[0] \tag{3}
$$

where $\mathbf{H}_i^2$ stores the transformed representations of words in the *i*-th review sentence.

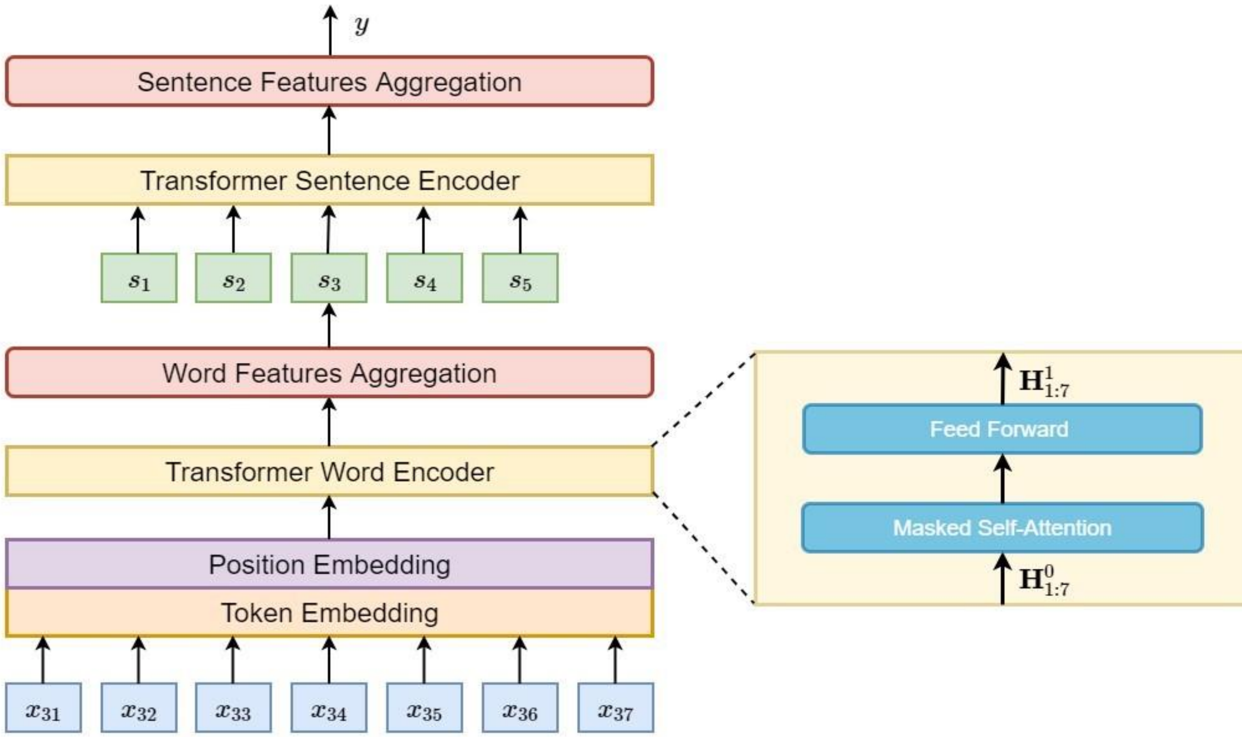

Figure 2: Architecture of the proposed model.

After obtaining the representations $\boldsymbol{s}_i$ ($i = 1; \cdots; M$) of the review sentence for a movie, we introduce sentence-encoder, another transformer-based encoder to calculate the overall review representation of movie. The computational process is as follows:

$$\begin{gathered}\hat{\mathbf{S}}^{l-1} = \mathrm{SLF-ATT}(\mathbf{Q}'^{l-1}, \mathbf{K}'^{l-1}, \mathbf{V}'^{l-1}) \\ \mathbf{Q}'^{l-1} = \mathbf{S}^{l-1}\mathbf{W}^{Q} \\ \mathbf{K}'^{l-1}, \mathbf{V}'^{l-1} = \mathbf{S}^{l-1}\mathbf{W}^{K}, \mathbf{S}^{l-1}\mathbf{W}^{V}\end{gathered} \tag{4}$$

$$\begin{gathered}\mathbf{S}^{l} = \mathrm{FFN}(\hat{\mathbf{S}}^{l-1} + \mathbf{S}^{l-1}) \\ \mathrm{FFN}(x) = \mathrm{relu}(x\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2\end{gathered} \tag{5}$$

Similar to word encoder, the sentence encoder firstly captures the intra-relation between the review sentences for the same movie via self-attention and employ feed-forward networks to refined the sentence-level representations. The final representation **r** for rating prediction is below:

$$\mathbf{r} = \mathbf{S}^{l}[0] \tag{6}$$

This overall representation is then fed to a linear layer with sigmoid activation to predict the overall movie score $\hat{y}$:

$$\hat{y} = \delta(r\mathbf{W}_3 + \mathbf{b}_3) \tag{7}$$

3.2 Model training

On the whole, the overall framework is fully differentiable and can be efficiently trained with gradient descent. We adopt Mean-Squared Error as loss function:

$$\mathcal{L} = \frac{1}{M}\sum_{j}\left(y_j - \hat{y}_j\right)^2 \tag{8}$$

Where M denotes the number of movies, the training objective is to minimize the value of the loss function $\mathcal{L}$

4 Experiment

4.1 Experiment Setup

We utilize the movie review dataset built by ourselves to evaluate the proposed framework. Since there are a lot useless reviews or fake reviews, we adopt a simple rule to filter such kind of review sentences under the assumption that the reviews with high "Like" are more likely to be useful for rating prediction. The statistics of the training/development/testing data set are given in Table 2.

Both the word encoder and the sentence encoder in our framework are 2-layer transformer. The hidden dimension of transformer and the number of head in multi-head self-attention are set as 768 and 8 respectively. To stabilize the training, we apply layer normalization [18] on each transformer output. We employ Adam [19] as optimizer and the initial learning rate is set as 1e-4. The batch size is 8 and we train the model up to 20 epochs. We do not use the word embeddings pre-trained on large-scale corpus to initialize the embedding layer but train it from scratch.

We employ jieba, a popular Chinese Word Segmentation toolkit to transform each review sentence to a word sequence.

Dataset	movies	reviews
Train	8141	421931
Dev	994	43859
Test	1062	26398

Table 2: Statistics for the used dataset.

4.2 Comparison Models

HCNN: HCNN models the internal structure of movie document (i.e., the collection of reviews for this movie) with hierarchical convolution based encoder. In our implementation, we employ multi-channel CNN [1] as feature extractor.

HLSTM: HLSTM sequentially encodes the word/sentence representations to obtain the high-level features for movie review prediction. In our implementation, we employ LSTM [20] as feature extractor.

LSTM-GRNN [6]: LSTMGRNN learns sentence representation with LSTM encoder first and then performing composition of sentence representations with gated recurrent neural networks.

HAN-CNN: HAN-CNN is a varient of Hierarchical Attention Networks [5] where both word encoder and sentence encoder are set as CNN.

HAN-LSTM [5]: HAN-LSTM is another variant of Hierarchical Attention Networks and LSTM is introduced to produce the word-level representations and sentence-level representations.

4.3 Evaluation Metrics

Since the problem studied in this paper is actually a regression problem, we adopt Mean-Squared Error (MSE),

Figure 3: *Movie ratings result of the proposed model*

one of the most commonly-used metric, to evaluate the model's capability of review rating. The formula of MSE, identical to Eq 8, is as follows:

$$\mathrm{MSE} = \frac{1}{M}\sum_{j}\left(y_j - \hat{y}_j\right)^2 \tag{9}$$

5 Result and discussion

5.1 Result of the model output

With the above algorithm and experiment methods, we got the output movie ratings using two proposed models. The first model named Model 1 is trained with additional data clearing procedures, like filtering with like parameters with the truth that reviews having small like values are less useful. The other model named Model 2 here is trained without data clearing procedures. Figure 3 shows ten movies with the score range from 0.3 to 0.9 as an example set. We want to find that if additional data clearing procedure could help movie lovers to remove the "water" of one movie. Here we compared the actual score of these movies and the rating results from our two models. The first column is the actual score, which we think is the ideal movie score, since it comes from the large number of reviewers. However, some short comment reviewers would like to occupy the comments with rubbish contents, these comments may make the score higher or lower than the actual score. In this case, the third column of figure 3 shows the big difference between rating of model 2 and original score of the movie. Fortunately, with the improvement of the clearing step, the second column shows better performance to get closer to the real score. The output shows that, this additional procedure introduces better performance, that is to say, less error between the exact movie rating.

Figure 4 shows the corresponding rating bias between our model and Douban rating. It can be easily seen that the bias of our model 1 (with additional data clearing procedure) has less mean value and shows narrower error limits. This figure is a supplement to the above conclusion.

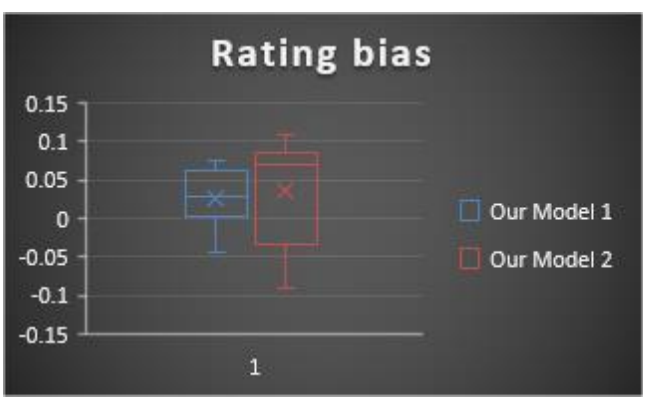

Figure 4: Movie rating bias of the proposed model

To examine what cause the better performance of model 1, we draw the attention graph as the function of *like* numbers. Figure 5 is the generated attention graph. Here we choose one movie with less reviews. The abscissa is each review with its "like" value, the ordinate is the attention of each review. This result proved that attention and the *like* value has highly relevant relationship. This result is also consistent with our common sense. After all, the review with more "like" means that more people agree with it, in general.

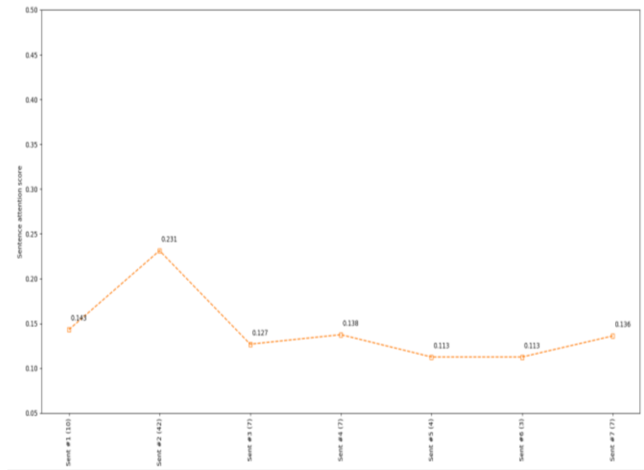

Figure 5: Generated attention graph

5.2 Discussion

As shown in Table 3, we compare our model with other five models (HCNN, HLSTM, LSTM-GRNN, HAN-CNN, HAN-LSTM). The criterion is mean-squared error (MSE).

Methods	*MSE (unit: 1e-3)*
HCNN	17.5
HLSTM	19.3
LSTM-GRNN	12.8
HAN-CNN	8.9
HAN-LSTM	6.5
Our Model	3.7

Table 3. Error analysis of semantic analysis model

Compared to other five models, our model obtains the best performance using our test set of Douban movie information.

6 Conclusion

In this project, we successfully perform rating evaluation on movie reviews. Compared with other common models, our model obtains the best performance.

In the comparison between two models training with/without additional data clearing procedure, the former model obtains better result. It reflects that the review which is more appreciated by others always expresses their emotion more accurately. Hence, the model training without useless reviews map emotion to rating more accurately.

At the same time, we successfully visualize attention map of each movie. Interestingly, we find that the final attention map is consistent with the "like" value of reviews. This shows that the review with more "like" value is always more accurate to the true rating of this certain object.

For further development, more efforts can be made to solve and remove junk reviews, such as manually labelling junk reviews or summarizing certain features. In this way, we can eliminate the interference of junk reviews and obtain accurate results.

On the other hand, much more cases can be trained to help improve the performance of this model, such as the reviews of Zhihu, Weibo, Youku and so on. The bigger, boarder datasets we utilize, the better performance we might obtain. Also, there may be more interesting conclusions or even some contrary to our common sense in our output.

REFERENCES

[1] Kim, Yoon: 'Convolutional Neural Networks for Sentence Classification'. Proc. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) pp. 1746-1751

[2] Johnson, Rie, and Zhang, Tong: 'Supervised and semi-supervised text categorization using LSTM for region embeddings'. Proc. Proceedings of the 33rd International Conference on International Conference on Machine Learning-Volume 48 pp. 526-534

[3] Johnson, Rie, and Zhang, Tong: 'Deep pyramid convolutional neural networks for text categorization'. Proc. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) pp. 562-570

[4] Zhang, Xiang, Zhao, Junbo, and LeCun, Yann: 'Character-level convolutional networks for text classification'. Proc. Advances in neural information processing systems pp. 649-657

[5] Yang, Zichao, Yang, Diyi, Dyer, Chris, He, Xiaodong, Smola, Alex, and Hovy, Eduard: 'Hierarchical attention networks for document classification'. Proc. Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies pp. 1480-1489

[6] Tang, Duyu, Qin, Bing, and Liu, Ting: 'Document modeling with gated recurrent neural network for sentiment classification'. Proc. Proceedings of the 2015 conference on empirical methods in natural language processing pp. 1422-1432

[7] Lai, Siwei, Xu, Liheng, Liu, Kang, and Zhao, Jun: 'Recurrent convolutional neural networks for text classification'. Proc. Twenty-ninth AAAI conference on artificial intelligence

[8] Howard, Jeremy, and Ruder, Sebastian: 'Universal Language Model Fine-tuning for Text Classification'. Proc. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) pp. 328-339

[9] Devlin, Jacob, Chang, Ming-Wei, Lee, Kenton, and Toutanova, Kristina: 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding'. Proc. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) pp. 4171-4186

[10] Martin, Lionel, and Pu, Pearl: 'Prediction of helpful reviews using emotions extraction'. Proc. Twenty-Eighth AAAI conference on artificial intelligence

[11] Yang, Yinfei, Yan, Yaowei, Qiu, Minghui, and Bao, Forrest: 'Semantic analysis and helpfulness prediction of text for online product reviews'. Proc. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers) pp. 38-44

[12] Liu, Haijing, Gao, Yang, Lv, Pin, Li, Mengxue, Geng, Shiqiang, Li, Minglan, and Wang, Hao: 'Using Argument-based Features to Predict and Analyse Review Helpfulness'. Proc. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing pp. 1358-1363

[13] Fan, Miao, Feng, Chao, Guo, Lin, Sun, Mingming, and Li, Ping: 'Product-Aware Helpfulness Prediction of Online Reviews'. Proc. The World Wide Web Conference pp. 2715-2721

[14] Chen, Cen, Qiu, Minghui, Yang, Yinfei, Zhou, Jun, Huang, Jun, Li, Xiaolong, and Bao, Forrest: 'Review Helpfulness Prediction with Embedding-Gated CNN', arXiv preprint arXiv:1808.09896, 2018

[15] Lin, Zhouhan, Feng, Minwei, Santos, Cicero Nogueira dos, Yu, Mo, Xiang, Bing, Zhou, Bowen, and Bengio, Yoshua: 'A structured self-attentive sentence embedding', arXiv preprint arXiv:1703.03130, 2017

[16] Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N, Kaiser, Łukasz, and Polosukhin, Illia: 'Attention is all you need'. Proc. Advances in neural information processing systems pp. 5998-6008

[17] He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, and Sun, Jian: 'Deep residual learning for image recognition'. Proc. Proceedings of the IEEE conference on computer vision and pattern recognition pp. 770-778

[18] Ba, Jimmy Lei, Kiros, Jamie Ryan, and Hinton, Geoffrey E: 'Layer Normalization', stat, 2016, 1050, pp. 21

[19] Kingma, Diederik P, and Ba, Jimmy: 'Adam: A method for stochastic optimization', arXiv preprint arXiv:1412.6980, 2014

[20] Hochreiter, Sepp, and Schmidhuber, Jürgen: 'Long short-term memory', Neural computation, 1997, 9, (8), pp. 1735-1780