Multi-view Representation Learning

A BTP Report

by

Swathi Kedarasetty

Roll No: S20180010172

Akula Haricha ndana

Roll No: S20180010009

Usarthi Harshini

Roll No: S20180010180

Under the guidance of

Dr. Viswanath P

INDIAN INSTITUTE OF INFORMATION

TECHNOLOGY SRICITY

16-12-2021

2nd Semester Report

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY SRICITY

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the BTP entitled "Multi-view Representation Learning" in the partial fulfillment of the requirements for the award of the degree of B. Tech and submitted in the Indian Institute of Information Technology SriCity, is an authentic record of my own work carried out during the time period from January 2021 to May 2021 under the supervision of Prof. Viswanath, Indian Institute of Information Technology SriCity, India.

The matter presented in this report has not been submitted by me for the award of any other degree of this or any other institute.

K. swathi

Signature of the student

(K. Swathi)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

P Viswanath
7.12.21

Signature of BTP Supervisor with date

(Prof. Viswanath)

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY SRICITY

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the BTP entitled "Multi-view Representation Learning" in the partial fulfillment of the requirements for the award of the degree of B. Tech and submitted in the Indian Institute of Information Technology SriCity, is an authentic record of my own work carried out during the time period from January 2021 to May 2021 under the supervision of Prof. Viswanath, Indian Institute of Information Technology SriCity, India.

The matter presented in this report has not been submitted by me for the award of any other degree of this or any other institute.

A. Hari Chandana

Signature of the student

(A.Harichandana)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

P Viswanath
7.12.21

Signature of BTP Supervisor with date

(Prof. Viswanath)

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY SRICITY

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the BTP entitled "Multi-view Representation Learning" in the partial fulfillment of the requirements for the award of the degree of B. Tech and submitted in the Indian Institute of Information Technology SriCity, is an authentic record of my own work carried out during the time period from January 2021 to May 2021 under the supervision of Prof. Viswanath, Indian Institute of Information Technology SriCity, India.

The matter presented in this report has not been submitted by me for the award of any other degree of this or any other institute.

U.Harshini

Signature of the student

(U.Harshini)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

P Viswanath 7.12.21

Signature of BTP Supervisor with date

(Prof. Viswanath)

ABSTRACT

Human beings perceive the external world via multiple senses, such as vision, hearing and touch. This implies that the information faced by the human brain is multi-view i.e., the brain is capable of multi-view learning. Therefore, it is important for the machines which are designed to tackle real-world problems for human beings to be able to learn representations from multi-view data. Learning an expressive representation from multi-view data, is a key step in various real world applications.

We've worked on the problem of personalized news recommendation on online platforms. News data being multi-view in nature, comprises different views like title, body, category. We've employed attentive multi-view learning to learn informative representations of the users and the news articles.

The objective of this project is to learn representations of news and users by incorporating information from different views like title, body, and topic category via an attentive multi-view learning framework, thereby achieving better performance in personalized news recommendation.

Contents

1. INTRODUCTION

The success of machine learning algorithms generally depends on the data representation, and we hypothesize that this is because different representations can entangle and hide more or less the different explanatory factors of variation behind the data. Multiview representation learning refers to settings where one has access to many “views" of data, at train time. Views often correspond to different modalities or independent information about examples. For instance, a scene that is represented as a series of audio and image frames, pictures taken from different angles of the same object etc

Online news platforms contain thousands of news articles and it is impossible and not necessary for every user to read all of these news articles. Therefore, personalized news recommendation is very essential for online news platforms to help users find their interested news and improve user experience. A core problem of such a recommendation system is to learn representations of news and users.

A news article usually contains different kinds of information such as title, body and topic category which are all useful for representing the news. These can be considered as the multiple views of a news article where each view may have different informativeness in representing different news articles. Therefore, we’ve built an attentive multi-view learning model to learn informative news and user representations by exploiting different views of the news data.

2. LITERATURE SURVEY

News recommendation is a well known problem and a good number of solutions have been proposed for this task. However, the existing methods rely on manual feature engineering for news and user representation learning. This needs a large amount of domain knowledge and effort to craft. Moreover, these methods cannot capture contexts and orders of words in news, which are important for learning news and user representations.

In the recent years, several deep learning based news recommendation methods have been proposed, like, learning news representations from news bodies using denoising autoencoders as in [], learning news representations via a knowledge-aware CNN from the titles of news articles etc.

However, these methods can only exploit a single kind of news information which may be insufficient for learning accurate representations of news and users. A representation that is able to explain many views of the data is more likely to capture meaningful variation than a representation that is a good fit for only one of the views. Therefore, multi-view representation learning with attention solves the above limitations and has given better performance for the news recommendation task by learning accurate representations of the news and user data.

3. METHODOLOGY

3.1 DATASET

MIND i.e Microsoft news dataset is widely used for news recommendation research which was collected from the anonymized behavior logs of Microsoft news website. It contains about 160k English news articles and more than 15 million impression logs generated by 1 million users. Impression logs contain information about a user's click behavior in a session. It is of the format:

Impression ID	ID of the impression
User ID	Anonymous ID of the user
Time	Impression time
History	News click history of this user before this impression
Impressions	List of news displayed in this impression and user's click behavior on them.

We've used a small version of MIND (MIND-small) which was obtained by randomly sampling 50k users and their impression logs.

3.2 MODEL

The recommendation model comprises of the following modules:

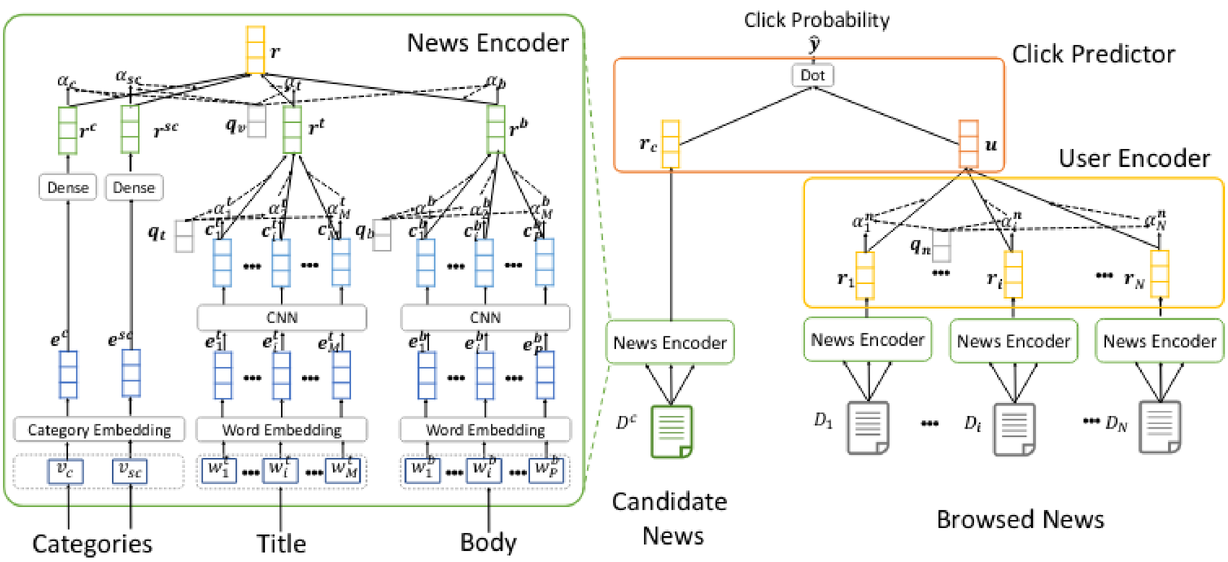
1. News encoder
2. User encoder
3. Click predictor

In the news encoder, we employed an attentive multi-view learning model to learn unified news representations from titles, bodies, and topic-categories by regarding them as different views of news. In addition, a view-level attention mechanism has been applied to select the important views to represent the news and word-level attention mechanism to select the important words in a particular view.

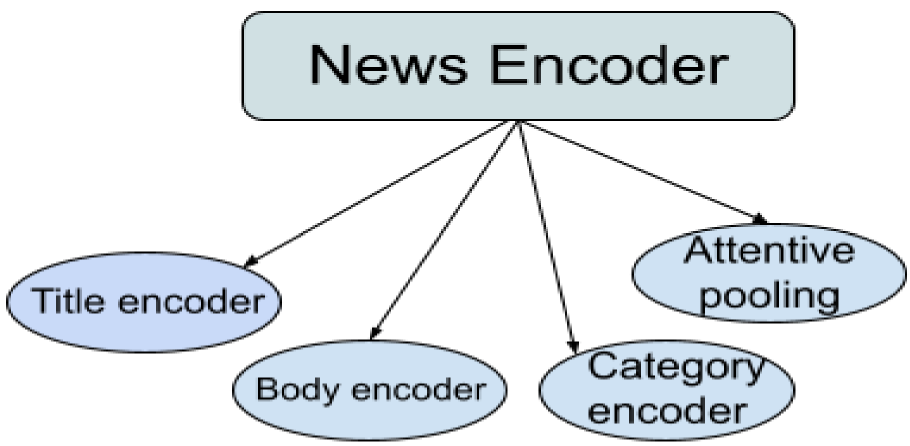
In the user encoder, the representations of the users have been learnt based on their browsed news. Attention mechanism has been employed to select the informative news to represent that user.

Click predictor is used to predict the probability of a user to browse a candidate news article. This is calculated based on the similarity of the user's representation and the news representation which can be computed by the inner product of the two vectors.

3.3 ARCHITECTURE



3.3.1 NEWS ENCODER



The four major components in news encoder are

1. Title encoder
2. Body encoder
3. Category encoder
4. Attentive pooling

Title encoder learns representations of news from title and body encoder learns representations of news from body of the news article. They both are similar, and have three layers:

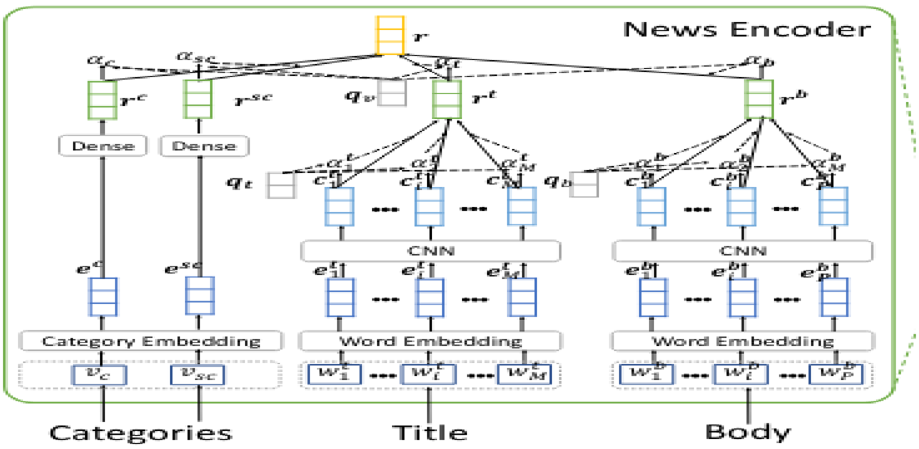
- The first layer is word embedding, which is used to convert the sequence of words(in title or in body) into a sequence of low dimensional semantic vectors.
- The second layer is CNN, which learns contextual word representations by capturing their local contexts. The output of this layer is a sequence of contextual word representations.
- The third layer is a word-level attention network. Different words in a news title/ body usually have different informativeness for learning news representations. Therefore, this attention network assigns attention weights to select important words within the context of each news title/body.

The final representation of the news title/body is the summation of the contextual word representations weighted by their attention weights.
Category encoder includes two layers

I. Category ID embedding layer : converts the discrete IDs of categories and subcategories into low dimensional dense representations.
II. Dense layer with ReLU activation function to learn the hidden category representations.

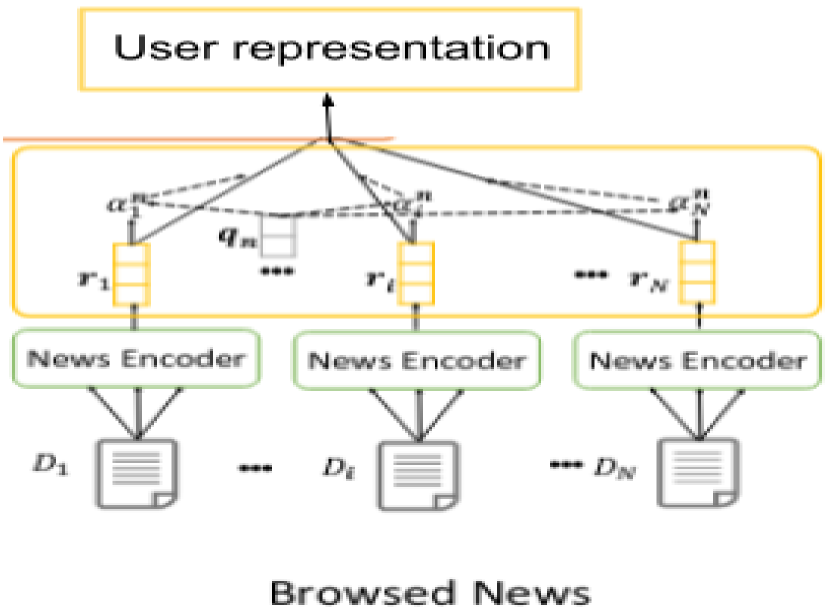
Attentive pooling includes a view-level attention network to select important views to represent each news article. This assigns attention weights to the different views i.e., title, body, category.

The final news representation learned by the news encoder module is the summation of the news representations from different views weighted by their attention weights.



3.3.2 USER ENCODER

Learns representations of users from the representations of the news browsed by them. The news encoder module is used to obtain the representations of the news that were browsed by a user, and then a news attention network is employed to give attention weights to the news browsed by the user.



The final representation of a user is the summation of the representations of the news browsed by him weighted by their attention weights.

3.3.3 CLICK PREDICTOR

Predicts the probability of a user to browse a candidate news article, based on their representations. This is calculated by the inner product of the representation vectors of the user and the candidate news article.

3.4 TRAINING

Model training is done using the negative sampling technique. From the “impression logs”, for each news browsed by the user, we randomly sampled K news which were presented in the same session and not clicked by the user.
We then jointly predicted the click probability scores of these K+1 news, and normalized them using soft-max.

The loss function used is the negative log-likelihood of the normalized click probability scores of all positive samples.

The negative sampling ratio K set to 4 has given us the appropriate results.

3.5 IMPLEMENTATION

Dataset split :

> The impression logs of the last week were used as the test set, and the rest logs were used for training.

Further, we randomly sampled 10% of the training samples for validation.

We’ve used pre-trained GloVE embedding for the word embedding.
The number of CNN filters was set to 400 and the window-size was 3. We further applied 20% drop-out to each layer in order to reduce overfitting.

After training and getting the predicted news on the test set, evaluation is done based on the following metrics:

- MRR (Mean reciprocal rank)
- AUC (Area under the ROC curve)
- NDCG @ 5 (Normalized discounted cumulative gain)
- NDCG @ 10

4. RESULTS

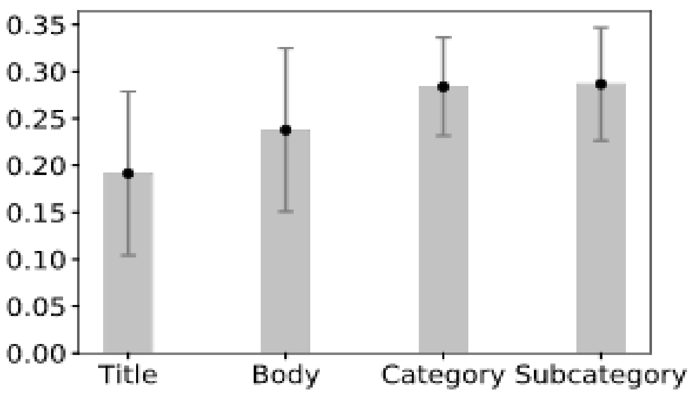
1. A sneak peek into the results obtained upon evaluating the model on the test set:

```
!python3 src/evaluate.py
```

```
Using device: cuda:0
Evaluating model NAML
Load saved parameters in ./checkpoint/NAML/ckpt-3000.pth
/usr/local/lib/python3.7/dist-packages/torch/utils/data/dataloader.py:481:
  cpuset_checked))
Calculating vectors for news: 100% 21/21 [00:03<00:00,  5.70it/s]
Calculating vectors for users:   0% 0/25 [00:00<?, ?it/s][W pthreadpool-cp
[W pthreadpool-cpp.cc:90] Warning: Leaking Caffe2 thread-pool after fork.
[W pthreadpool-cpp.cc:90] Warning: Leaking Caffe2 thread-pool after fork.
[W pthreadpool-cpp.cc:90] Warning: Leaking Caffe2 thread-pool after fork.
Calculating vectors for users: 100% 25/25 [00:46<00:00,  1.88s/it]
Calculating probabilities:   0% 0/73152 [00:00<?, ?it/s][W pthreadpool-cpp
[W pthreadpool-cpp.cc:90] Warning: Leaking Caffe2 thread-pool after fork.
[W pthreadpool-cpp.cc:90] Warning: Leaking Caffe2 thread-pool after fork.
[W pthreadpool-cpp.cc:90] Warning: Leaking Caffe2 thread-pool after fork.
Calculating probabilities: 100% 73152/73152 [02:32<00:00, 480.30it/s]
AUC: 0.6518
MRR: 0.3072
nDCG@5: 0.3380
nDCG@10: 0.4022
```

Therefore, the AUC = 0.6518, MRR = 0.3072, nDCG@5 = 0.3380, nDCG@10 = 0.4022.

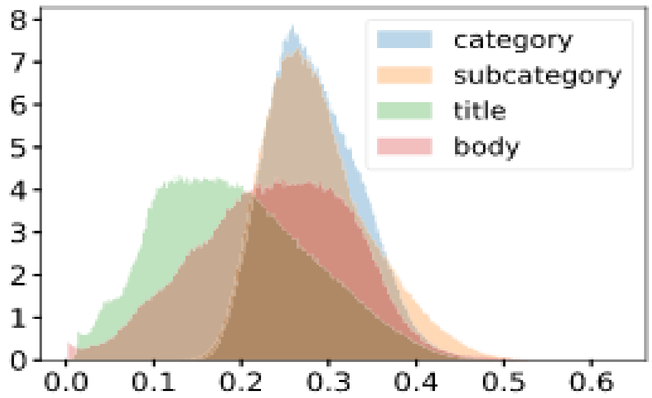
2. The mean and standard deviation of attention weights of the views is obtained as shown.



From this, it can be inferred that, the average attention weights of the body view are higher than those of the title view. This may be because bodies usually convey the original meanings of news and contain richer information than titles.

The category view gains the highest attention weight among all views.

3. The distribution of attention weights of the views is as shown.



From this, we conclude that the attention weights on the title and body views are small for many samples. This can be because some news titles and bodies are vague and uninformative. In these cases, categories contribute more in learning the news representation.

5. CONCLUSION

News recommendation is very important for online news platforms, and the accuracy of the recommendation algorithm depends on learning informative representations of the user and the news. Most of the classic and existing methods fail to incorporate information from the different views of the news data like title, body, category into its representation. Different from these baseline methods, our approach uses a multi-view framework to incorporate the information from different views. In addition, different words, different news and views may have different informativeness in representing news and users. The attention mechanism applied ensures that the important words, views and news are chosen which can build more informative news and user representations.

6. SOURCE CODE

The source code of our implementation can be found at https://github.com/swathi-vennela/NewsRecommendation-NAML repository on GitHub.

7. LIST OF FIGURES

8. LIST OF ABBREVIATIONS

MIND: MIcrosoft News Dataset

CNN : Convolutional neural networks

ReLU: Rectified Linear activation Unit

AUC: Area under the ROC Curve

MRR: Mean Reciprocal Rank

nDCG: Normalized Discounted Cumulative gain

9. ACKNOWLEDGEMENTS

[1]Li, Yingming, Ming Yang, and Zhongfei Zhang. "A survey of multi-view representation learning." IEEE transactions on knowledge and data engineering 31.10 (2018): 1863-1883.

[2]Okura, Shumpei, et al. "Embedding-based news recommendation for millions of users." Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017.

[3]Wu, Chuhan, et al. "Neural news recommendation with attentive multi-view learning." arXiv preprint arXiv:1907.05576 (2019).

[4]Wang, Hongwei, et al. "DKN: Deep knowledge-aware network for news recommendation." Proceedings of the 2018 world wide web conference. 2018.