Data Mining Final Project
Team # 3
NYCU 2024 Spring

511557010 李逸帆
411551039 陳柏翔
312605018 郭珈源

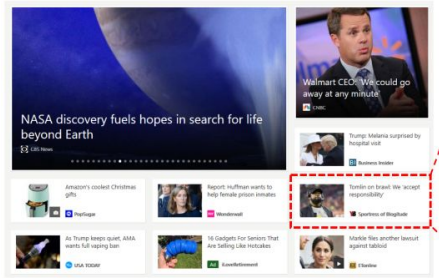
Abstract

- What we've done in this project
  - Identify origin dataset and recover ground truth for in-training evaluation
  - Build an efficient pipeline for news encoder
  - Reimplement and fix existing open-source MINER implementation
  - Flexible modularized architecture for future extensions
  - Intensive ablation studies on dos and don'ts
  - Detail error analysis on per category AUC and per user inspection

- Codes / Pre-trained weights are here for reproducibility
  - http://kertansul.synology.me:30000/playground/mimn

Introduction

- Situation
  - There are tons of news generated through web every day
  - As shown in the upper right figure, each news is composed by “Title”, “Category”, “Abstract” and “Body”
- Task
  - Given a history of news clicked by an user, we want to predict whether a user will be interested in a news (impression) or not



(a) An example Microsoft News homepage

Title	Mike Tomlin: Steelers ‘accept responsibility’ for role in brawl with Browns
Category	Sports
Abstract	Mike Tomlin has admitted that the Pittsburgh Steelers played a role in the brawl with the Cleveland Browns last week, and on Tuesday he accepted responsibility for it on behalf of the organization.
Body	Tomlin opened his weekly news conference by addressing the issue head on. “It was ugly,” said Tomlin, who had refused to take any questions about the incident directly after the game, per Brooke Pryor of ESPN. “It was ugly for the game of football. I think all of us that are involved in the game, particularly at this level, ...

(b) Texts in an example news article

News Click Sequence

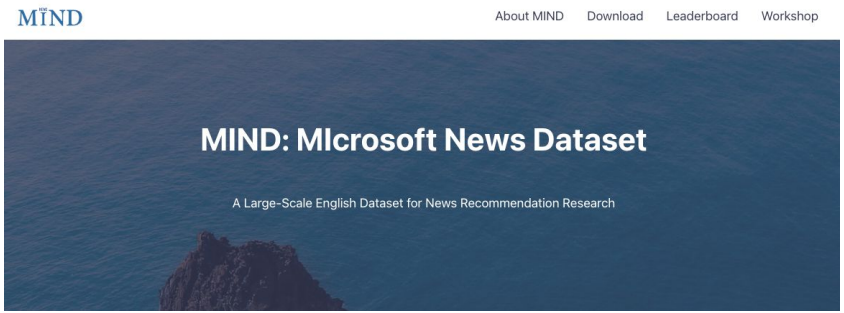
Category	*Title*
Finance	5 excellent dividend stocks to buy for the holiday season.
Sports	Should NFL be able to fine players for criticizing officiating?
Sports	5 takeaways from the 49ers' dominant win over the Panthers.
Movies	Francis Ford Coppola says Marvel movies are ‘despicable‘.
Sports	Magic vs. Cavs Preview: Magic basketball is finally back.
Fitness	This guy altered his diet and training to drop 65 pounds and pack on muscle.

Figure 1: The news click history of one user, who has various interests including finance, sports and movies.

MIND: A Large-scale Dataset for News Recommendation, Fangzhao Wu et al., 2020
MINER: Multi-Interest Matching Network for News Recommendation, Jian Li et al., 2022

Phase 1: Standing on the shoulders of others

- Challenge
  - We have little prior knowledge about news recommendation
- Action
  - Do some paper & benchmark survey
- Result
  - We identified that the dataset origins from Microsoft MIND
  - Therefore
    - We could narrow down to works in the benchmark table
    - We could try to recover ground truth labels for in-training model evaluation



https://msnews.github.io/

Phase 2: Prototyping - with open-source codes

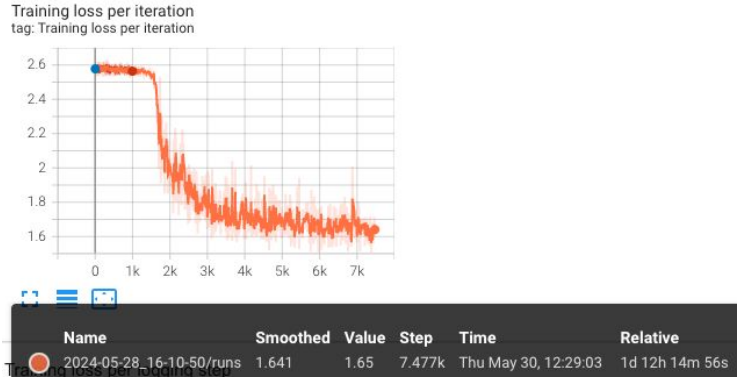
- Motivation
  - To quickly understand the difficulty of the problem
- Action
  - We found there are two open-source implementations
  - duynguyen-0203/miner
  - reczoo/RecZoo/pretraining/news/UNBERT

Leaderboard

Rank	Team	AUC	MRR	nDCG@5	nDCG@10
1 OCT. 05, 2021	UniUM-Fastformer-Pretrain	0.7304	0.3770	0.4180	0.4718
2 SEPT. 02, 2021	MINER	0.7275	0.3724	0.4102	0.4661
3 AUG. 08, 2021	UniUM-Fastformer	0.7268	0.3745	0.4151	0.4684
15 FEB. 01, 2023	kevin_zzh	0.7208	0.3676	0.4046	0.4608
16 FEB. 26, 2021	UNBERT	0.7207	0.3677	0.4041	0.4602
17 JAN. 19, 2022	only2233	0.7201	0.3688	0.4059	0.4617

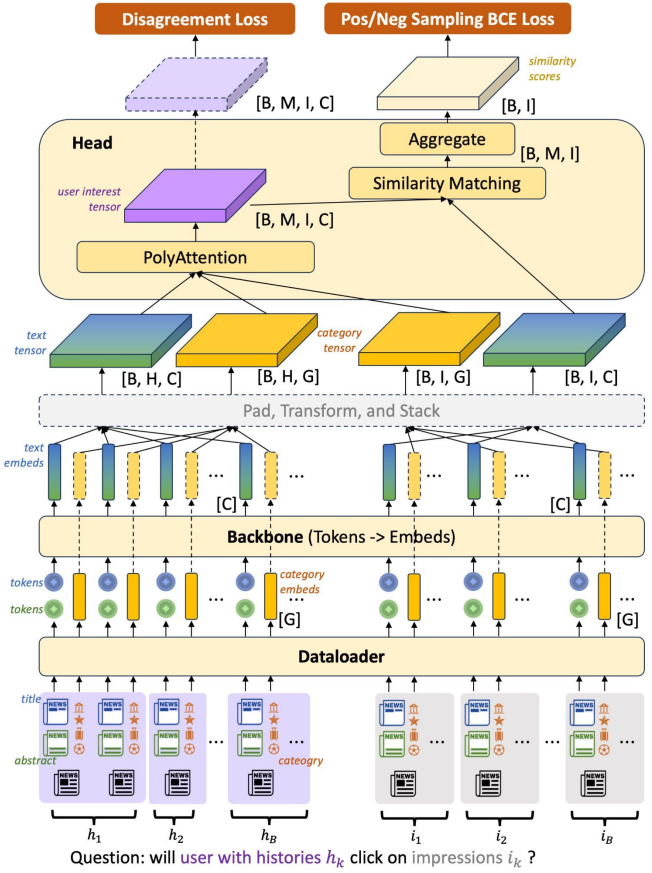
Phase 2: Prototyping - with open-source codes (cont.)

- Result
  - However, things didn't go well with the open-source implementations
  - For MINER
    - Only support single GPU
    - Require ~1 day to train a single epoch (iteration is too slow)
  - For UNBERT
    - Code cannot be executed directly;
      Some bugs exist and need to be resolved
    - After solving those bugs,
      training performance is not comparable
      with the numbers mentioned in paper
    - For more details, please check Appendix



Phase 3: Build an efficient pipeline

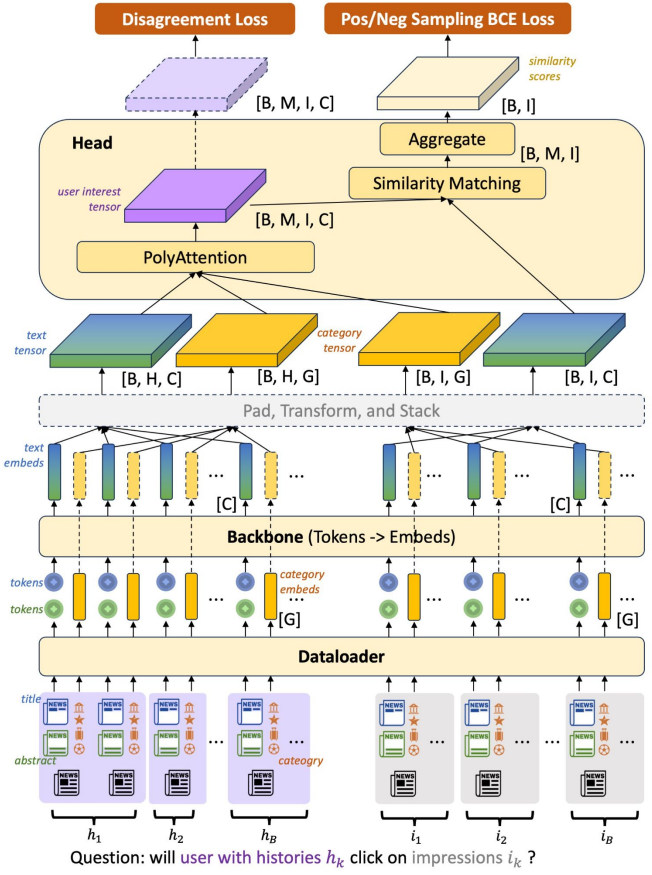
- Challenge
  - To support quicker experiment iterations, we need to find a way to speedup training
- Action
  - Major bottleneck is the news encoder layer
    - Since different behavior rows contains different history (clicked_news) length
    - It will be quite a waste of computation power if we pad history to “max_history_length” then send it into encoder
    - Instead, all news (history or impression) will be view separately
    - Pad and transform will be done after encoder
  - Include multi-GPU training + batch accumulation for best efficiency and performance
- Result
  - ~1.5 hr. per epoch after optimization



http://kertansul.synology.me:30000/playground/mimn

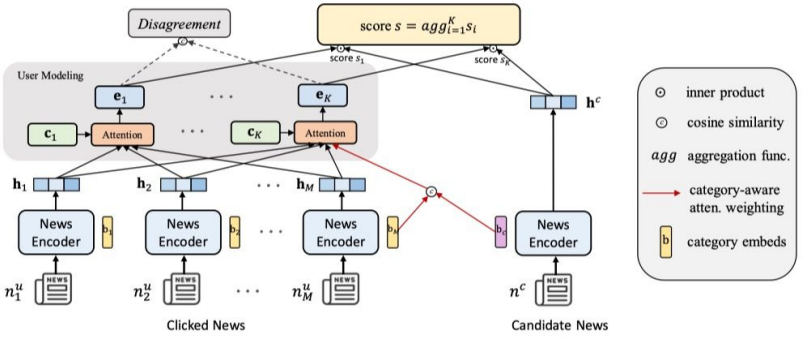
Phase 3 (cont.) pipeline details

- Architecture
  - There are 3 major components in MIMN
  - **Dataloader**
    - Read behavior.tsv and convert to
      - News token ids (title + abstract)
      - Glove embeddings (category)
  - **Backbone**
    - Build on top of Huggingface's transformers
    - Convert news token ids -> News embeddings
  - **Head**
    - Input
      - History embeddings & Impression embeddings
    - Output
      - User modeling embeddings
      - Similarity scores



Phase 4: Re-implement MINER

- Challenge
  - Open-source implementation (unofficial) is not reliable
- Action
  - Read the paper thoroughly then re-implement it ourselves
  - To be specific, we implemented the following modules
    - PolyAttention
    - TargetAwareAttention
    - MultiInterestMatchingNetwork
    - PosNegNceLoss
    - DisagreementLoss



$$\mathbf{e}_i = \sum_{j=1}^{M} w_j^{c_i} \mathbf{h}_j,\ w_j^{c_i} = \mathrm{softmax}(\phi_h^{c_i}(\mathbf{h}_j) + \underline{\lambda \cos(\mathbf{b}_j, \mathbf{b}_c)}),$$

$$\phi_h^{c_i}(\mathbf{h}_j) = \mathbf{c}_i^\top \tanh(\mathbf{W}^h \mathbf{h}_j)$$

$$\mathcal{L}_{NCE} = -\sum_{i=1}^{|\mathcal{D}|} \log \frac{\exp(s_i^+)}{\exp(s_i^+) + \sum_{j=1}^{L} \exp(s_i^j)}.$$

$$\mathcal{L}_D = \frac{1}{K^2} \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{\mathbf{e}_i^\top \mathbf{e}_j}{\|\mathbf{e}_i\| \|\mathbf{e}_j\|},$$

MINER: Multi-Interest Matching Network for News Recommendation, Jian Li et. al, 2022

Phase 5: Settings for Experiment

- Text preprocess
  - We follow the preprocess used by google-t5
  - Only remove some special characters and convert to lowercase
- Model selection
  - Most of our experiments are based on “bert-base-uncased” and “distilbert-base-uncased”
  - We’ve also tried “google-t5/t5-small”, “google-t5/t5-base”, “roberta-base”, “albert-base-v2” but didn’t yield stronger performance
- Hyperparameters
  - To grasp all our settings, please refer to our config file
  - An example snapshot is shown on the right

```
    max_num_history: 50
    max_title_length: 24
    max_abstract_length: 40
    history_order: last
    batch_size: 2

  architr:
    type: ForwardAllNewsThenSplitPn
    ForwardAllNewsThenSplitPn:
      backbone:
        type: NewsAutoEncoder
        NewsAutoEncoder:
          model_name: bert-base-uncased
          pool_method: cls
          out_dim: 512
          embed_dropout: 0.0
      head:
        type: MultiInterestMatchingNetwork
        MultiInterestMatchingNetwork:
          in_dim: 512
          num_interest: 4
          interest_dim: 256
          aggregation_method: mean
          use_category_bias: true
          category_dropout: 0.
          use_self_attn: false
          selfattn_dropout: 0.1
      loss:
        PosNegNceLoss:
          weight: 1.
          pn_ratio: 4
        DisagreementLoss:
          weight: 0.0
    pretrained_weights:

  optimizer:
    type: Adam
    Adam:
      lr: 2.e-4
      weight_decay: 1.e-6
```

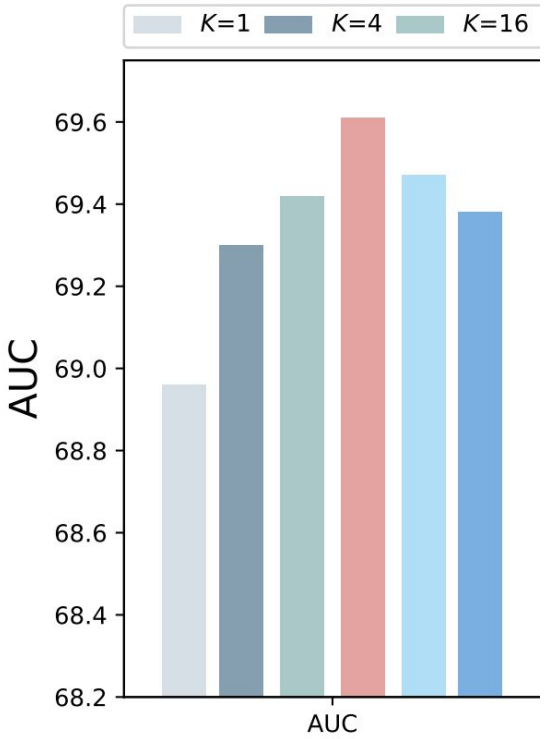
Question 2

針對於模型的架構與超參數，我們在架構與超參數上做了些實驗。

	AUC score
BERT-base-uncased	0.7057
DistilBERT	0.68274
DistilBERT num_interest_16	0.44858
DistilBERT with_Synonym Augmenter	0.67840
DistilBERT impression_news_without_dropout	0.67424
DistilBERT with_XLNet_encoder	

Num_interest_16

在 MINER 的論文中指出，the number of interest 在 32 時，效果最好，為此我們把參數從 4 先調整到 16 作一個階段性實驗，但在 num_intereset 16 時，效果卻從 0.68274 掉到 0.44858，猜測可能是因為新聞種類數量少，所以其實 num_intereset 數字太大，效果反而會變差。



With_Synonym Augmenter

我們在訓練的過程中，發現模型有 overfiting的現象，因此，我們使用了 Data Augmentation in NLP 專案中所提供的套件，其中的同義詞增強方法。

最後訓練的結果卻從 0.68274 掉到了 0.67840，我們猜想，可能是網路新聞的標題，存在某些領域用的特定詞彙，導致資料增強沒有如預期般上升，反而下降。但是也可能是我們硬體資源有限，訓練 epoch數量不多，或許再多跑1倍的時間，就會出現顯著的差異。

Synonym Augmenter

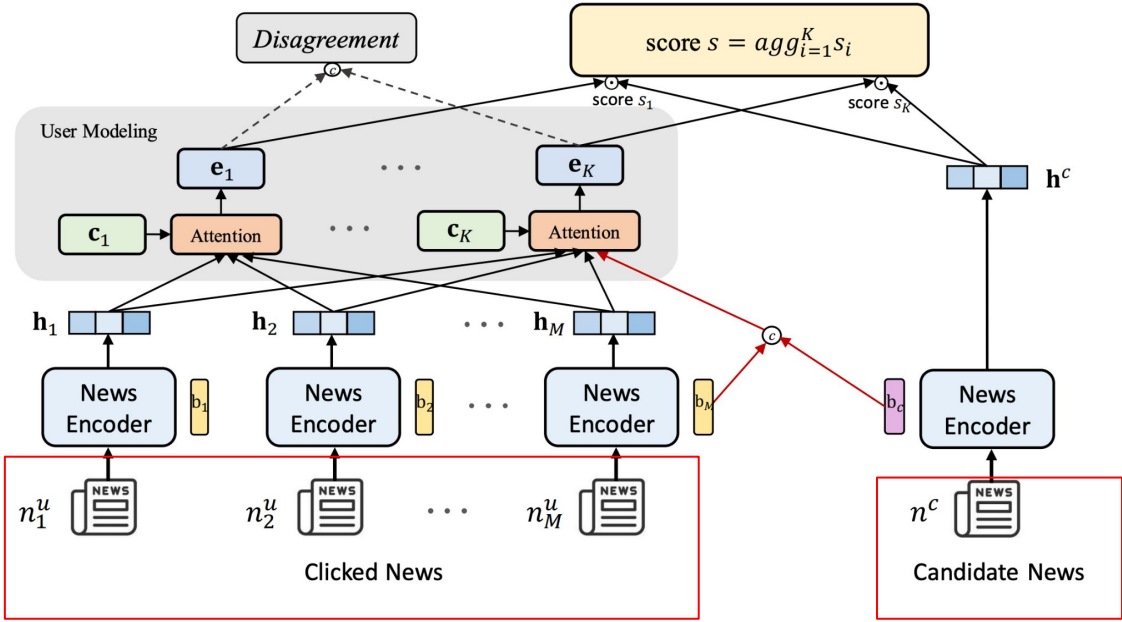
Substitute word by WordNet's synonym

```
aug = naw.SynonymAug(aug_src='wordnet')
augmented_text = aug.augment(text)
print("Original:")
print(text)
print("Augmented Text:")
print(augmented_text)
```

```
Original:
The quick brown fox jumps over the lazy dog .
Augmented Text:
The speedy brown fox jumps complete the lazy dog .
```

Impression_news_without_dropout

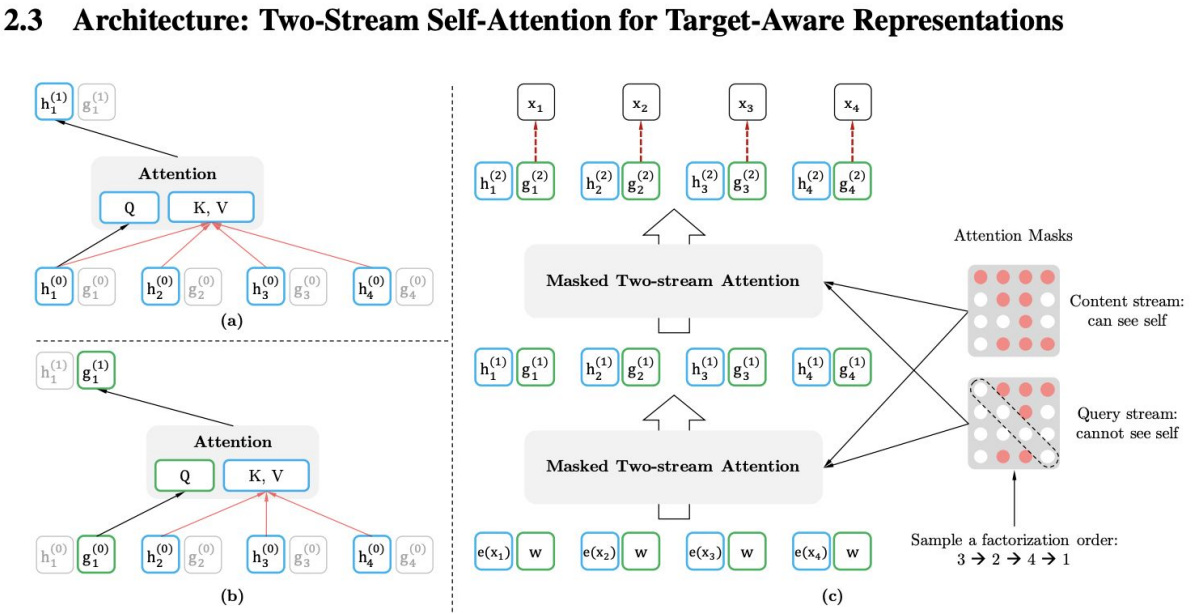
因為 Dropout 的挖空有時可能會導致效果變差，對於這點，我們保留了Clicked News的dropout，並對Candidate News的Dropout去除掉。結果從0.68274降到0.67424，也有可能只是在誤差範圍內。



DistilBERT with_XLNet_encoder

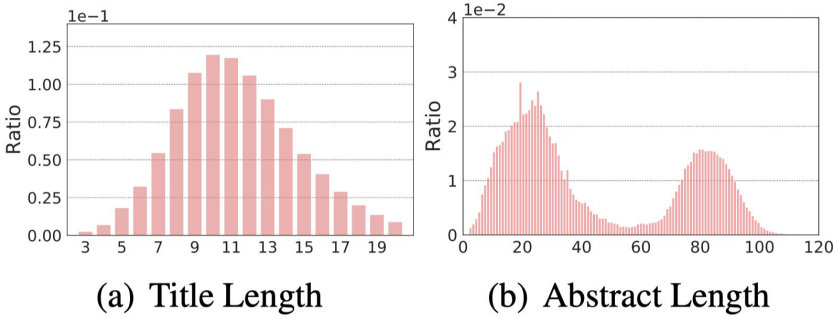
XLnet 是一種類似於BERT的模型，採用二階訓練的技術。與BERT的差別在於，XLNet採用了 Autoregressive 方式訓練，而 BERT使用了 Autoencoding。

訓練中採用了 Permutation Language Modeling(PLM)方法，消除了 Mask 同時遮蔽掉重要相關詞彙的問題。

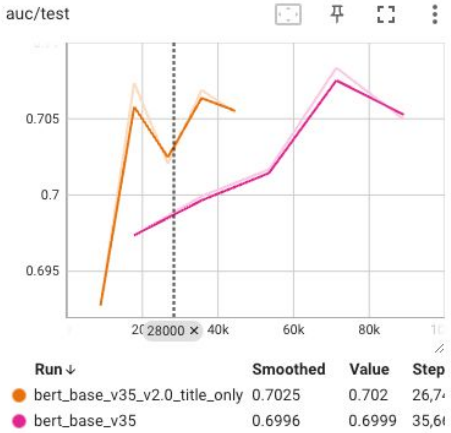


Empirical Findings 1 - title only is sufficient for the task

- Observation
  - Graph on the right is the title length & abstract length histogram from MIND's paper
  - However, in MINER's paper, author only uses title information
- Experiment
  - We did a comparison between the following settings
    - Title: 24 + Abstract: 40
    - Title: 32
      (both are trained with history 50)
- Result
  - Performance is comparable
  - However, title 24 + abstract 40 got a slightly lead



(a) Title Length (b) Abstract Length



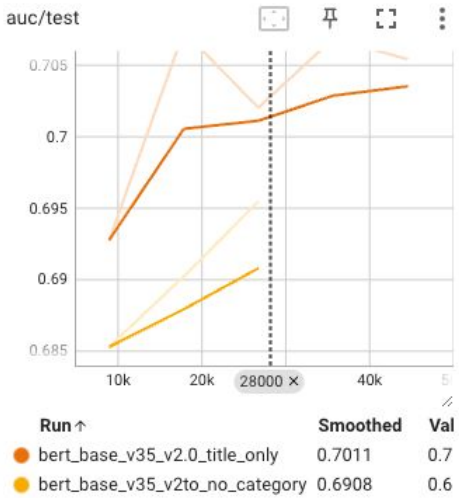
MIND: A Large-scale Dataset for News Recommendation, Fangzhao Wu et al., 2020

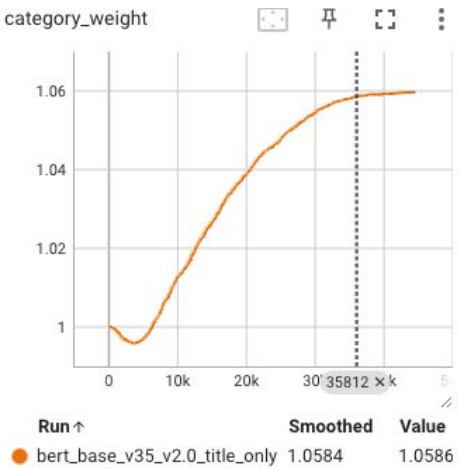
Empirical Findings 2 - some categories must be renamed

- Observation
  - Following MINER's method, we use Glove 6B with 300-dimension to encode news categories
  - However, there are 3 category names that are not in the vocabulary, leading to zero vector
  - Thus, we change
    - northamerica -> north-america
    - middleeast -> middle-east
    - foodanddrink -> food
- Result
  - We do find a slight performance boost after the fix

Empirical Findings 3 - category bias slightly boost performance

- Experiment
  - To understand the importance of “category” information, we turn on/off category bias to see its difference
- Observation
  - There’s ~0.01 AUC performance gap between them (/w category information is better)
  - We also monitors the category bias’ weight
    - Slightly grows over time
- Conclusion
  - Category information is important, must include





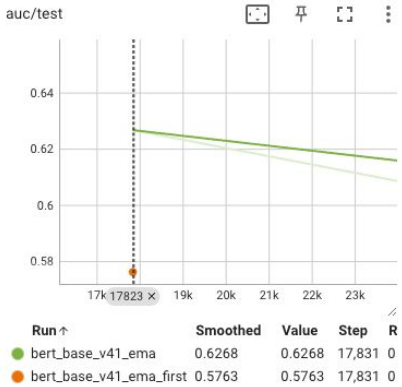
Empirical Findings 4 - history order matters

- Observation
  - Although not mentioned in paper, we suspect that the history is in-order (not reverse-order)
  - The most recently clicked news is located at the last position of the sequence
  - Therefore, we think that it is crucial to reverse the history first, then do any cutoffs if necessary
- Result
  - There's a huge performance difference between "last-first" & "first-first" history order
  - As suspected, "last-first" history ordering has a ~0.05 performance lead

News Click Sequence

Category	*Title*
Finance	5 excellent dividend stocks to buy for the holiday season.
Sports	Should NFL be able to fine players for criticizing officiating?
Sports	5 takeaways from the 49ers' dominant win over the Panthers.
Movies	Francis Ford Coppola says Marvel movies are 'despicable'.
Sports	Magic vs. Cavs Preview: Magic basketball is finally back.
Fitness	This guy altered his diet and training to drop 65 pounds and pack on muscle.

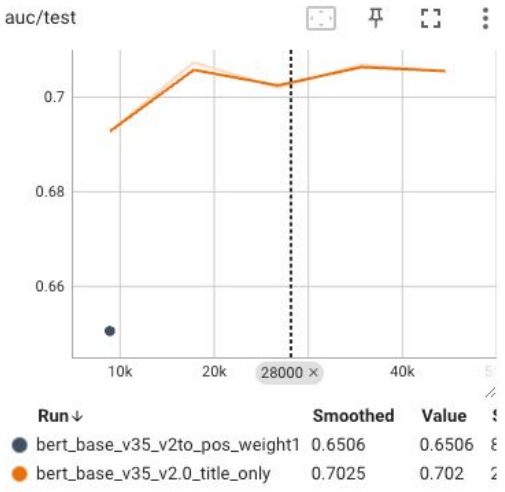
Figure 1: The news click history of one user, who has various interests including finance, sports and movies.



Empirical Findings 5 - sampling & pos_weight is crucial

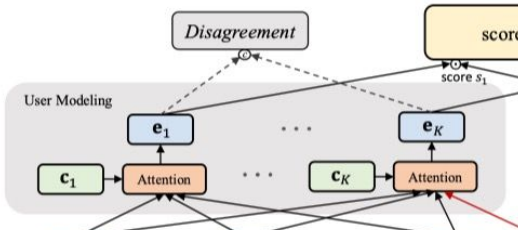
- Observation
  - In PosNegNceLoss, we found that it is crucial to
    - Sample positive and negative pairs
      - We use pn_ratio = 4
    - Should give the positive sample a higher weight for precision-recall balancing
      - We use pos_weight = pn_ratio
- Result
  - As shown in bottle right, huge performance gap between pos_weight = 1 vs. pos_weight = 4
    - ~0.04 AUC difference

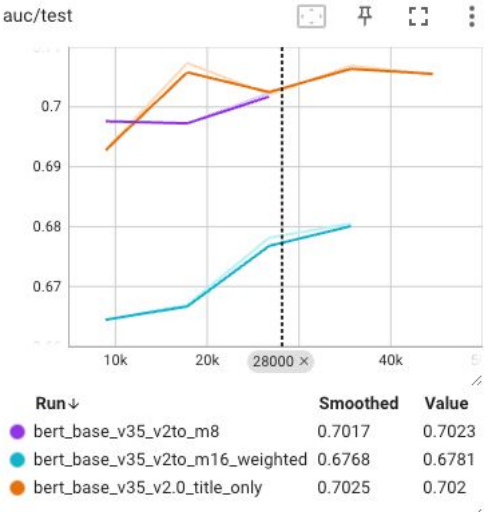
$$\mathcal{L}_{NCE} = -\sum_{i=1}^{|\mathcal{D}|} \log \frac{\exp(s_i^+)}{\exp(s_i^+) + \sum_{j=1}^{L} \exp(s_i^j)}.$$



Empirical Findings 6 - larger query size damage performance

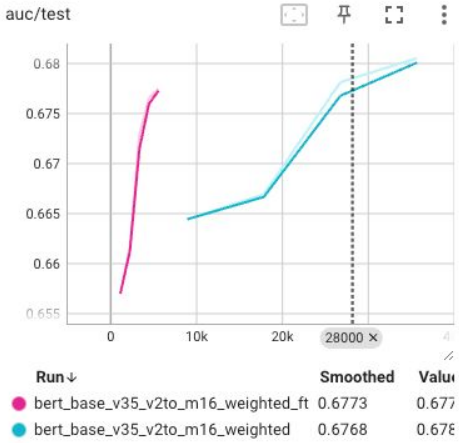
- Observation
  - Strangely, throughout our experiment, increasing the number of user interest leads to lower performance
- Experiment
  - Comparison between
    - Title_only: M = 4
    - M = 8
    - M = 16
- Result
  - M=4 leads to best performance, M=8 slightly lower, M=16 drops significantly
- Discussion
  - This contradicts with the phenomenon mentioned in MINER's paper. We did multiple verifications on our implementation but it doesn't seem wrong. May conduct more investigation into this if we have more time.





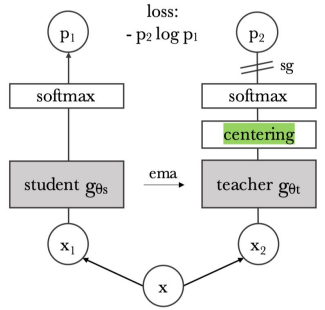
Empirical Findings 7 - don’t partially freeze backbone

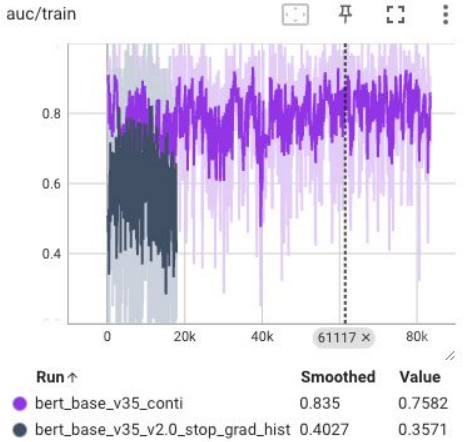
- Motivation
  - In the fear of model overfitting, we tried several methods to restrict model strength
- Experiment
  - Partially freeze backbone layers vs. end-to-end
  - Dropouts on news embeddings vs. no dropouts
- Result
  - Neither partial frozen or dropout gain performance
  - Instead, both of them got lower performance than the one using full model capability
- Thoughts
  - We suspect that because Bert is not originally trained for single sentence classification
  - It requires quite a lot weight tuning to achieve good performance
  - Therefore, restricting model strength turns down AUC score



Empirical Findings 8 - don't stop gradient on single side

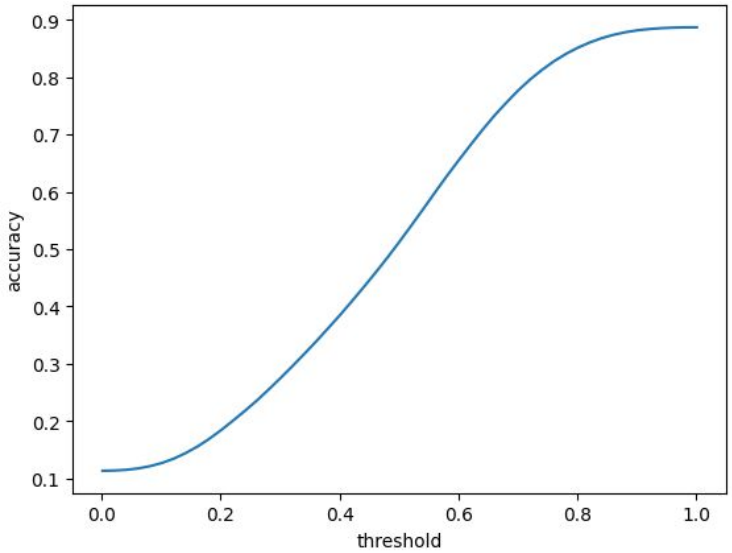
- Motivation
  - Inspired by BYOL and RL methods, we were wondering if stopping gradient on either side (history or impression) will gain stability on the learned embeddings
- Result
  - Nope! No good even in AUC/Train.
  - Don't do this in practice





Error Analysis

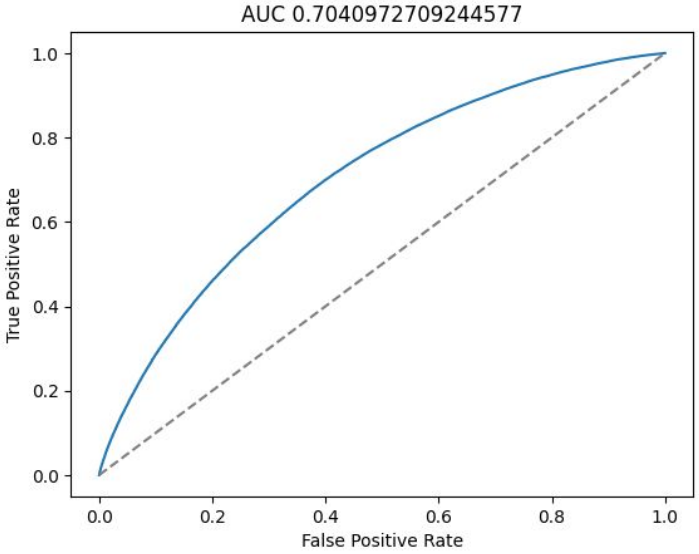
- Overview
  - At first, we tried to find a threshold that maximizes overall accuracy
  - However, since the dataset is biased toward negative impressions
    - # positive : # negative impressions ~ = 1 : 10
  - The best threshold is actually the one that outputs all zeros
  - Therefore, we cannot but to
    - Run analysis based on ROC (for high-level overview)
    - Show some example results (for detail-level inspection)



NOTE: the following results are based on bert_base_v35

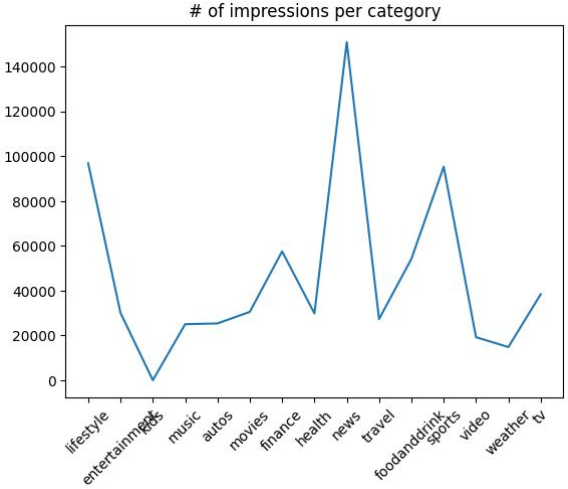
Error Analysis - Overall Performance

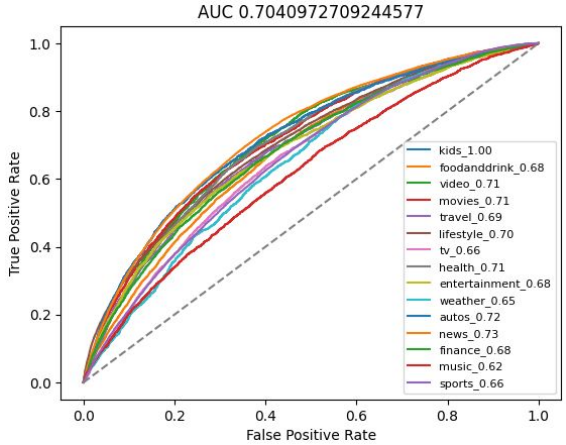
- Situation
  - Here we plot the TPR-FPR ROC
    - Blue line is our model's prediction
    - Gray dashed line is the ROC for all zero output
- Challenge
  - The curve is relatively smooth and hard to provide insights on how should we improve our model



Error Analysis - Category Performance

- Action
  - We tried to plot the per category ROC & per category count
- Observation
  - There are 3 major categories in the test dataset
    - News, Lifestyle, Foods and Drinks
  - Among the top 3 categories
    - News AUC: 0.73
    - Lifestyle AUC: 0.70
    - Foods and Drinks: 0.68
- Thoughts
  - We're highly underperforming in the "Foods and Drinks" section
    - Might because "Foods and Drinks" is a relatively large category, people interested in fast foods might not be interested in vegetarian
  - An improvement might be to make use of the "subcategory" information





NOTE: for kids category, there is no positive impression, AUC is set to a dummy value 1.0

Random inspection - 1

- Observation
  - User's history
    - Filled with sports category
    - 1 news related to celebs
  - Our prediction
    - high probability
      - Sports related news
      - News related to celebrity
    - Low probability
      - Homeless news
      - Food and Drinks

row index: 19822

===== history =====

	category	title
0	sports	Why ex-NFL star Kellen Winslow II finally plea...
1	sports	Luke DeCock: After Cole Anthony's spectacular ...
2	sports	Bears' Khalil Mack calls out Eagles C Jason Ke...
3	sports	Reports: LSU LB Michael Divinity removed from ...
4	sports	Browns stumble again, 9 other disappointments ...
5	entertainment	Celebs celebrate Halloween 2019
6	finance	The world's largest nuclear power producer is ...
7	sports	Former NBA first-round pick Jim Farmer arreste...
8	sports	Warriors' Russell ejected for going after offi...
9	sports	Brock Osweiler retires

===== impressions =====

	category	title	gt	pred
0	news	Homeless woman living in SUV with dogs moves i...	0	0.118186
1	tv	Pete Davidson, Kaia Gerber Are Dating, Trying ...	0	0.729664
2	lifestyle	Cows swept away by Hurricane Dorian found aliv...	0	0.424534
3	lifestyle	6 Ways You Can Unclog a Toilet Bowl Without a ...	0	0.760117
4	sports	No. 2 Ohio State heads to Rutgers as 51-point ...	0	0.816683
5	sports	Bold predictions for Week 12 in college football	0	0.770369
6	foodanddrink	85 Thanksgiving Recipes You Can Make Ahead	0	0.242003
7	lifestyle	Meghan Markle and Hillary Clinton Secretly Spe...	0	0.345688
8	tv	The Kardashians Face Backlash Over 'Insensitiv...	0	0.472882
9	entertainment	THEN AND NOW: What all your favorite '90s star...	0	0.755440
10	lifestyle	Meghan Markle's Lawyers Debunk Multiple False ...	0	0.540843
11	news	U.S. Troops Will Die If They Remain in Syria, ...	0	0.639977
12	news	5 arrested in connection with deadly shooting ...	0	0.643912
13	foodanddrink	Wendy's Is Turning 50 Years Old, And Is Giftin...	0	0.587785
14	sports	It's not just the Browns vs. Steelers. NFL has...	1	0.845905

Random inspection - 2

- Observation
  - User's history
    - Multiple interests: sports, celeb, music, food
  - Our prediction
    - High probability
      - Sport and celeb
    - Low Probability
      - Food, Finance, and Travel

row index: 35469

===== history =====

	category	title
0	sports	Clippers' Kawhi Leonard on load management cri...
1	tv	Tristan Thompson Drops Flirty Comment on Ex Kh...
2	tv	Bachelor in Paradise Stars Tayshia Adams and J...
3	music	Fans are crazy in love with Ciara and Russell ...
4	foodanddrink	How to Make Taco Mac n Cheese
5	sports	Why Frank Vogel and LeBron James both believe ...
6	sports	Kawhi Leonard scores 30, as Clippers overwhelm...
7	sports	Ian Rapoport: MRI results are good news for Pa...

===== impressions =====

	category	title	gt	pred
0	sports	7 possible landing spots for Anthony Rendon	0	0.650182
1	finance	Get the most out of your credit card with thes...	0	0.148968
2	foodanddrink	11 Regional Thanksgiving Recipes That Food Blo...	0	0.424334
3	sports	Lakers' Avery Bradley sidelined by hairline fr...	1	0.932721
4	finance	How to Set Boundaries When Your Family Is Bad ...	0	0.360295
5	travel	Five Great Holiday Cruises Worth Taking	0	0.294094
6	foodanddrink	41 Delicious Sweet Potato Recipes You'll Want ...	1	0.338790
7	foodanddrink	Simple Tip: Heat Leftover Pasta on the Stove	0	0.510905
8	tv	Pete Davidson, Kaia Gerber Are Dating, Trying ...	0	0.887354
9	foodanddrink	How to Make Chicken Parm Spaghetti Squash	1	0.429864
10	travel	Greek donkeys are still being abused as 'touri...	0	0.118484
11	sports	Lakers' Avery Bradley suffers hairline fractur...	0	0.952064
12	sports	Opinion: Kitchens unqualified to lead undiscip...	0	0.610393
13	health	Ready to Try CrossFit? Give This Beginner's WO...	0	0.568962
14	sports	Opinion: NFL had no choice but to send a clear...	0	0.789392

Random inspection - 3

- Observation
  - User's history
    - Sport, News, and some Lifestyle / Health
  - Our prediction
    - High probability
      - News and sport
    - Low Probability
      - Finances, Car and Celeb lifestyle

row index: 128
===== history =====

	category	title
0	sports	Why ex-NFL star Kellen Winslow II finally plea...
1	news	Dangerous fugitive in case of slain couple cau...
2	sports	Winners and losers in College Football Playoff...
3	sports	AP Top 25: Navy gives AAC 4 teams, 3rd-most by...
4	news	More Californians Could Lose Home Insurance Af...
5	lifestyle	Photographer Crosses Paths With A Black Cat Un...
6	finance	Most adults over 50 would rather die than do this
7	video	Where have Cape Town's great whites gone?
8	news	Mississippi woman found after being missing fo...
9	news	Florida needs python hunters. A man in Iran is...
10	sports	Forde-Yard Dash: This is how bad the playoff p...
11	sports	The Effect of Illinois' Upset Went Far Beyond ...
12	news	Tucson homeowner fatally shoots 2 men during a...
13	health	Flu Shot Seekers 'Chase After' Senior Dose Dur...
14	news	Hard Rock Hotel New Orleans collapse: Former s...
15	news	Man charged with 5 counts of first-degree murd...

===== impressions =====

	category	title	gt	pred
0	lifestyle	Archie's Photo Album: Prince Harry, Duchess Me...	0	0.197152
1	news	3 Indiana judges suspended after a night of dr...	1	0.735332
2	autos	Ford v Ferrari: the forgotten car at the heart...	0	0.525641
3	news	Liver transplant leads to unbreakable bond bet...	0	0.185341
4	news	California man convicted of torture-murder die...	0	0.737739
5	lifestyle	I'm A Queer Woman Dating A Trans Man & No On...	0	0.310154
6	travel	Here's how much and who you should be tipping ...	0	0.361777
7	sports	LeBron James on he and Tom Brady: 'We're gonna...	0	0.399850
8	news	Why did Santa Clarita shooting happen? Detecti...	0	0.470666
9	finance	Billionaires' success boils down to 3 simple t...	0	0.368789
10	autos	Mustang: all the wild and wonderful offshoots ...	0	0.570574
11	sports	Pittsburgh Steelers quarterback Mason Rudolph ...	0	0.825099
12	travel	Secrets for Scoring Hotel Room Upgrades	0	0.522495
13	news	Texas parole board recommends delaying Rodney ...	0	0.592461
14	tv	The Kardashians Face Backlash Over 'Insensitiv...	0	0.448544

Random inspection - 4

- Observation
  - User's history
    - Sports and Cars
  - Our prediction
    - High probability
      - Cars
    - Low Probability
      - Lifestyle and Travel

row index: 44504
===== history =====

	category	title
0	sports	With loss to Raiders, is Chargers QB Philip Ri...
1	autos	SEMA 2019 – About That Yokohama Booth #MTSEMA19
2	news	Trump's allies turned to online campaign in qu...
3	autos	Everything We Think We Know About the 2020 For...
4	news	Amelia Bambridge: Body of missing backpacker f...
5	weather	Wild fall freeze: Utah falls close to minus-35...
6	autos	What do the Corvette codes Z51, Z06, and ZR1 m...
7	lifestyle	17 photos that show the ugly truth of living i...
8	sports	This might be most ridiculous TD pass of Aaron...
9	finance	Jeff Bezos lost about $7 billion on Thursday
10	travel	Stowaway Discovered in Couple's Carry-On Luggage
11	autos	Ken Block's Home Garage Is Surprisingly Tastef...
12	autos	How Much Power Does the 2020 Chevrolet Corvett...
13	autos	The Bugatti Chiron Super Sport 300+ Briefly We...
14	lifestyle	Duchess Meghan Describes 'Really Challenging' ...
15	news	'Serial Stowaway' Marilyn Hartman Held Without...

===== impressions =====

	category	title	gt	pred
0	autos	Ford v Ferrari: the forgotten car at the heart...	1	0.853010
1	travel	Five Great Holiday Cruises Worth Taking	0	0.343469
2	tv	The Kardashians Face Backlash Over 'Insensitiv...	0	0.491250
3	lifestyle	Kate Middleton Took Public Transit to Her Roya...	0	0.397959
4	lifestyle	66 Cool Tech Gifts Anyone Would Be Thrilled to...	0	0.240839
5	autos	Ford v Ferrari: the real story	0	0.764060
6	finance	This stately home is having the ultimate yard ...	0	0.646861
7	lifestyle	Meghan Markle's Lawyers Debunk Multiple False ...	0	0.608930
8	news	U.S. Troops Will Die If They Remain in Syria, ...	1	0.630473
9	news	California and nearly two dozen other states s...	0	0.581771
10	sports	Former North Carolina State, NBA player Anthon...	0	0.588664
11	sports	Report: Police investigating woman's death aft...	0	0.658669
12	travel	Maze of tunnels reveals remains of ancient Jer...	0	0.653801
13	health	I Overate at Dinner Until I Started Doing This...	0	0.469623
14	movies	13 Reasons Why's Christian Navarro Slams Disne...	0	0.499792

Random inspection - 5

- Observation
  - User’s history
    - Lifestyle, news, and some finance
  - Our prediction
    - High probability
      - Health, news
    - Low Probability
      - Celeb, and music

row index: 27842

===== history =====

	category	title
0	lifestyle	Couple Cancels Wedding, Keeps $30K as "Donatio...
1	lifestyle	40 Etiquette Mistakes You Need to Stop Making ...
2	news	The News In Cartoons
3	news	Texas attorney convicted of scamming drug traf...
4	news	5 charged in alcohol poisoning death of UC Irv...
5	news	Fox News contributor: 'Most likely' outcome is...
6	lifestyle	Divers Find Giant Mysterious 'Egg' Floating In...
7	news	Deer fatally attacks hunter who shot him
8	finance	Caterpillar just flashed the latest warning si...
9	autos	Bloodhound Shows The Car That Will Attempt 1,0...
10	news	Democrats' 2020 race has a new shadow: Hillary...
11	autos	Shut Down: 100s of Muscle Cars, EVs, and More ...
12	news	California governor pardons 3 convicted immigr...
13	finance	Harley-Davidson halts production of new electr...

===== impressions =====

	category	title	gt	pred
0	news	Trump attacks ambassador on Twitter as she tes...	0	0.600116
1	news	House to hear from US official who said he ove...	0	0.433694
2	finance	Nurses face an epidemic of workplace violence	0	0.353717
3	finance	10 reasons it's better to rent rather than buy...	0	0.423126
4	entertainment	How the biggest stars of the decade have changed	0	0.087974
5	sports	Opinion: Colin Kaepernick is about to get what...	1	0.646592
6	lifestyle	Please Don't Yell at Your Dog!	0	0.693438
7	entertainment	Best celebrity hair moments of 2019 (so far)	0	0.232171
8	news	Cellphone call from Ukraine could compound Tru...	0	0.683246
9	health	Cannabis Use Disorder is Rising in U.S. States...	1	0.877213
10	entertainment	Stars They're Just Like Us!	0	0.375036
11	travel	What Happens When Your Cruise Has to Rescue An...	0	0.382277
12	news	Multiple Lawmakers Under Investigation Over Et...	0	0.468745
13	movies	13 Reasons Why's Christian Navarro Slams Disne...	0	0.482071
14	music	Fashion hits and misses from the 2019 Country ...	0	0.232788

Conclusion

- To conclude, after random inspection on test data
  - We think our model learns what users do like and don't
  - However, user's behavior depend on various factors (not just by browsing history)
  - We think that more information is required to further boost performance (e.g. user's stay time on specific news (like tiktok) ...etc)
- As mentioned in Abstract, what we've done in this project
  - Identify origin dataset and recover ground truth for in-training evaluation
  - Build an efficient pipeline for news encoder
  - Reimplement and fix existing open-source MINER implementation
  - Flexible modularized architecture for future extensions
  - Intensive ablation studies on dos and don'ts
  - Detail error analysis on per category AUC and per user inspection
- Codes / Pre-trained weights are here for reproducibility
  - http://kertansul.synology.me:30000/playground/mimn

References

- MIMN Gitlab
- MIND dataset
- MINER: Multi-Interest Matching Network for News Recommendation, Jian Li et. al, 2022

Thank You

Appendix A: UNBERT

1. 解決什麼問題？ -> UNBERT注意到了，新聞的標題內容作為 embedding 是不足夠的，用戶在選擇推薦的新聞時，也會關注標題的關鍵字，而全是整個語意。

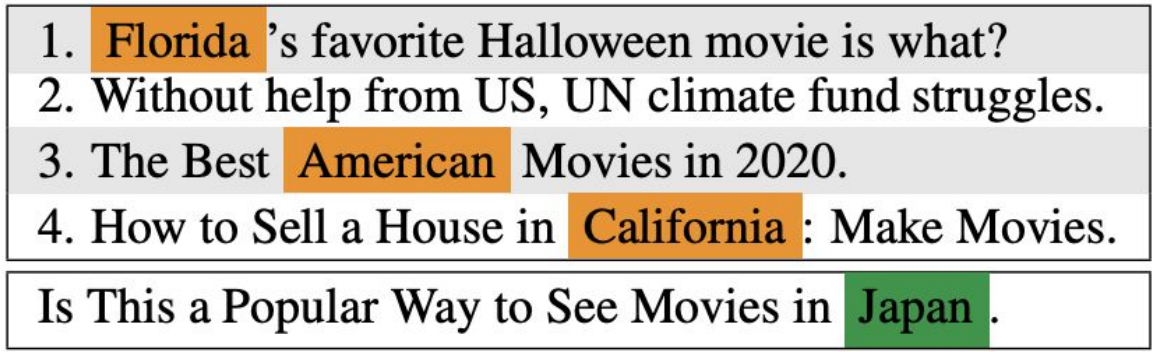
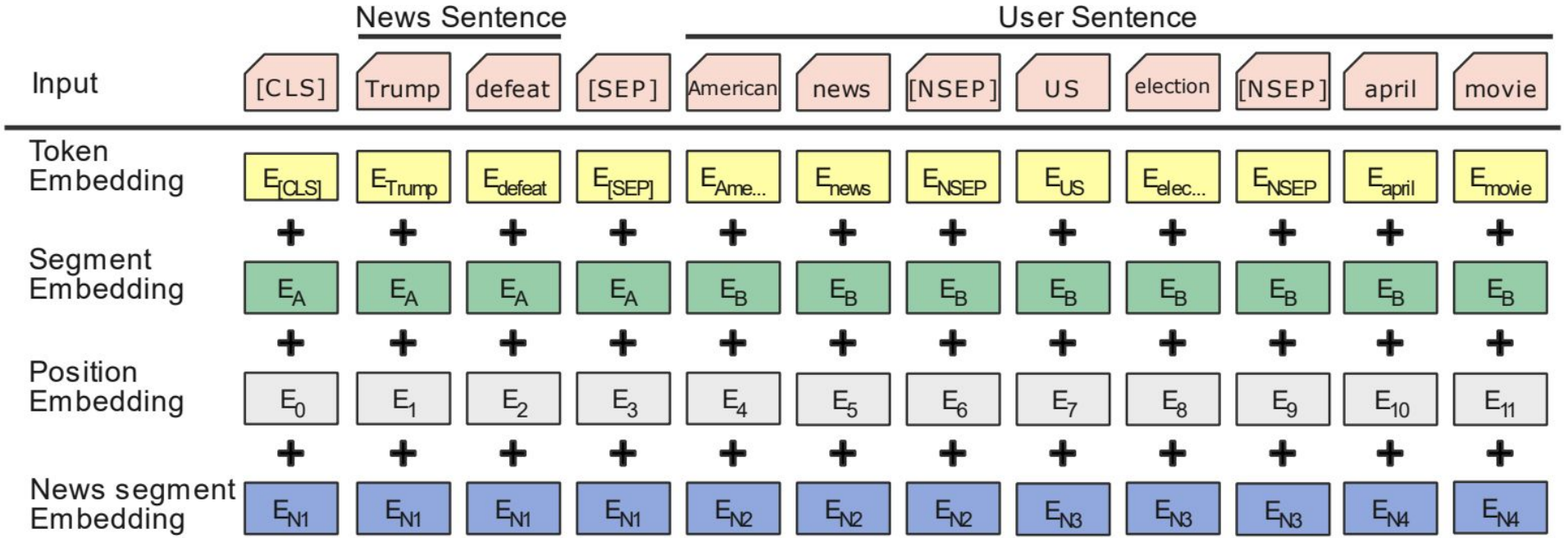


Figure 1: A negative example: several news browsed by a user (upper box) and a candidate news (lower box). Orange bars represent the important signals related with green bar that should be captured.

UNBERT

2. 模型的輸入資訊，包含，候選的新聞標題，與用戶過去點擊新聞的標題。除Token embedding，還有，區分候選與歷史紀錄(Seggment)、位置資訊(Position)、每個新聞的範圍(News segment)。



UNBERT

3. 模型如何運行 -> 模型在運行上，同時考量了文字(World-Level)與新聞語意(News-Level)。模型經過world-level的過程，會再WL與NL端，匯聚成一篇新聞的語意(根據範圍)。最終，WL與NL會輸出e(w)與e(n)，並進行點擊的預測。

在這裡TL為Transformer Layer模塊。

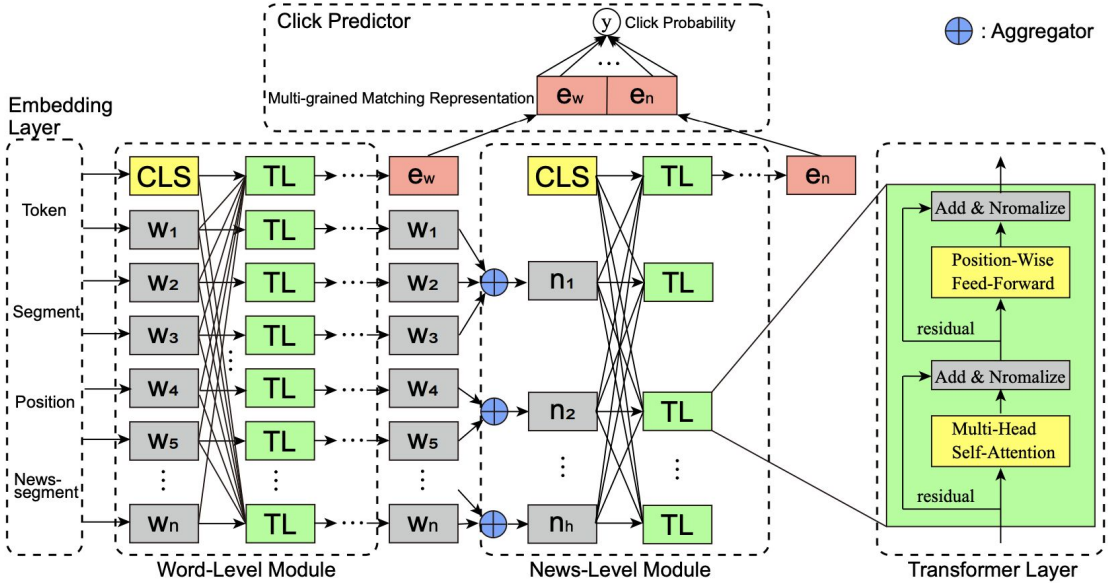


Figure 3: The overall architecture of our UNBERT approach.

Result

根據了論文所提供的數據，我們設計不同的參數做實驗，結果皆與論文的分數，0.6762 相差甚遠。

With	AUC score
	0.5528
History news: title + abstract	**0.5844**
Predicting news: title +abstract	0.5537
Predicting news: title +abstract History max length 30	0.5577