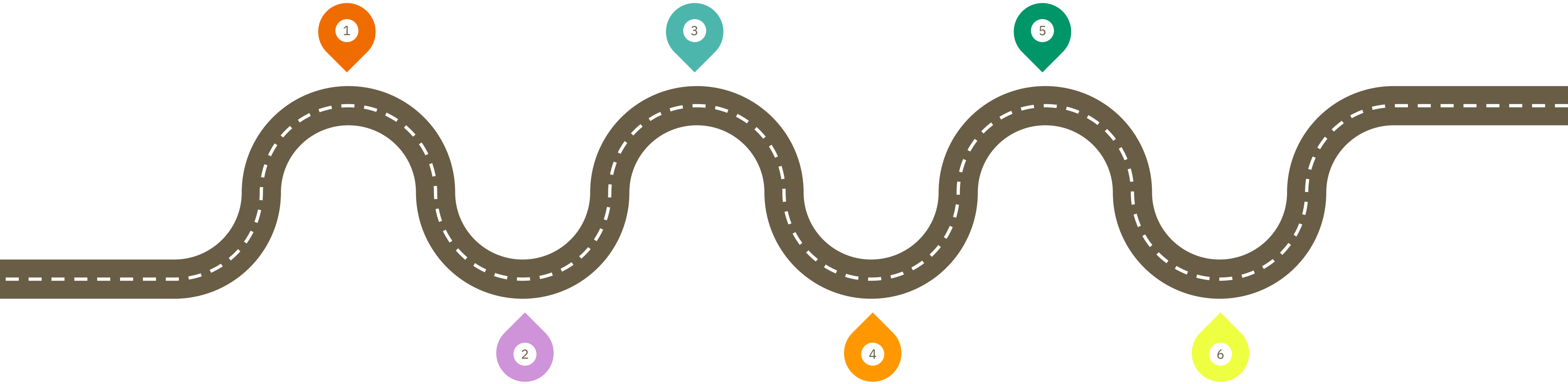
Multi View Representation Learning

B21PV02

Progress



1. Literature review on multi view representation learning
2. Understanding different CCA techniques for multi-view learning
3. Implementation of deep CCA i.e., two view non-linear representation learning on noisy MNIST dataset
4. Exploring a real-world problem that can be efficiently solved using multi-view learning
5. Implementation of multi-view representation techniques on the real world problem
6. Analysis on the improvement in results upon incorporating information from multiple views

Problem Statement

Personalized news recommendation

News recommendation

- Objective : To improve user experience with personalized news recommendation
- News - multi view data
  - Title
  - Body
  - Category etc
- Learn informative representations of users and news by exploiting information from the multiple views of the news data

Dataset

MIND Dataset -

- Around 160k English news articles
- Around 15 million impression logs generated by around 1 million users

The performance of recommendation algorithm depends on the informativeness of the news and user representations learnt

Attentive Multi view learning approach

To incorporate information from different views as per their importance

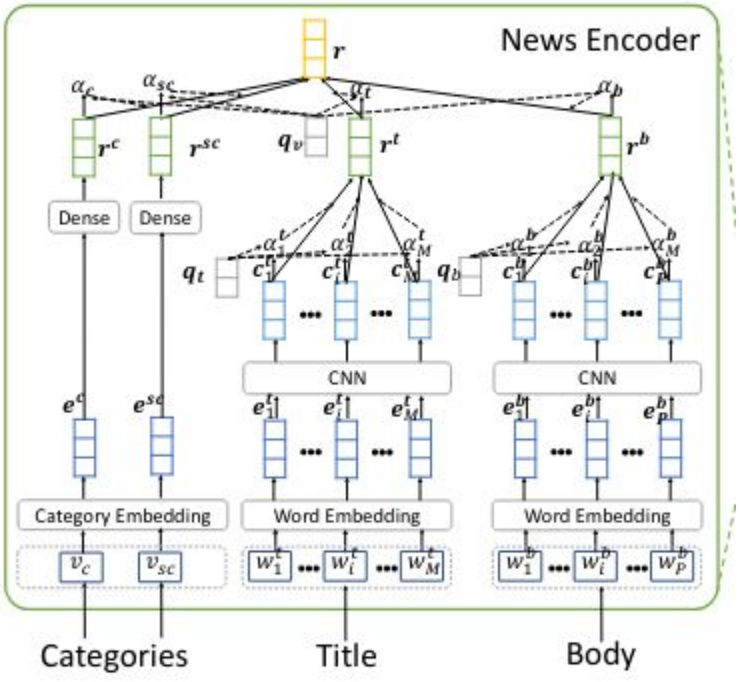
Attentive Multi-view learning framework

News Encoder → User Encoder → Click Predictor

News Encoder

Four major components -

- ➢ Title Encoder
- ➢ Body Encoder
- ➢ Category Encoder
- ➢ Attentive pooling



Title Encoder

1. Word Embedding

Converts a word sequence into a sequence of low-dimensional semantic vectors

$[w_1^t, w_2^t, \ldots, w_M^t]$ is converted into $[e_1^t, e_2^t, \ldots, e_M^t]$

2. CNN

Contextual word representation of the i-th word is

$c_i^t = \mathrm{ReLU}(F_t \times e^t_{(i-k):(i+k)} + b_t)$

The o/p of this layer is a sequence of contextual representations.

3. Word level attention

To select important words within the context of each news title.

Say, α_i is the attention weight obtained for the i-th word,

Final representation of the news title:

$r^t = \Sigma \alpha_i c_i^t$ for i = 1 to M

Body Encoder

- Three layers - word embedding, CNN and attention network similar to the title encoder
- Final representation of news body is the summation of contextual word representations weighted by their attention weights

$r^b = \Sigma \alpha_i c_i^b$ for i = 1 to P, where P is the number of words in the news body

Attentive Pooling

- A view-level attention network to learn attention weights of title, body, category and sub-category

$$a_t = \mathbf{q}_v^T \tanh(\mathbf{U}_v \times \mathbf{r}^t + \mathbf{u}_v),$$

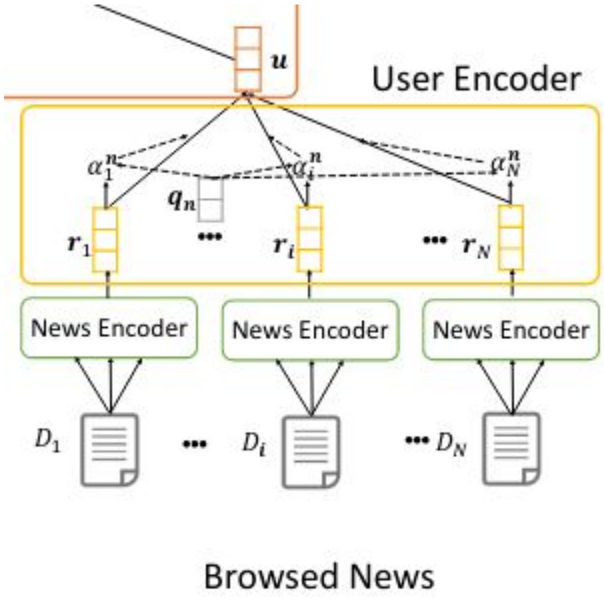
$$\alpha_t = \frac{\exp(a_t)}{\exp(a_t) + \exp(a_b) + \exp(a_c) + \exp(a_{sc})}$$

- Final news representation

$$r = \alpha_c r^c + \alpha_{sc} r^{sc} + \alpha_t r^t + \alpha_b r^b$$

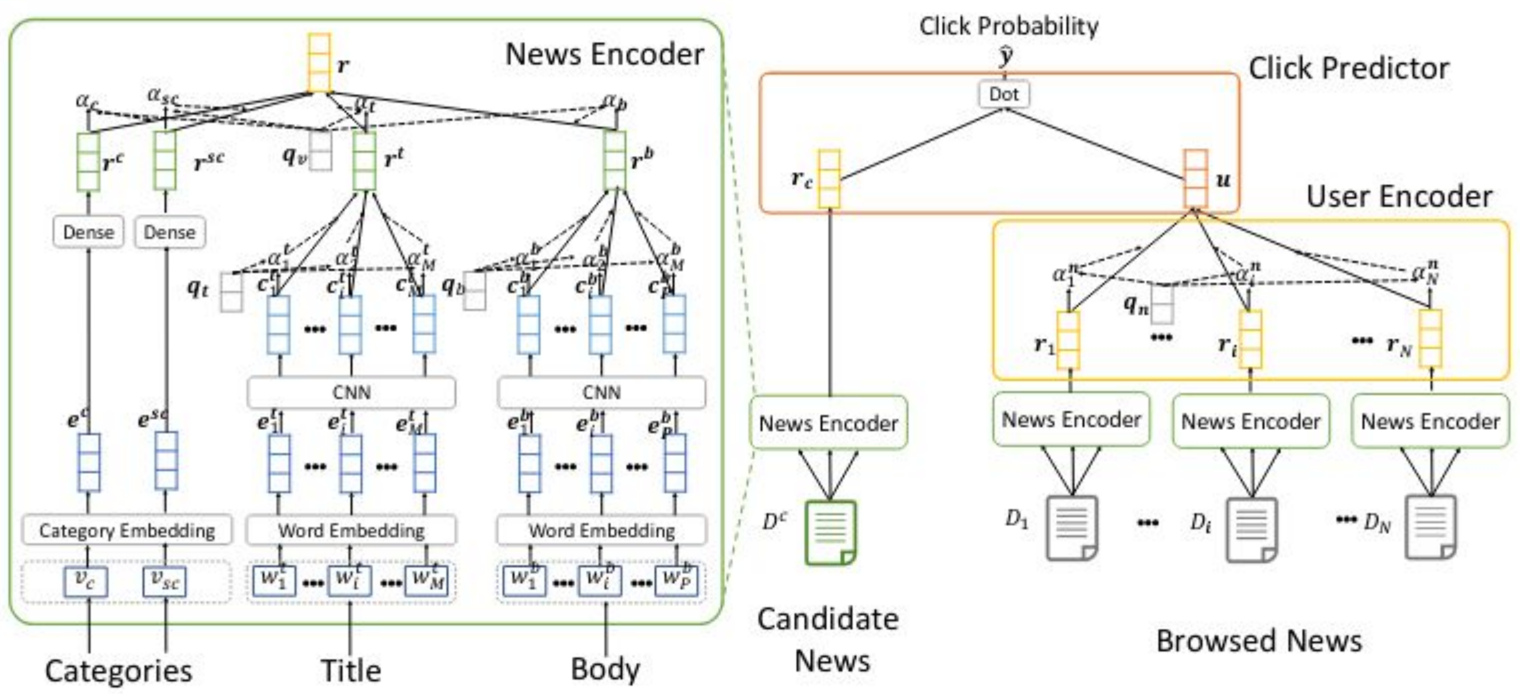
where $\alpha_c, \alpha_{sc}, \alpha_t, \alpha_b$ are the attention weights of each view

User Encoder



Learns representations of users from the representations of their browsed news

Attentive Multi-view learning framework



Training

- Negative sampling technique - with K = 4
- For each news article clicked by the user, randomly sample K articles that are presented in the same session and are not clicked by the user
- Jointly predicting the click probability scores of the positive news article and the k negative news articles.
- Normalize the click-probability scores using softmax
- Loss function : negative log likelihood of all positive samples

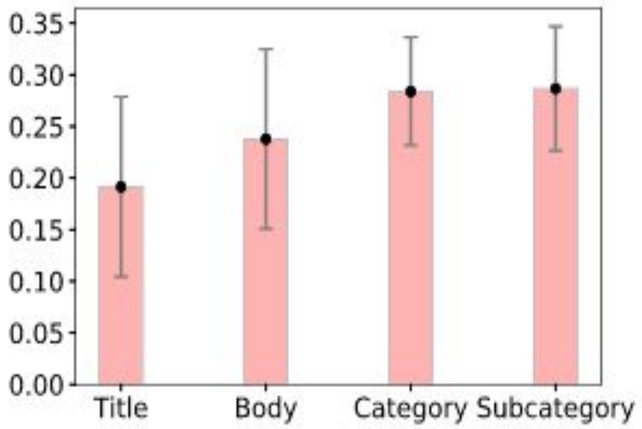
$$L = -\Sigma\log(p_i) \text{ for } i \in S$$ where S is the set of positive training samples

Results

- MRR (Mean reciprocal rank) = 0.6518
- AUC Area under the ROC curve = 0.3072
- nDCG @ 5 = 0.3380
- nDCG @ 10 = 0.4022

Results

Plot of the mean and standard deviation of attention weights of the views



Observations

- The category view has the highest attention weight among all views for most samples.
- Attention weights on the title and body views are small for many samples.
- The over-fitting has reduced upon applying 20% drop-out to each layer.

Conclusion

Incorporating information from multiple views and applying attention mechanism has learnt useful representations thereby improving the performance of the news recommendation task.

References

[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.

Thank You

Any Questions?