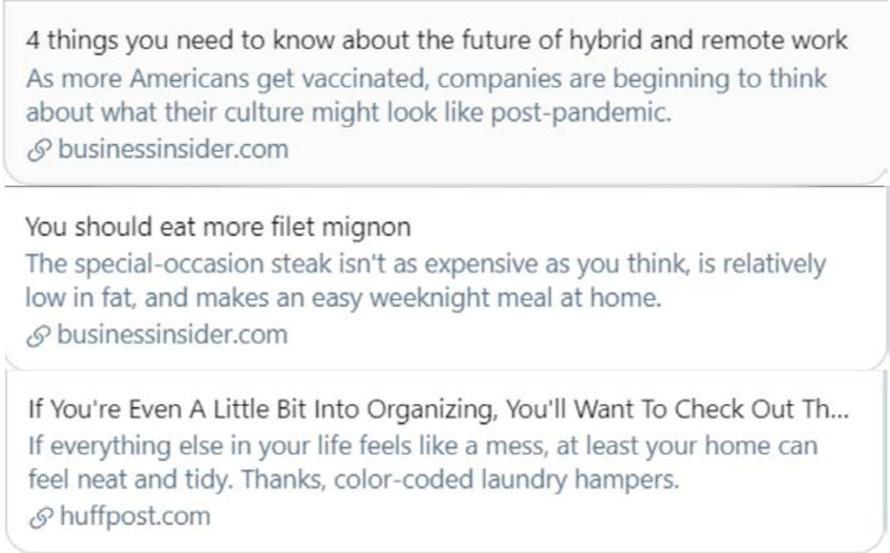
Click-BERT: Clickbait Detector with Bidirectional Encoder Representations from Transformers

EECS 498-004 Final Project

Changyuan Qiu, Chenshu Zhu, Haoxuan Shan

{peterqiu, jupiterz, shanhx}@umich.edu

What is Clickbait? & Why we want to detect it?



4 things you need to know about the future of hybrid and remote work

As more Americans get vaccinated, companies are beginning to think about what their culture might look like post-pandemic.

businessinsider.com

You should eat more filet mignon

The special-occasion steak isn't as expensive as you think, is relatively low in fat, and makes an easy weeknight meal at home.

businessinsider.com

If You're Even A Little Bit Into Organizing, You'll Want To Check Out Th...

If everything else in your life feels like a mess, at least your home can feel neat and tidy. Thanks, color-coded laundry hampers.

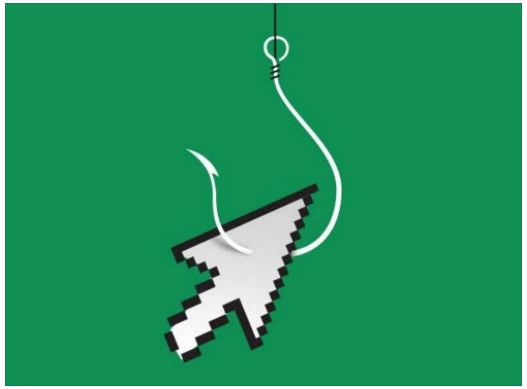
huffpost.com

The $$$ Moneymaking Secret that Banks Don't Want You To Know

Bankfacts

Leading Doctor Reveals the No. 1 Worst Carb You Are Eating

Mediconews



- Time-wasting
- Money-Cost
- Misleading

Related Works

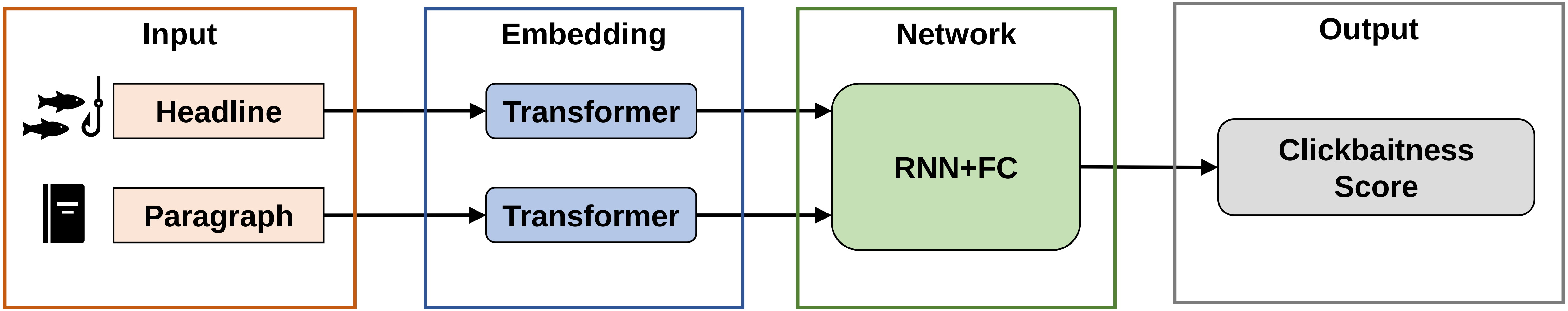
Reference:

- Input Format:
  - ➢ Only Headline (Biyani et al., 2016)
  - ➢ Headline + linked Content (Potthast et al. 2016)
- Representations & Architecture:
  - ➢ Hand-crafted features: (Cao et al., 2017, Elyashar et al., 2017)
  - ➢ Word Embeddings:
    - word2vec + LSTM (Thomas, 2017)
    - GloVe + BiGRU (Omidvar et al., 2018)

Our Contributions:

- Propose a parallel model architecture
- Use advanced pre-trained models like BERT and Longformer (long text)

Model Architecture - Pipeline



Headline: The sentence attract you.
Paragraph: The content of the linked article.
Both are used for better prediction.

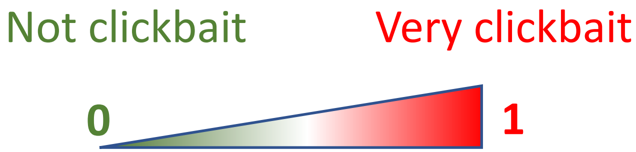
A network to transfer the embedded result into a score.

Challenge: How to precisely embed the (long) text.
Solution: BERT / Longformer

A score in [0,1].
1 - clickbait
0 - non-clickbait

Dataset and Data preparation

Webis-Clickbait-17 Dataset (19538 Tweets)



Not clickbait

Very clickbait

0

1

Tweet id: 858464162594172928

Headline: "UK response to modern slavery leaving victims destitute while abusers go free"

Content: "Thousands of modern slavery victims have, ..., possibility of falling victim again."

Media: "modern-slavery-rex.jpg"

Score*: 0.133

Data Cleaning: discard tweets with **0.3 < score < 0.7**

Low confidence labels only confuse the model

	#tweets	**#clickbaits**	**#non-clickbaits**
Before Cleaning	19538	3133	16405
After Cleaning	12963	2230	10733

We use 0.5 as classification threshold here.
Could be altered in the inference stage to attain a more rigorous or tolerating model.

Note the 1:5 class imbalance. This is nonissue since we expect same imbalance in real life.

*Averaged from 5 human annotators, [0.3, 0.0, 0.3, 0.0, 0.0] in the example above.

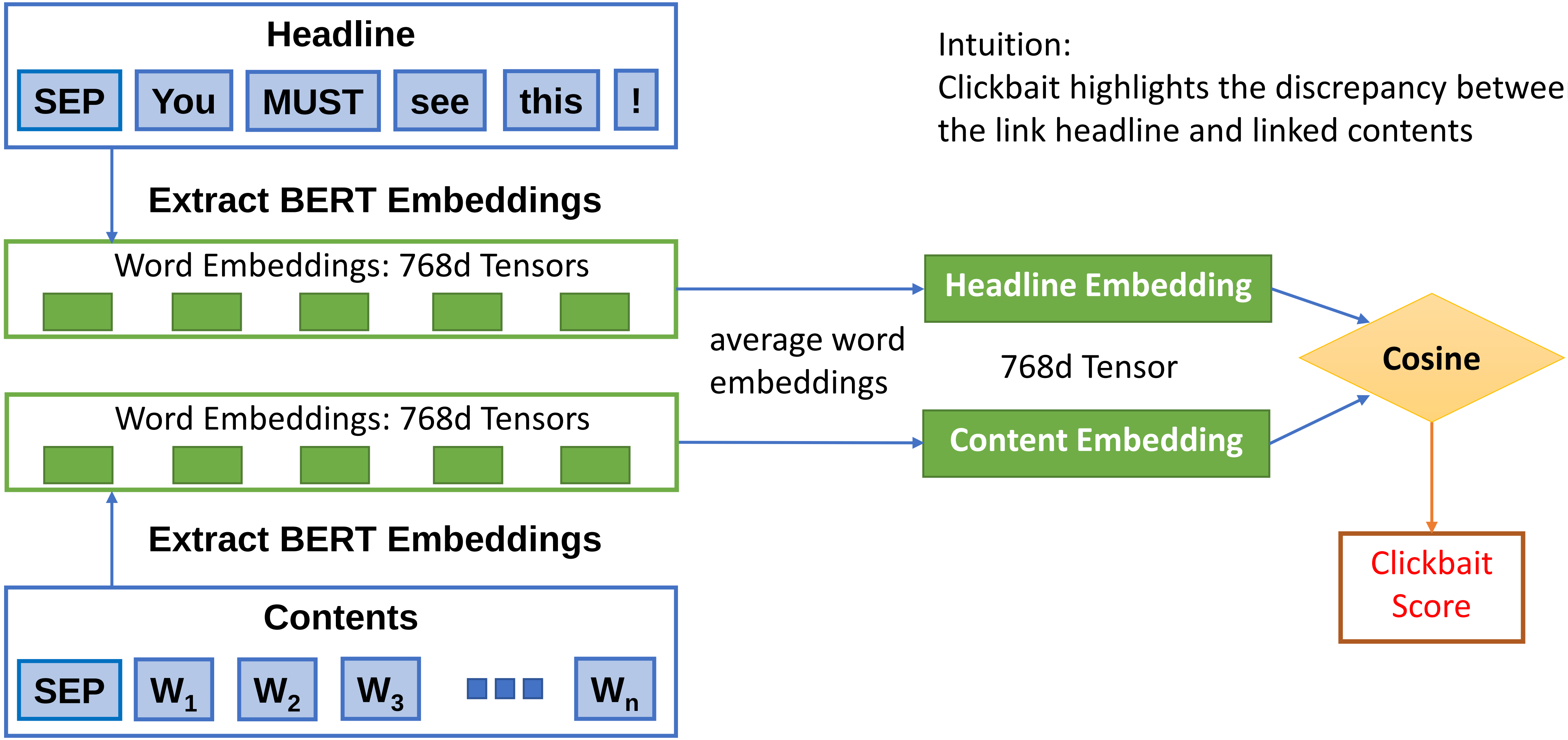
Proposed Models

- **Baseline Model 1:** Naïve Cosine Similarity with BERT Embeddings
  - Use cosine similarity between headline and content for classification
- **Baseline Model 2:** Headline BERT Embeddings with RNN Network
  - Use headline alone with RNN networks for regression
- **Proposed Model:** BERT and Longformer with Parallel Structure
  - Encode headline with pretrained BERT model
  - Encode content (long text) with pretrained Longformer model
  - Pass embeddings from two networks though shared RNN structure

- Detailed model diagrams to follow.

Naïve Cosine Similarity with BERT

Intuition:
Clickbait highlights the discrepancy between the link headline and linked contents



Headline

SEP You MUST see this !

Extract BERT Embeddings

Word Embeddings: 768d Tensors

average word embeddings

Headline Embedding

768d Tensor

Content Embedding

Cosine

Clickbait Score

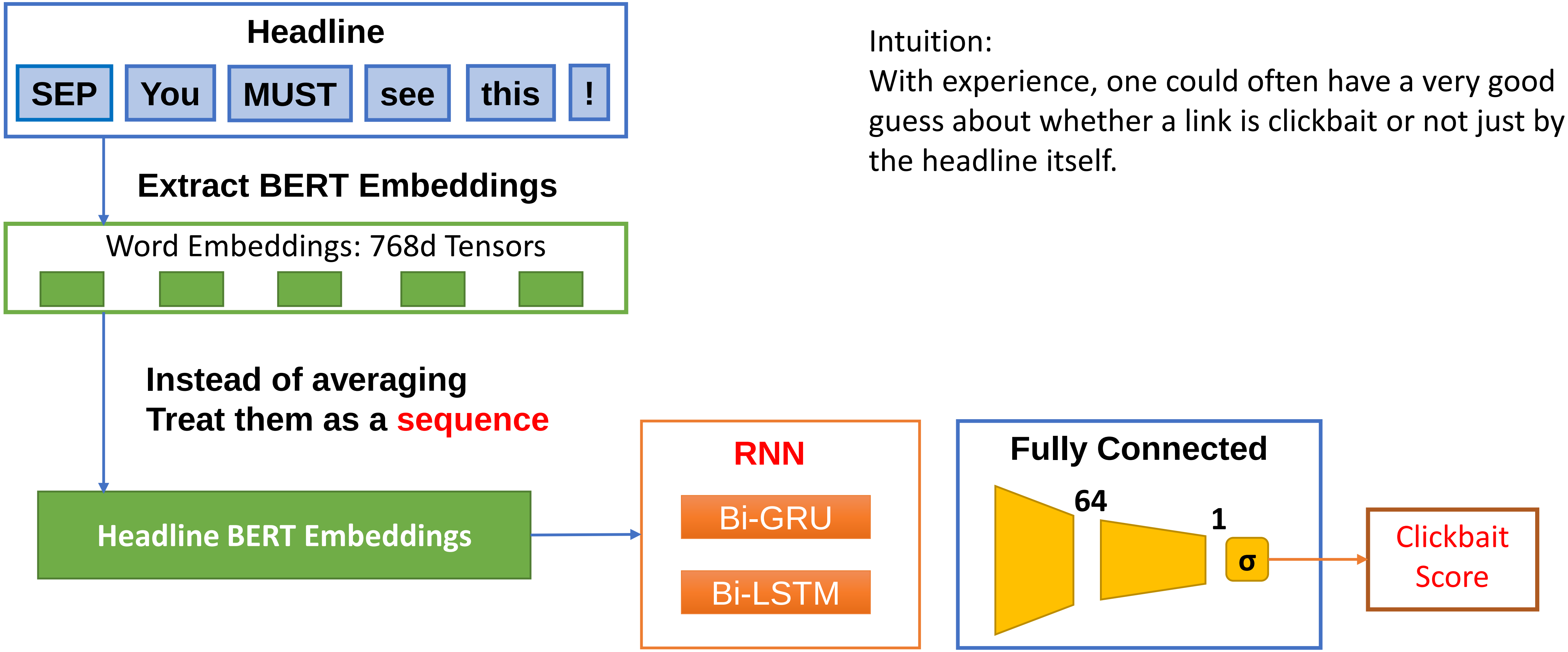
Word Embeddings: 768d Tensors

Extract BERT Embeddings

Contents

SEP W_1 W_2 W_3 W_n

Headline BERT with RNN Network



Headline

SEP You MUST see this !

Extract BERT Embeddings

Word Embeddings: 768d Tensors

Instead of averaging
Treat them as a **sequence**

Headline BERT Embeddings

RNN

Bi-GRU

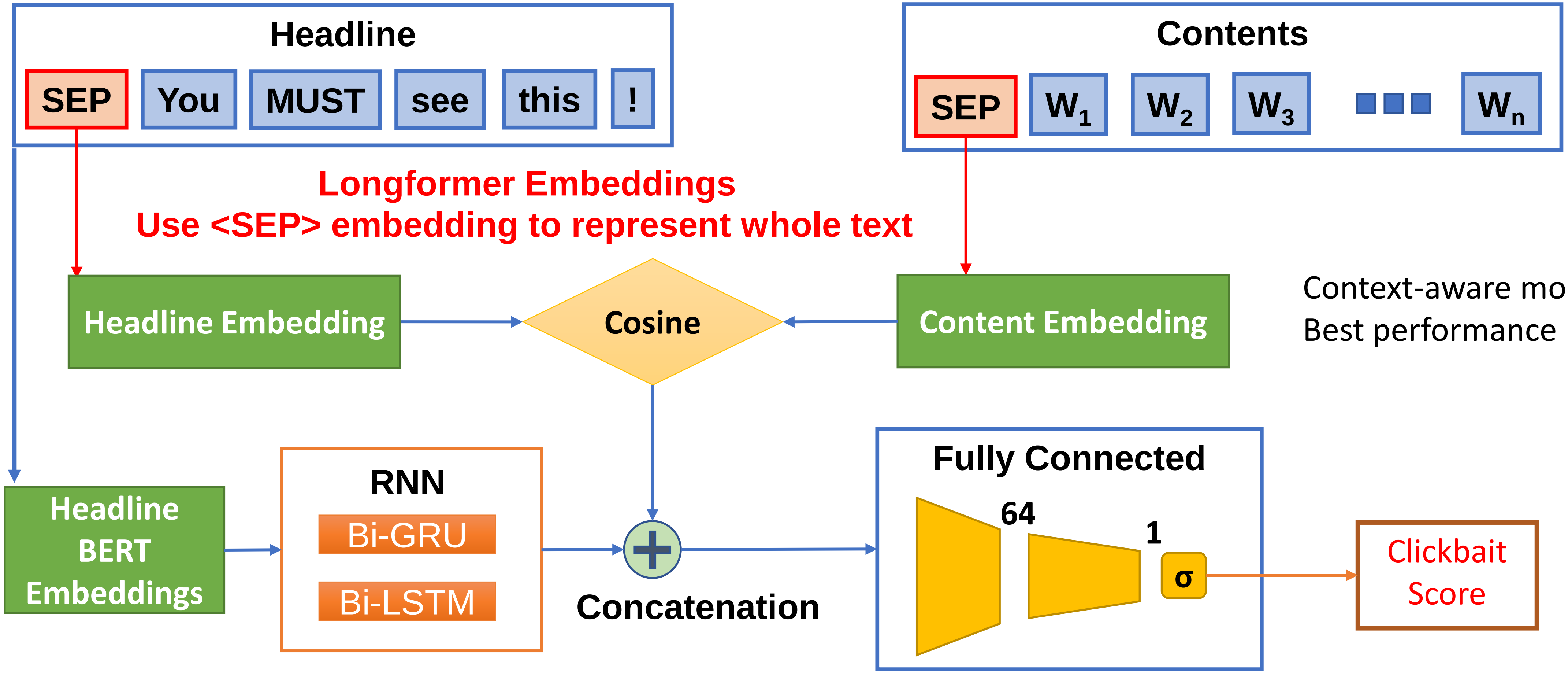
Bi-LSTM

Fully Connected

Clickbait Score

Intuition:
With experience, one could often have a very good guess about whether a link is clickbait or not just by the headline itself.

BERT + Longformer with Parallel Structure



Headline

SEP You MUST see this !

Contents

SEP W_1 W_2 W_3 ■■■ W_n

Longformer Embeddings

Use <SEP> embedding to represent whole text

Headline Embedding

Cosine

Content Embedding

Headline BERT Embeddings

RNN

Bi-GRU

Bi-LSTM

Concatenation

Fully Connected

64

1

σ

Clickbait Score

Context-aware model

Best performance

Experiments - Training/Evaluation Details

- **Train/Validation Split:**

	#tweets	#clickbaits	#non-clickbaits
Training	11663	2027	9636
Validation	1300	203	1097
Total	12963	2230	10733

- **Evaluation Metrics:**
  - Benchmark with previous models *on the Clickbait 2017 Challenge for Webis-17 dataset*
  - Regression: Mean Square Error (MSE)
  - Classification: Accuracy & F1-Score

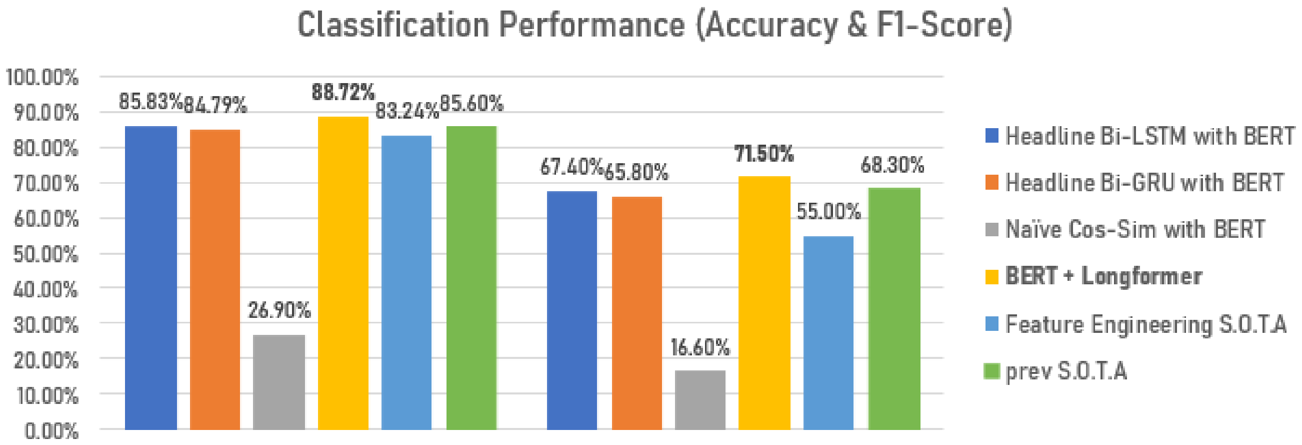
Experiments - Training/Evaluation Details

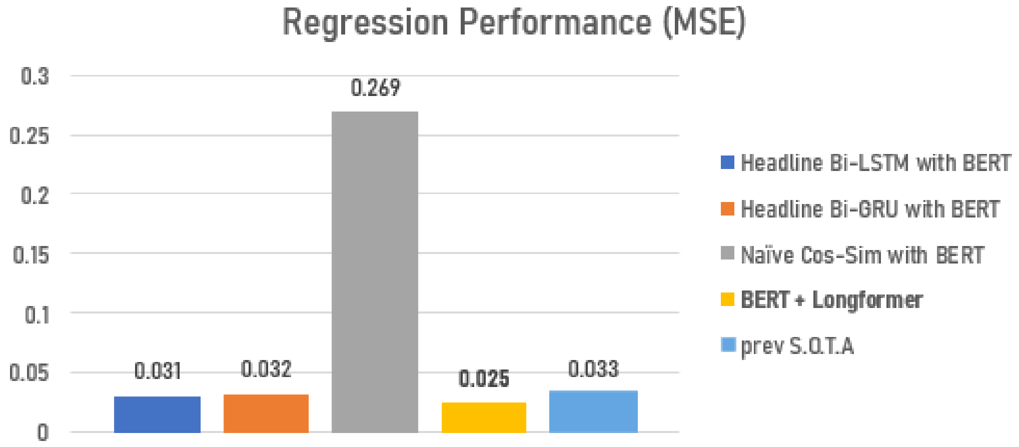
- Hyperparameters for Bi-LSTM / Bi-GRU:
  - Hidden dimension: 50
  - # of layers: 2
  - Dropout: 0.2
- Hyperparameters for FC layer:
  - Hidden dimension: 64
  - Dropout: 0.2
- Loss function: MSE
- Optimizer: Adam
  - Learning rate: 1e-4
  - Adaptive Learning rate (Learning rate scheduler): decay by factor of 0.25 with patience of 2
  - Weight Decay: 1e-3
- Mini-batch size: 8 (limited RAM on Google Colab)
- # Epochs: 20
- Training time: Headline BERT – 0.5 day on T4, BERT + Longformer – 4.5 days on P100

Experiments - Results

Achieve S.O.T.A Performance on Webis-17!

- **Surpass previous S.O.T.A**
  - ~3% w.r.t classification metrics
  - ~25% w.r.t regression MSE
- **Baseline 1 (naïve cos-sim) failed**
  - Average pooling across words failed to represent headline precisely
- **Baseline 2 (only headline) achieve comparable performance with previous S.O.T.A**





Experiments – Case Study

Success case - non-clickbait

Headline: "Tokyo's subway is **shut down** amid fears over an imminent North Korean **missile attack** on Japan"
Content: "One of Tokyo's major subways systems says it **shut down** all lines for 10 minutes after receiving warning of a North Korean **missile launch**. Tokyo Metro official Hiroshi Takizawa says the temporary suspension affected 13,000 passengers this morning. Service was halted on all nine lines at 6:07 am and was resumed at 6:17 am after it was clear there was no threat to Japan. Takizawa said it was the first time service had been stopped in response to a missile launch…. "
Truth Score: 0 (non-clickbait)
Predict Score: 0.074 (non-clickbait)

Success case - clickbait

Headline: "26 pictures guaranteed to make you laugh every time“

Content: "Just trust me. We asked the **BuzzFeed Community** to send us the funniest pictures on the internet. Want to be featured in similar BuzzFeed posts? **Follow the BuzzFeed Community on Facebook and Twitter!** BuzzFeed Home © 2017.…. "

Truth Score: 1.0 (clickbait)
Predict Score: 0.895 (clickbait)

Experiments – Case Study

Failure case - idiom

Headline: "CenturyLinkVoice: **New product launch**: **Testing the waters with social media**"
Content: "Back in the day, companies assembled focus groups to **gather feedback on product prototypes**. Changes were then made based on this group's advice. Today, the same kinds of opinions are being collected **through social media**, making prelaunch research vastly more efficient and cost-effective. Trusted customers who offer their frank opinions often become valuable promoters of a product before and after launch informing their social media followers at no cost to its maker…"
Truth Score: 0.13 (non-clickbait)
Predict Score: 0.674 (clickbait)

Failure case - human label error

Headline: "18 **uplifting documentaries** guaranteed to put a smile on your face"
Content: "The real world isn't all trash. We asked the BuzzFeed Community to tell us their **favourite uplifting documentaries.** Here are the results.
1. Twinsters (2015) "It's an uplifting story of two twins finding each other after they'd been adopted to families in different countries. An easy watch, and so heartwarming!" Watch on: Netflix Worldwide
2. Iris (2014) "I cannot speak highly enough of this film… "
Truth Score: 0.73 (clickbait)
Predict Score: 0.221 (non-clickbait)

Conclusion & Future Work

In this project, we proposed **Click-BERT**: **C**lickbait **D**etector with **B**idirectional **E**ncoder **R**epresentations from **T**ransformers:

- Achieved S.O.T.A performance on the Webis-17 dataset
- Able to distinguish clickbait vs. non-clickbait contents with high accuracy

Future Work:

- Directly fine-tune BERT-like model (compute limits)
- Improve language understanding abilities for media idioms
- Incorporate feature engineering

Reference

[1] Ankesh Anand, Tanmoy Chakraborty, and Noseong Park. We used neural networks to detect clickbaits: You won’t believe what happened next! In Joemon M. Jose, Claudia Hauff, Ismail Seng¨or Alting¨ovde, Dawei Song, Dyaa Albakour, Stuart N. K. Watt, and John Tait, editors, Advances in Information Retrieval - 39th European Conference on IR Research, ECIR 2017, Aberdeen, UK, April 8-13, 2017, Proceedings, volume 10193 of Lecture Notes in Computer Science, pages 541–547, 2017.

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014.

[3] Iz Beltagy, Matthew E. Peters, and Arman Cohan. Longformer: The long-document transformer. CoRR, abs/2004.05150, 2020.

[4] Prakhar Biyani, Kostas Tsioutsiouliklis, and John Blackmer. ”8 amazing secrets for getting more clicks”: Detecting clickbaits in news streams using article informality. In Dale Schuurmans and Michael P. Wellman, editors, Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA, pages 94–100. AAAI Press, 2016.

[5] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, JeffreyWu, ClemensWinter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In Hugo Larochelle, Marc’Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, 2020.

[6] Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. Stop clickbait: Detecting and preventing clickbaits in online news media. CoRR, abs/1610.09786, 2016.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In 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), pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[8] Tom´as Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In Yoshua Bengio and Yann LeCun, editors, 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings, 2013.

[9] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, 2014.

Reference (continued)

[10] Martin Potthast, Tim Gollub, Matthias Hagen, and Benno Stein. The clickbait challenge 2017: Towards a regression model for clickbait strength. CoRR, abs/1812.10847, 2018.

[11] Martin Potthast, Tim Gollub, Kristof Komlossy, Sebastian Schuster, MattiWiegmann, Erika Patricia Garces Fernandez, Matthias Hagen, and Benno Stein. Crowdsourcing a large corpus of clickbait on Twitter. In Proceedings of the 27th International Conference on Computational Linguistics, pages 1498–1507, Santa Fe, New Mexico, USA, Aug. 2018. Association for Computational Linguistics.

[12] Martin Potthast, Sebastian K¨opsel, Benno Stein, and Matthias Hagen. Clickbait detection. In European Conference on Information Retrieval, pages 810–817. Springer, 2016.

[13] Philippe Thomas. Clickbait identification using neural networks. CoRR, abs/1710.08721, 2017.

[14] Matti Wiegmann, Michael V¨olske, Benno Stein, Matthias Hagen, and Martin Potthast. Heuristic feature selection for clickbait detection. CoRR, abs/1802.01191, 2018.

[15] Yiwei Zhou. Clickbait detection in tweets using selfattentive network. CoRR, abs/1710.05364, 2017.

[16] Indurthi, Vijayasaradhi & Syed, Bakhtiyar & Gupta, Manish & Varma, Vasudeva. Predicting Clickbait Strength in Online Social Media. 4835-4846. 10.18653/v1/2020.coling-main.425, 2020.

UNIVERSITY OF
MICHIGAN

Thanks for listening!
Any Questions?