The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)
Emerging track on artificial intelligence for social impact

Detecting Incongruity Between News Headline and Body Text via a Deep Hierarchical Encoder

*Seunghyun Yoon, *Kunwoo Park, Joongbo Shin
Hongjun Lim, Seungpil Won, Meeyoung Cha and Kyomin Jung

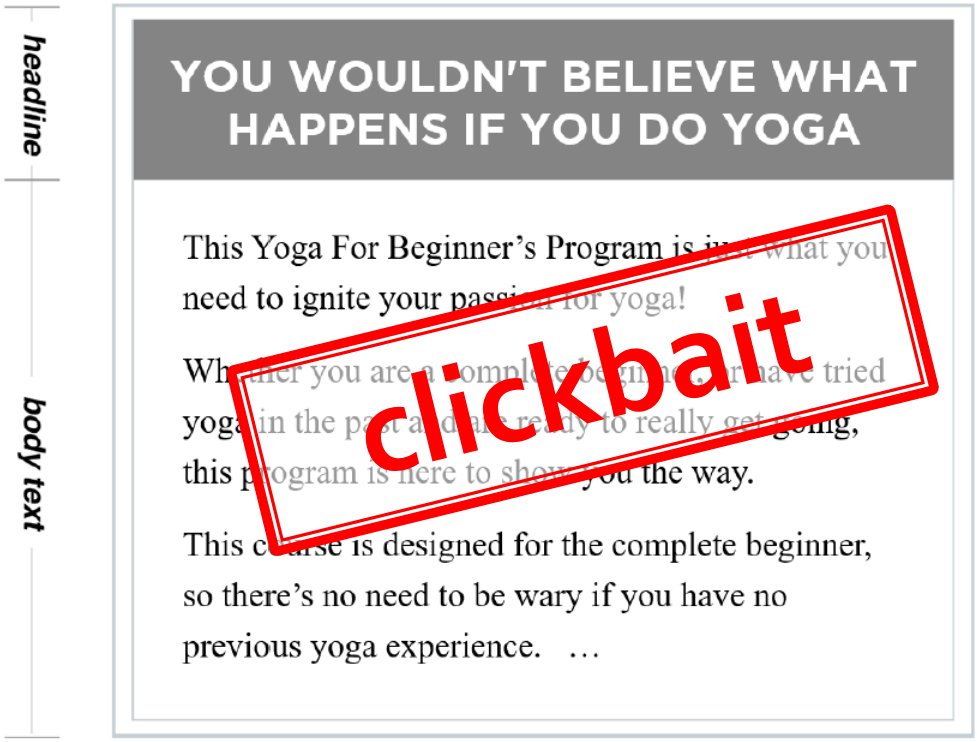
SEOUL NATIONAL UNIVERSITY

Index

- Problem to Solve
- Related Works & Limitations
- Our Contributions
  - New Dataset for Research
  - Proposed Approaches
  - Evaluation Results in the Real World
- Conclusion

Research Problem

- **Detect incongruity between news headline and body text**
  (i.e., when a news headline does not correctly represent the story as in advertisements, clickbaits, fake news, hijacked stories, etc.)



Why Is This an Important Problem?

- News headlines are known to play an important role in making first impressions to readers, and thereby deciding the viral potential of news stories within social networks[1]

- People are less likely to read or click on the whole contents but just read news headlines[2]

much of news sharing is headline based

[1] Reis et at. 2015. Breaking the news: First impressions matter on online news. In Proceedings of the ICWSM.
[2] Gabielkov et al. 2016. Social clicks: What and who gets read on twitter? ACM SIGMETRICS Performance Evaluation Review

Why Is This an Important Problem?

An Initial impression gained from the headline is persistent such that its stance remains even after reading the whole news content[3]

[3] Ecker et al. 2014. The effects of subtle misinformation in news headlines. Journal of experimental psychology: applied

What Has Been Done so Far?

challenge

The **Fake News Challenge** (FNC-1) 2017

- Estimate the stance of a news article
- 50K pairs of headline and body text
- From 1,683 original news articles

http://www.fakenewschallenge.org/

What Has Been Done so Far?

challenge

The **AI R&D Challenge** 2017

- Total Prize : 10M USD
- Detecting fake news
- Training set is not provided

http://airndchallenge.com

What Has Been Done so Far?

academy

Wei et al. 2017, **Learning to identify ambiguous and misleading news headlines**, IJCAI

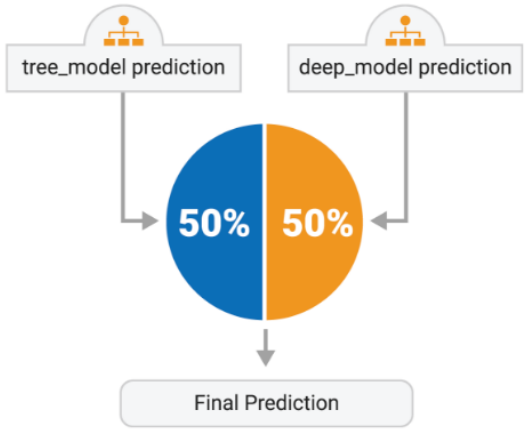
- Detecting ambiguous headlines from the pair of title and body text
- 40,000 articles, only 2,924 articles are annotated
- Feature based

Feature	Description
Wordcnt	Count the number of words
Number	Count the number of numerals
Baitword	Count the number of clickbait words
Slang	Count the number of internet slang
Punctuation	Count the number of !, ? and ...
SentDegree	Respectively count 2 sets of degree adverbs expressing "很" (very) and "非常" (extremely)
SentPolar	Respectively count words expressing positive evaluation, negative evaluation, positive emotion, and negative emotion
Distance	Compute the average distance between governing and dependent words (identified by LTP parser[5])
WHword	Count the number of Chinese interrogative pronouns
ForwardRef	Count words expressing forward-reference, including demonstratives (this,that,...) and personal pronouns (he,she,it,...)

Limitations in Previous Research

A sophisticated model could not be proposed or fully-evaluated due to lack of available corpus for research

- Simple model won the competition
- Non-realistic data (FNC-1)
  - 50,000 pairs of headline and body text from 1,683 articles
  - (Avg) 29.7 samples have an identical headline



FNC-1 Winning model

Our Contributions

① Generate dataset for research

- 4 M news articles[1,2] : Jan 2016 ~ Oct. 2017 (Korean)
- 100 K news articles (English)

② Propose new deep learning models & method

- Attentive Hierarchical Dual Encoder (AHDE)
- Independent Paragraph (IP) Method

③ Evaluate the models in the real world

- With newly crawled 232,261 news articles[2]
- Find 250 incongruent articles with 0.82 precision

[1]http://www.naver.com, [2]http://www.daum.net

① Generate Dataset for Research

whole-corpus

Injecting negative sample (paragraph),

into original article to generate inconsistent article

Rule (1) : Sample *N* consecutive paragraphs from an article and insert them into the {body text} of the target article.

Rule (2) : Sample *N* non-consecutive paragraphs from one article and insert them randomly into the {body text} of the target article.

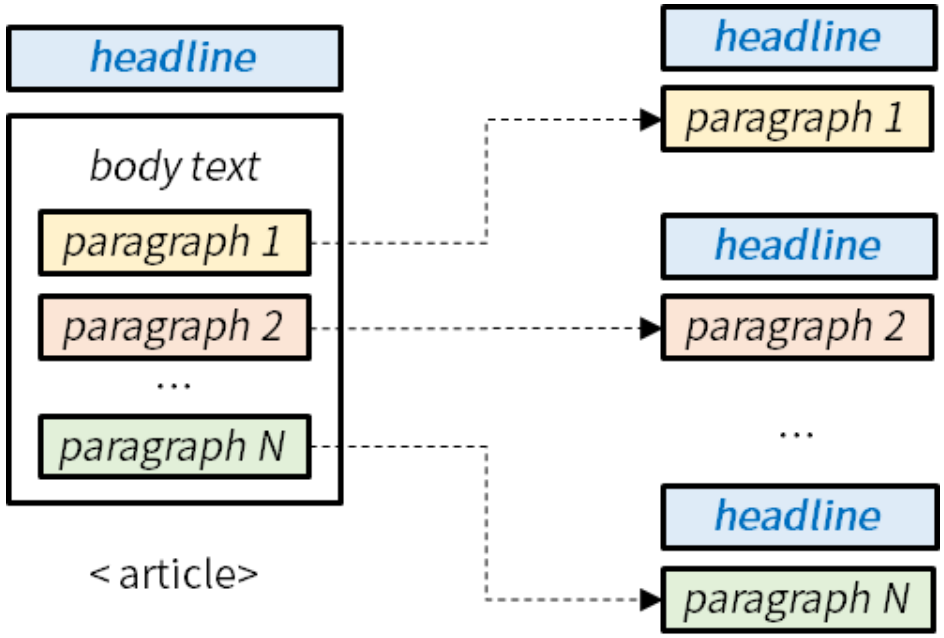
※ We hired human annotators to manually read 1,000 randomly sampled articles from the created dataset and check whether their headlines are incongruent with the article content.

① Generate Dataset for Research

paragraph-corpus

Transform pair of {headline} and {body text}

into **multiple sub-pairs** of {headline} and {paragraph}



① Generate Dataset for Research

paragraph-corpus

The length of sequence that a model handles

The number of training instances

Dataset	# Samples			Headline (Avg.)			Body Text (Avg.)		
	Train	Dev	Test	# tokens	# chunk	# tokens /chunk	# tokens	# chunk	# tokens /chunk
whole	1.70M	100,000	100,000	13.71	1	13.71	518.97	8.37	62.00
paragraph	14.20M	834,064	100,000	13.71	1	13.71	62.00	2.03	30.05

Table 1: Properties of the dataset. The chunk in the body text implies paragraphs and sentences for the whole and the paragraph dataset, respectively.

② Proposed Approaches

- **Attentive Hierarchical Dual Encoder (AHDE)**

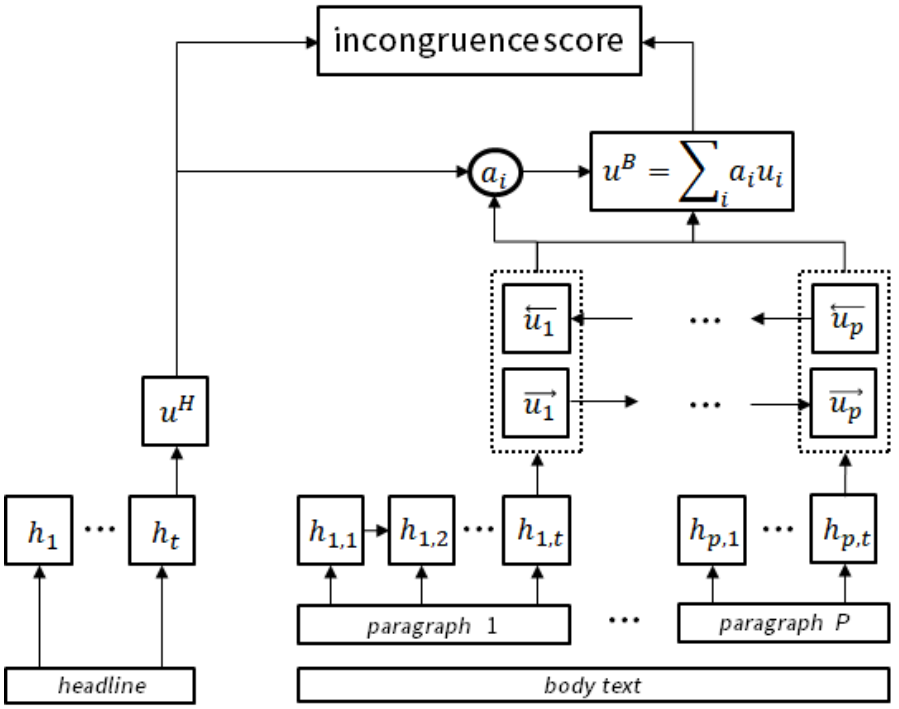
Deep hierarchical models encode the full news article
from a word-level to a paragraph-level with attention mechanism

Passage encoding

$$h_{p,t} = GRU(h_{p,t-1}, w_{p,t})$$
$$u_p = GRU(u_{p-1}, h_p)$$

Attention

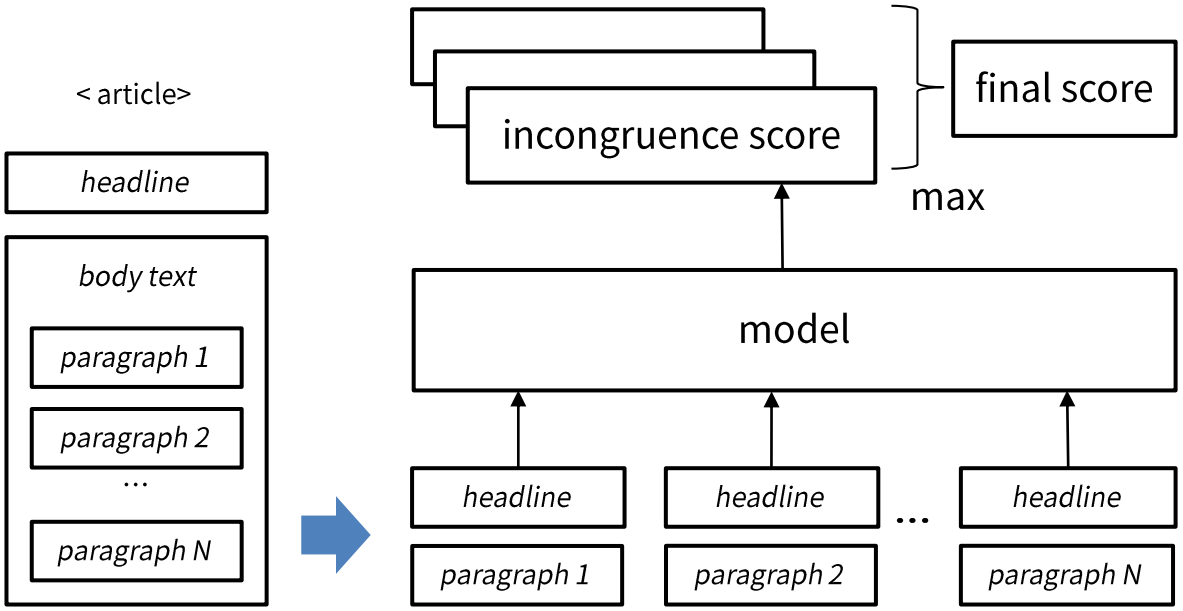
$$s_p = v^T \tanh(W_u^B u_p^B + W_u^H u_p^H),$$
$$a_i = \frac{e^{s_i}}{\sum_p e^{s_p}},$$
$$u^B = \sum_p a_i u_i^B$$



② Proposed Approaches

- **Independent Paragraph (IP) Method**

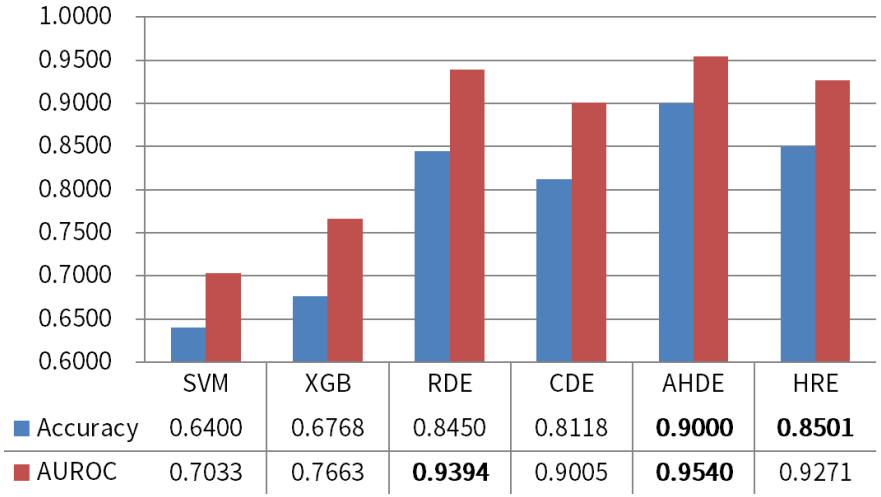
Splits paragraphs in the {body text} and learns the relationship between each paragraph and headline independently



② Proposed Approaches

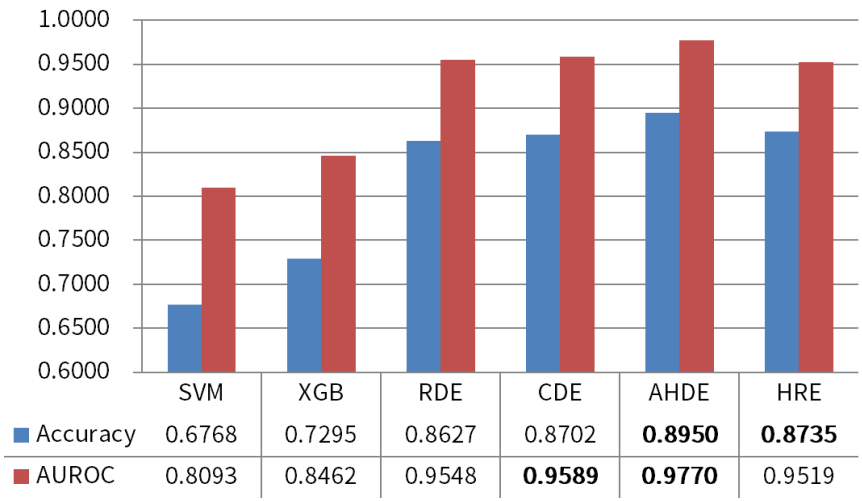
- **Performance Comparison**

The newly proposed AHDE achieved the best performance
IP method provides additional performance gain



	SVM	XGB	RDE	CDE	AHDE	HRE
Accuracy	0.6400	0.6768	0.8450	0.8118	**0.9000**	**0.8501**
AUROC	0.7033	0.7663	**0.9394**	0.9005	**0.9540**	0.9271

without IP method
(whole corpus)



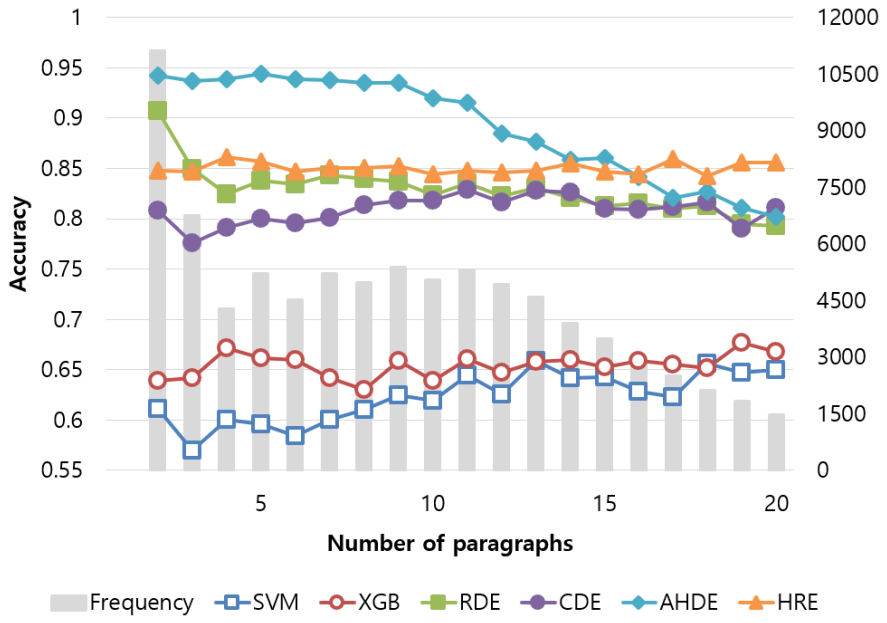
	SVM	XGB	RDE	CDE	AHDE	HRE
Accuracy	0.6768	0.7295	0.8627	0.8702	**0.8950**	**0.8735**
AUROC	0.8093	0.8462	0.9548	**0.9589**	**0.9770**	0.9519

with **IP** method
(paragraph corpus)

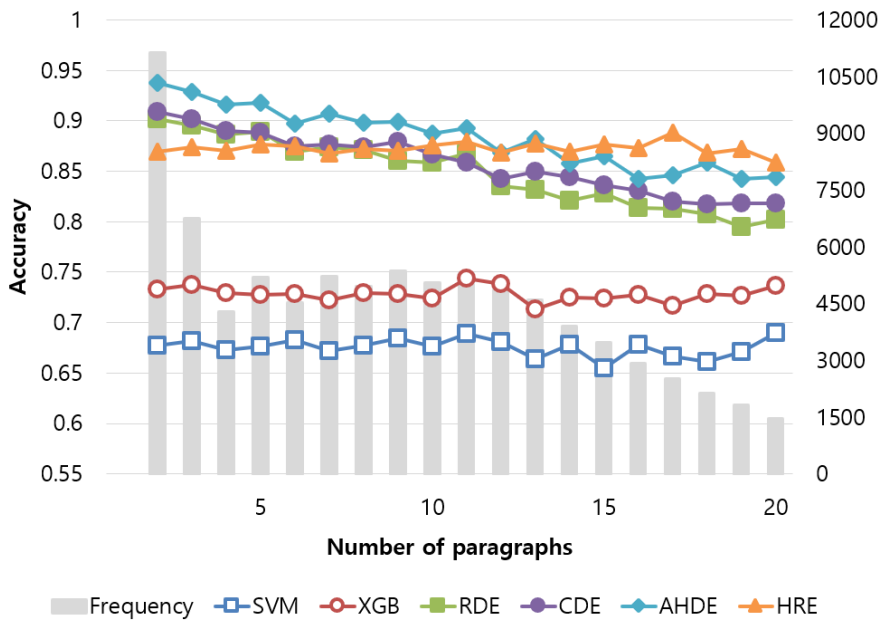
② Proposed Approaches

- **Performance over Long Text Input**

Robustness of the proposed models in handling long sequential input



without **IP** method



with **IP** method

③ Evaluation in the Real World

Evaluation in the Wild (1)

Crawl, recently released news articles (232,261)

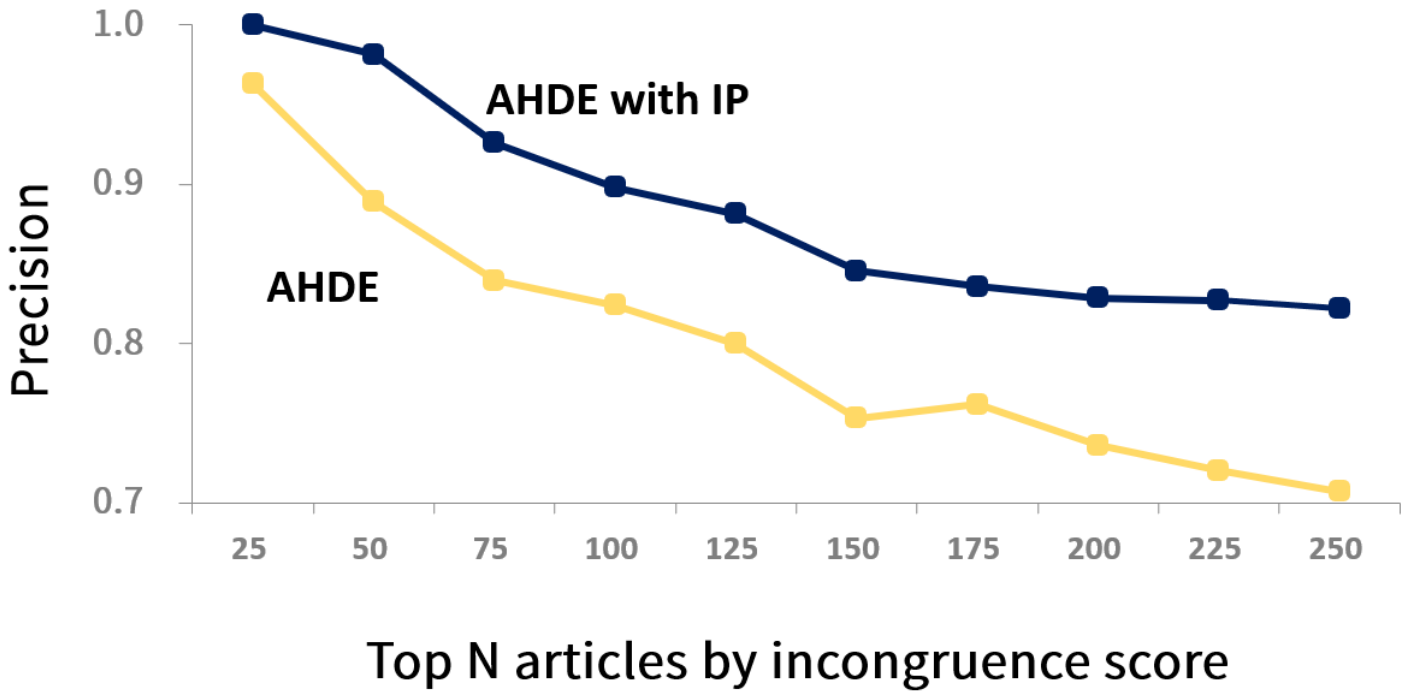
January to April of 2018

Validate top *N* articles by incongruence scores

that are given by model prediction

③ Evaluation in the Real World

The proposed models effectively detect the incongruent articles
among newly gathered 232,261 news articles



③ Evaluation in the Real World

Evaluation in the Wild (2)

Gather news articles from two media outlets

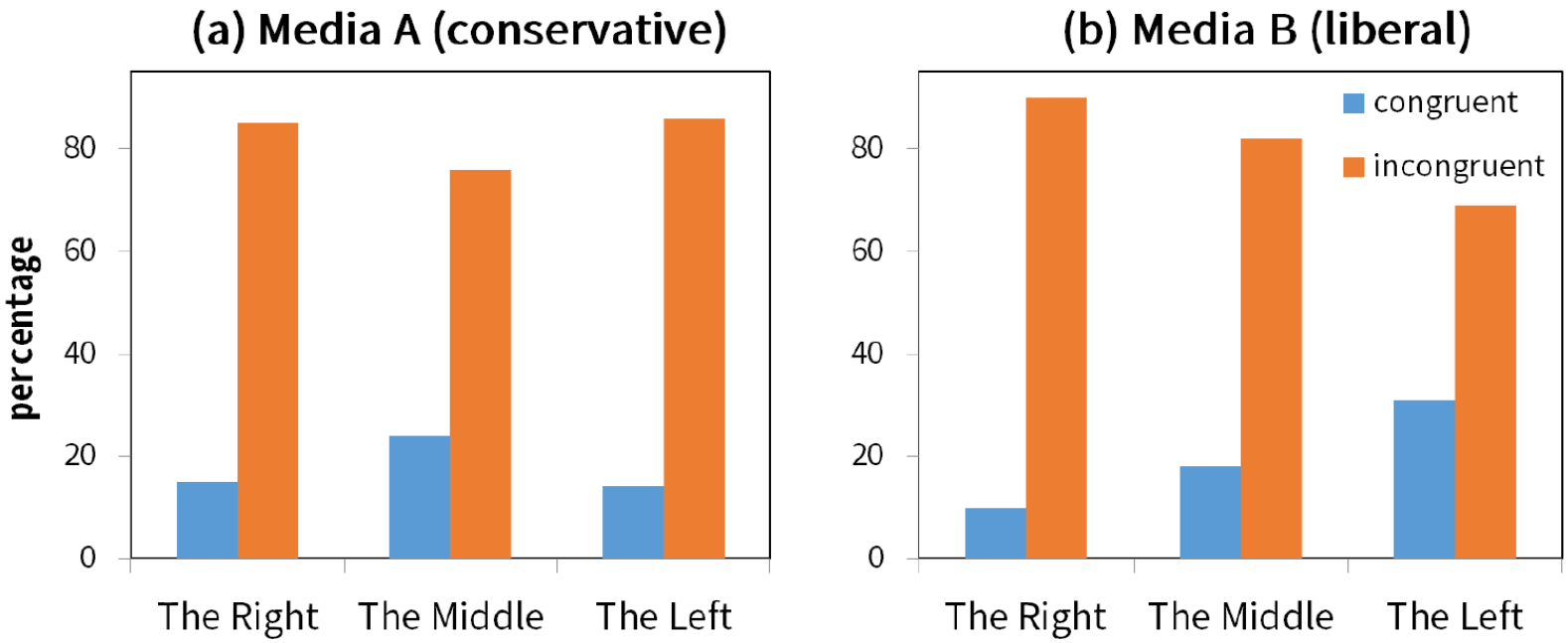
(**conservative** / **liberal** media)

Ask to 100 Amazon Mechanical Turk workers

"Do you think the headline of the above article is incongruent with its body text?"

③ Evaluation in the Real World

- **Varying perceptions on headline incongruence**



③ Evaluation in the Real World

Evaluation in the Wild (3)

Conducted an additional experiment on the

FNC-1 dataset (fnc 2017),

to have a similar setting with our task,

"unrelated", *"agree"*, *"disagree"*, and *"discuss"*

"unrelated" and *"others"*

③ Evaluation in the Real World

- **Hierarchical Encoders for Stance Detection**

Model	Accuracy
SVM	0.7501
XGB	**0.9279**
CDE	0.756
RDE	0.8228
AHDE	0.8444
HRE	0.8088
XGB+CDE (The winning model)	0.9304
XGB+RDE	0.9407
XGB+AHDE	**0.9433**
XGB+HRE	0.9368

Performance of models that are trained on the FNC-1 training set and evaluated on the separated test set.

Conclusion

We study the problem of incongruent headline detection

- RELEASE a **million-scale** data corpus for research
- PROPOSE **two neural networks** that efficiently learn the textual relationship between headline and body text
- SHOW that the models trained on our released corpus show decent performances on the **real-world dataset**

Thank you

We'd like to invite you to the poster session for further discussion (Jan 29)

code, data, paper → http://david-yoon.github.io