Stay on Topic, Please: Aligning User Comments to the Content of a News Article

Jumanah Alshehri[1*] **Marija Stanojevic**[1*] **Eduard Dragut**[1] **Zoran Obradovic**[1]

{jumanah.alshehri, marija.stanojevic, edragut, zoran.obradovic}@temple.edu

[1] Center for Data Analytics and Biomedical Informatics
Temple University, Philadelphia, PA, USA

43rd European Conference on Information Retrieval (ECIR 2021)

* Contributed equally

Motivation, Goals, and Challenges

Motivation:

- Virtual discussions offer an insight into the public opinion
- 20% - 50% of users comments are irrelevant to the news article [1,2]
- This noise in the data affects downstream applications such as opinion mining

Goals:

- Introduce the Article-Comment Alignment Problem (ACAP)
- Define a set of article-comment relevance classes and propose a methodology to classify article-comments pairs automatically

[1] He, L., Shen, C., Mukherjee, A., Vucetic, S., Dragut, E.: Cannot predict comment volume of a news article before (a few) users read it. In: ICWSM. AAAI Press (2021)
[2] Singer, J.B.: Separate spaces: Discourse about the 2007 scottish elections on a national newspaper website. The International Journal of Press/Politics 14(4), 477-496 (2009).

Motivation, Goals, and Challenges

Challenges:

- Comments are informal
- Comments are short
- Some comments are hard to categorize

"It's not Europe anymore. It's Eurabia. This should not be a news story anymore."

* Full article: https://dailym.ai/2Qz7RG9

Hypothesis

- It is possible to capture the extent of a connection and semantics between an article and its comments using globally pre-trained models, jointly fine-tuned with local data
- Considering the natural order of labels (relevant, same entities, same category, and irrelevant) during training will boost the algorithm learning process
- Constraints:
  - Limited amount of labeled data
  - Bounded number of tokens
  - Finding relevance by comparing formally and informally written text

Data

Dataset:

- We collect News articles and their comments between 2015 and 2017 [3] from multiple news outlets
- Dataset has over 19K articles and 9M comments
- We choose five outlets with a broad range of lengths and different number of articles and comments

Outlet	(A) Dataset Statistics			
	#Art.	#Comm.	ALA	ALC
FN	0.3K	72K	250	22
TG	1.6K	428K	797	54
MW	1.7K	65K	512	42
WSJ	3.6K	309K	164	57
DM	10K	1,012K	487	28

[3] He, L., Han, C., Mukherjee, A., Obradovic, Z., Dragut, E.: On the dynamics of user engagement in news comment media. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 10(1) (2020).

Data

Labeling:

- We randomly select 1K article-comment pairs from each outlet
- Three annotators manually and independently label the article-comment pairs in four classes: **Relevant**, **Same Entities**, **Same Category**, and **Irrelevant**
- The final label is assigned using an average aggregation schema
- We created a binary version of each dataset

Outlet	(B) Classes proportion			
	Relevant	Same Ent.	Same Cat.	Irrelevant
FN	3%	21%	29%	47%
TG	5%	39%	32%	24%
MW	7%	51%	20%	22%
WSJ	8%	25%	34%	33%
DM	15%	17%	20%	48%

Data

Labeling:

First Part of the article	Comment	Class
The *Clinton Campaign* at *Obama Justice* Emails on *WikiLeaks* show a top federal lawyer giving *Hillary* a quiet heads up. *President Obama* and *Attorney General Loretta Lynchat* the *White House* in July. The most obnoxious pin of the 2016 campaign cames this week, as *Democrats*, their media allies and even *President Obama* accused the *FBI* of stacking the election. It's an extraordinary claim, coming as it does from the same that has -we now know – been stacking the crew election all along in the corridors of the *Justice Department*. This is the true November surprise. For four months, *FBI Director James Comey* has been the public face of the investigation into *Hillary Clinton*'s email server. He played that role so well, putting the *FBI* so front and center, that the country forgot about *Mr.Comey*'s bosses.	As a practicing lawyer this is just embarrassing. Lawyers are governed by a *Code of Professional Conduct*. The *Justice Department* has enormous power in our country. Politics is not supposed to be part of the equation. It is now abundantly clear that politics is now game on for the *Obama Justice Department*.	Relevant
	The news article that never was written is how *Obama* has corrupted the government agencies and how *Clinton* will continue the process? *Kim* has done the best job of placing the blame for the political corruption of these agencies? The *IRS State Dept. Justice Dept.* and compass because he has perpetuated this for political gain	Same Entity
	We allowed *binladen* family to fly out during 911 blackout as soon I read that in the news I swore never to vote *Bush* again.	Same Category
	I always look liked *Joe Friday*	Not Relevant

Data

User Agreement Study:

Dataset	WSJ	TG	DM	MW	FN
Fleiss Kappa	0.22	0.36	0.37	0.40	0.45
Krippendorff's α	0.42	0.60	0.61	0.64	0.66

- **WSJ was the hardest dataset to label**
- **WSJ, TG, DM, and MW = Fair Agreement**
- **FN = Moderate Agreement**

$$FK = \frac{\sum_{i=1}^{N}\sum_{j=1}^{k} v_{ij}^2 - Nm}{Nm(m-1)} \quad [4]$$

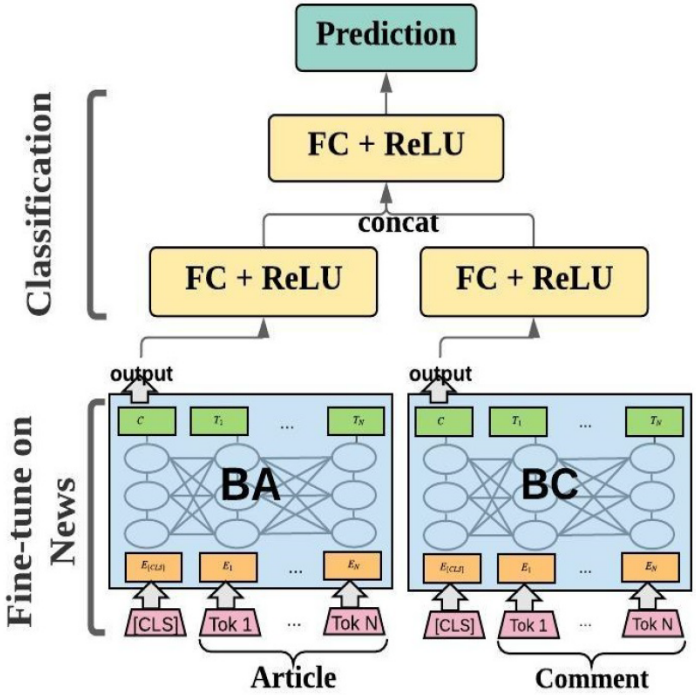
$$\alpha = 1 - \frac{(n-1)\sum_i \sum_j o_{ij} \times \delta_{ij}^2}{\sum_i \sum_j v_i \times v_j \times \delta_{ij}^2} \quad [5]$$

[4] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. \textit {In Psychological Bulletin}, 76(5), page 378
[5] Klaus Krippendorff. 2011. Computing krippendorff's alpha-reliability. \textit{ In Scholarly Commons}.

Methodology

- **BERTAC Model** - Joint Modeling of Article and Comments
- **BA-BC Model** - Disjoint Modeling of Article and Comments
- **Ordinal Classification Loss**

$$weight = 1 + \frac{|\bar{y}_i - y_i|}{k - 1}$$



BA-BC Model

Experimental Setup

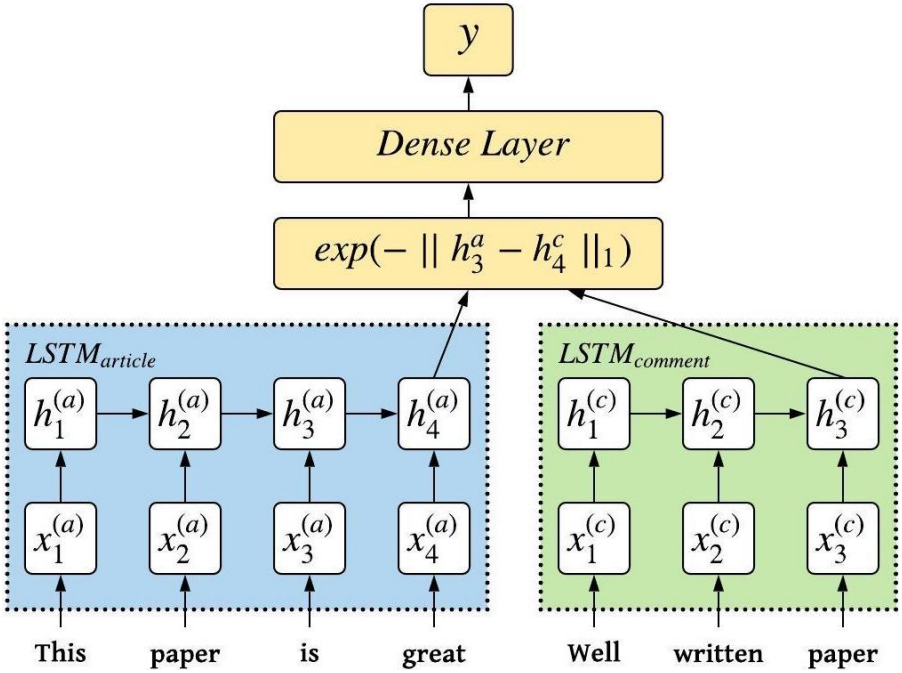
Evaluation:

- **Weighted Accuracy :**

$$WACC = \frac{\sum_{i=1}^{m} |\widetilde{s}_i - \bar{s}_i|}{mD}$$

Baselines:

- **Doc2Vec**
- **Siamese LSTM**



Siamese LSTM

Results and Discussion

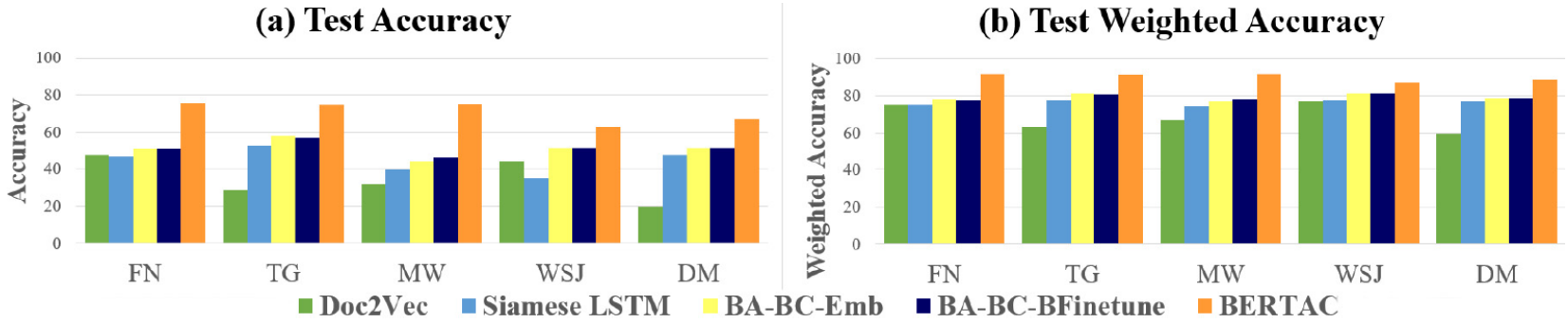
Binary versus Multiclass ACAP:

Model	Dataset	FN	TG	MW	WSJ	DM
BERTAC	B	88.30(1.42)	92.45(1.45)	88.64(2.50)	85.07(0.97)	90.46(1.75)
	M	75.60(1.81)	74.58(6.49)	75.26(4.52)	63.17(2.44)	67.36(3.46)

- Maximal performance is around 92% in the Binary setting
- Accuracy drops between 13%-23% in the Multiclass setting
- It is harder for the model to capture the semantics and knowledge in the Multiclass setting.

Results and Discussion

Models Comparisons on Multiclass ACAP:

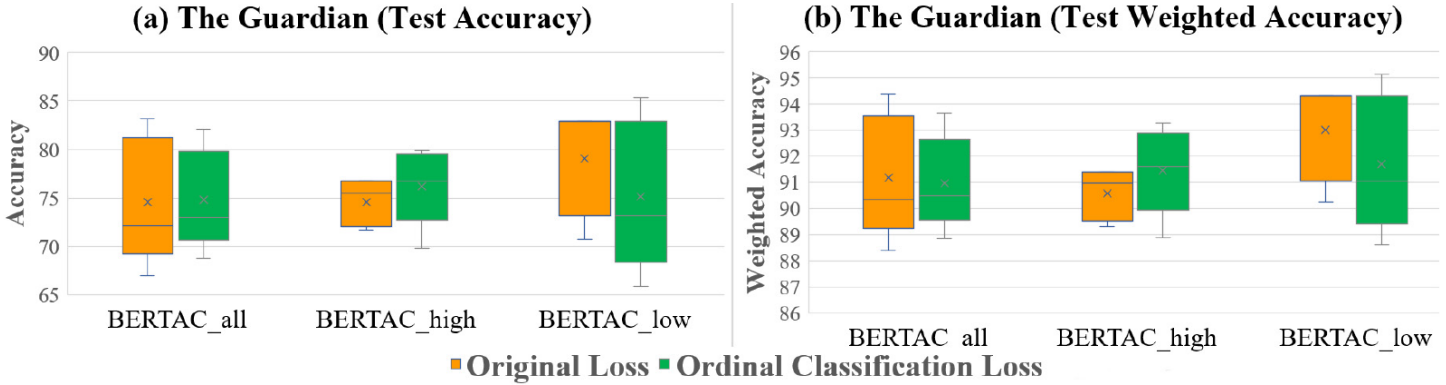


- **Doc2Vec** performance is the worst
- **Siamese LSTM** accuracy is 1%- 27% better
- **BA-BC** accuracy is 4%- 17% higher
- **BERTAC** outperforms all models

- The model learns better when trained on the same article with different types of comments

Results and Discussion

Ordinal Classification Loss:



- **Ordinal loss** has no significant advantage
- Investigate this phenomenon using **high agreement** examples and **low agreement** examples

- **BERTAC high** with ordinal loss is outperforming original loss
- **Quality** affects ordinal loss more than **quantity**

Results and Discussion

Ordinal Classification Loss:

Model	FN	TG	MW	WSJ	DM
BERTAC$_{ord}$	75.08(4.19)	74.78(5.15)	71.08(3.47)	64.45(3.36)	68.42(1.49)
BERTAC$_{vote}$	**76.73 (2.15)**	**76.00 (6.16)**	**74.00 (3.40)**	**64.00 (2.77)**	**69.02 (1.87)**

- What if the model was capable of finding the best prediction from different models?
- Vote = Average vote prediction for BERTAC uncased trained with ordinal loss and original loss, and BERTAC cased trained using ordinal loss.

- The voting system improves the results concerning the accuracy and standard division

Conclusion and Future Work

- Introduced article-comment alignment problem (ACAP)
- Propose an effective approach to predict the level of relatedness between a comment and an article
- Joint modeling of article-comments, i.e., BERTAC, can capture a deeper level of semantic relatedness between news articles and their comments
- With the ordinal loss, we can identify common mistakes made by annotators; use them to improve the performance of downstream applications, which we will explore in the future

Thank you!

Questions?