Clustering Document Parts: Detecting and Characterizing Influence Campaigns from Documents

Zhengxiang Wang, Owen Rambow
{zhengxiang.wang, owen.rambow}@stonybrook.edu
Department of Linguistics, Institute for Advanced Computational Science
Stony Brook University, Stony Brook, NY, USA

PROBLEM DEFINITION

- **Influence campaign**: A *coordinated* and *strategic* effort to *influence* the views of the target audience on certain matters of interest to the influencers.
- As such, it *cannot possibly be inferred from any single document in isolation*!

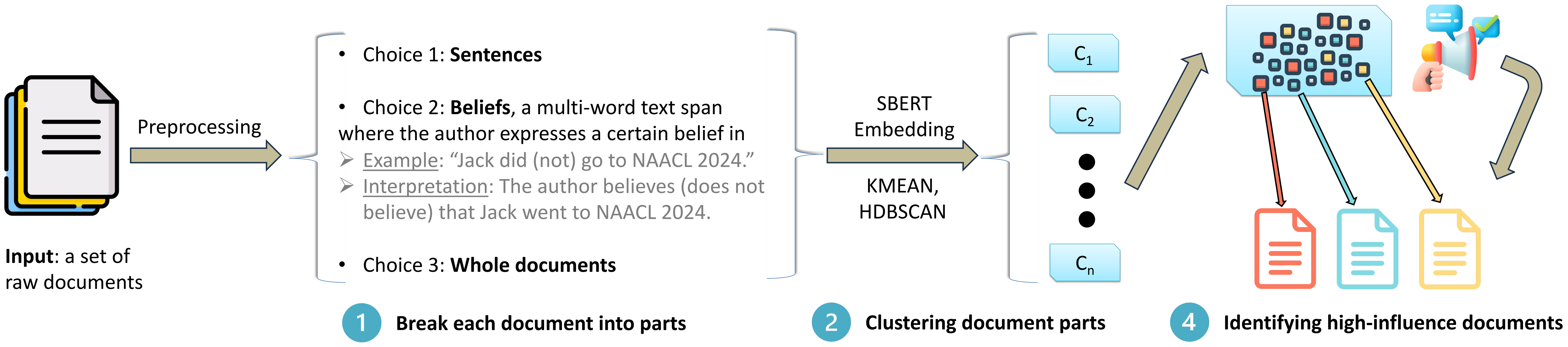
- **Detecting an influence campaign has two aspects**:
  - Classification → if a document reflects an influence campaign
  - Clustering → grouping a cluster of documents reflecting an influence campaign
- **Problems**: (1) classification: biased to keywords; (2) clustering: hard to evaluate

- **Assumption**: influence campaigns → spreading a shared belief/theme of the influencers in a highly organized and thus consistent way
- **Our approach**: cluster document parts → detecting influence clusters/documents

Media	Positive	Negative
Twitter	...Putin **cleans** up the bioweapons labs installed by the deep state... (44 tks)	...RT @EmmanuelMacron France strongly **condemns** Russia's decision to wage war on Ukraine... (19 tks)
Forum	...a secret NATO **laboratory** for biological weapons...Biological weapons tests were **carried** out in the laboratories of this facility... (638 tks)	...[NATO] has **blocked** Ukraine's plan to enter...Item 3: Ukraine **was** a pawn that the Westerners deliberately sacrificed to strengthen NATO... (703 tks)
News	...a NATO secret biological **laboratory** with biological weapons...The biological laboratory under the Azovstal plant in Marioupol in the so-called PIT-404 facility was **built**...In the laboratories of the facility, tests were **carried** out to create biological weapons... (1497 tks)	...Russia's **demand** for neutrality...But NATO members **said** that Ukraine's membership was at best a distant option... [The leader of the Ukrainian separatist region of Lugansk said he could hold a referendum on integration into Russia,] a **decision** immediately criticized by Kiev...(1152 tks)

Positive: documents reflecting an influence campaign. Negative: otherwise.

PIPELINE



DATA

	Train	Test
# Docs	5334	1333
	(416; 7.8%)	(56; 4.2%)
# Sents	72,330	14,370
	(15,394; 21.3%)	(2,182; 15.2%)
# Targets$_{\mathrm{ALL}}$	270,818	50,781
	(61,652; 22.8%)	(8,531; 16.8%)
# Targets$_{\mathrm{AT}}$	155,238	29,793
	(34,703; 22.4%)	(4,905; 16.5%)

Statistics of the train and test sets

- From a DARPA INCAS project
- **Annotated on document collection level**: if a collection of documents contain an influence campaign: US biolabs in Ukraine for bioweapons
- **Six genres**: Twitter, Forum, News, Blog, Reddit, and Other, from Jan 31 to June 30, 2022

RESULTS

- ✓ **Our clustering approach outperforms the direct-document approach**
- ✓ **Clustering document parts outperforms clustering whole documents**
- ✓ **Cluster aggregation helps in virtually all cases (except document-level + FNN)**
- ✓ **Clustering with beliefs can be useful, but clustering sentences + XGBoost + Aggregation achieves best results**

	FNN			XGBoost		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Direct-document	$20.2_{\pm 2.2}$	$18.9_{\pm 14.9}$	$17.1_{\pm 8.6}$	$77.3_{\pm 9.3}$	$37.9_{\pm 7.9}$	$50.7_{\pm 9.1}$
Document-level (mean)	$0.3_{\pm 0.1}$	$0.7_{\pm 0.1}$	$0.4_{\pm 0.1}$	$90.7_{\pm 2.5}$	$25.4_{\pm 3.3}$	$38.2_{\pm 4.6}$
+ Aggregation	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$94.1_{\pm 0.7}$	$28.6_{\pm 3.1}$	$43.8_{\pm 3.8}$
Sentence-level (mean)	$28.3_{\pm 4.1}$	$44.1_{\pm 4.7}$	$32.8_{\pm 4.2}$	$69.4_{\pm 10.9}$	$50.4_{\pm 2.7}$	$56.7_{\pm 4.1}$
+ Aggregation	$74.5_{\pm 16.4}$	$43.2_{\pm 4.1}$	$54.3_{\pm 7.7}$	$86.5_{\pm 1.8}$	$70.7_{\pm 2.4}$	$77.8_{\pm 2.0}$
Target$_{\mathrm{ALL}}$-level (mean)	$25.4_{\pm 6.7}$	$35.2_{\pm 8.5}$	$27.0_{\pm 6.9}$	$78.2_{\pm 3.7}$	$73.8_{\pm 2.4}$	$75.3_{\pm 1.3}$
+ Aggregation	$72.5_{\pm 4.5}$	$40.0_{\pm 5.7}$	$51.5_{\pm 5.7}$	$81.1_{\pm 3.5}$	$71.1_{\pm 7.3}$	$75.5_{\pm 3.8}$
Target$_{\mathrm{AT}}$-level (mean)	$60.7_{\pm 7.1}$	$66.8_{\pm 10.5}$	$62.4_{\pm 8.5}$	$63.5_{\pm 2.2}$	$49.5_{\pm 2.8}$	$54.8_{\pm 2.1}$
+ Aggregation	$64.8_{\pm 4.6}$	$61.8_{\pm 8.6}$	$63.1_{\pm 6.0}$	$80.2_{\pm 3.5}$	$71.4_{\pm 1.8}$	$75.5_{\pm 0.9}$

EXPERIMENTAL SETUPS

- **Task**: predicting if a document reflects an influence campaign (from the positive document collection) without using lexical features (e.g., word embeddings)

- **Metrics**: precision, recall, F1, given the imbalances of labels
- **Classification algorithms**: (1) Feedforward Neural Network (2) XGBoost
- **Features**: frequency counts of 95 general linguistic features + number of words
- **Cluster features**: 7 features, such as average cosine similarity, cluster size

- **Baselines**: (1) Direct-document: applying the two classification algorithms on documents; (2) Direct-level: applying our clustering pipeline on whole documents

- **Clustering aggregation**: aggregating the clusters from different clustering setups to enhance the classification of both high-influence clusters and documents

- **Each experiment was run for fives times with means + stdev results reported**

FUTURE WORK

- **Datasets**: we are not aware of any other datasets that have document-collection-level annotation on influence campaign to further test our pipeline

- **Incorporating non-textual information**: our clustering pipeline is a **text-only** system. Leveraging non-textual information, such as social interactions may help us create a more complicated system (e.g., graph neural network)

- **Automatic characterization of influence campaigns**: e.g., leveraging LLMs to characterize the themes of the high-influence clusters

SELECTED REFERENCES

- Diego A Martin, Jacob N Shapiro, and Julia G Ilhardt. 2023. Introducing the online political influence efforts dataset. Journal of Peace Research, 60(5):868– 876.
- Ricardo J. G. B. Campello, Davoud Moulavi, Arthur Zimek, and Jörg Sander. 2015. Hierarchical density estimates for data clustering, visualization, and outlier detection. 10(1).
- Nicolas Guenon des Mesnards and Tauhid Zaman. 2018. Detecting influence campaigns in social networks using the ising model.