Deep Headline Generation for Clickbait Detection

Kai Shu[1], Suhang Wang[2], Thai Le[2], Dongwon Lee[2], and Huan Liu[1]

[1]Arizona State University, [2]Penn State University

Clickbait

- Clickbaits are catchy social media posts or sensational headlines that attempt to lure the readers to click

- Clickbaits can have negative societal impacts
  - clickbaits may contain sensational and inaccurate information to mislead readers and spread fake news
  - clickbaits may be used to perform click-jacking attacks by redirecting users to phishing websites

It's important to detect clickbaits!

Clickbait Detection

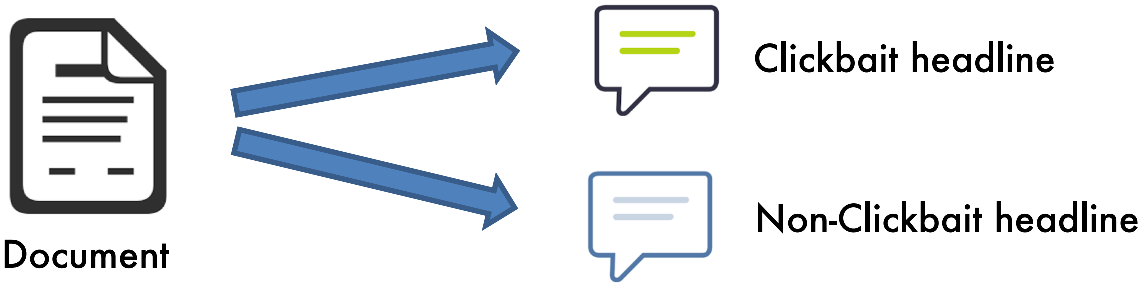
- Existing approaches mainly focus on extracting hand-crafted linguistic features or building complex predictive models such as deep neural networks

- However, these methods may face following limitations
  - Scale: datasets with labels are often limited
  - Distribution: imbalanced distribution of clickbaits and non-clickbaits

We aim to generate synthetic headlines with specific styles and exploit the utility to improve clickbait detection

Headline Generation from Documents

- Goal: Generate stylized headlines that also preserve document contents



  - Stylized headlines can help augment training data for clickbait detection
  - Content preserved headlines make it possible to suggest a non-clickbait headline to readers after we detect a clickbait

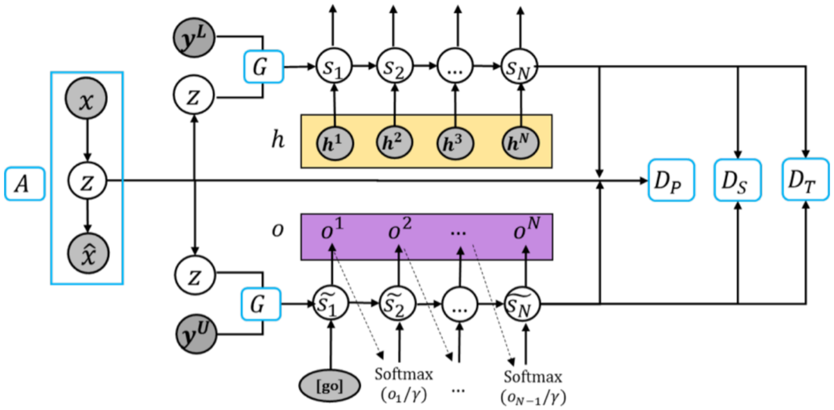
Problem Definition

- Let $\{x_1, x_2, \ldots, x_m\}$, $\{h_1, h, \ldots, h_m\}$, and $\{y_1, y_2, \ldots, y_m\}$ denote the set of m documents, the corresponding headlines, and style labels
- Given $S = \{(x_i, h_i) | i = 1, \ldots, m\}$, learn a generator f that can generate stylized headlines given a document and a style label, i.e., $o_i = f(x_i, y_i)$

- Challenges:
  - How to generate realistic and readable headlines from original document to improve clickbait detection
  - How to generate headlines that can preserve the content of documents and transfer the style of headlines?

Stylized Headline Generation (SHG)

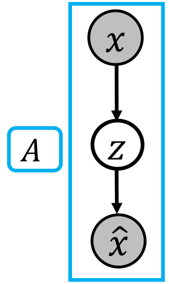
- We propose a deep learning model to generate both click-bait and non-clickbait with style transfer
  - Generator Learning: a document autoencoder $\boldsymbol{A}$, a headline generator $\boldsymbol{G}$
  - Discriminator Learning: a transfer discriminator $\boldsymbol{D_T}$, a style discriminator $\boldsymbol{D_S}$, a pair discriminator $\boldsymbol{D_P}$



Generator Learning

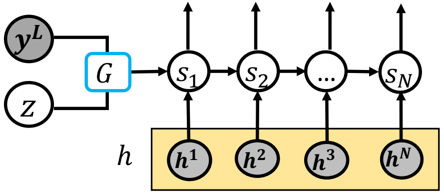
- Document autoencoder ***A***: extract document representation by minimizing the reconstruction error

$$\mathcal{L}_{rec}(\theta_e, \theta_d) = -\sum_{i=1}^{m} \log p(\hat{x}_i | x_i; \theta_d, \theta_e)$$

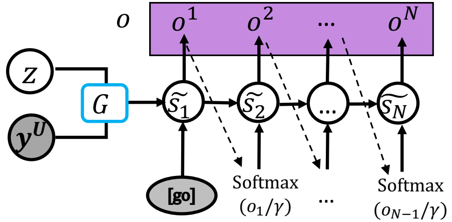


- Headline generator ***G***:
  - generate stylized headline by minimizing the reconstruction error of original headline

$$\mathcal{L}_G(\theta_G) = \mathbb{E}_{(x,h)\in\mathcal{S}}[-\log p_G(h|\mathbf{y}^L, \mathbf{z}))]$$

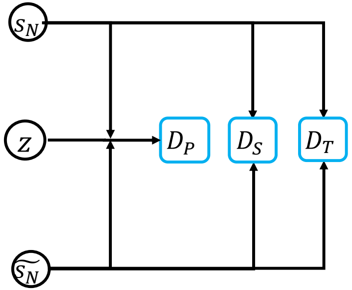


  - generate a set of new headlines o with the styles y^U opposite with original headlines



Discriminator Learning

- Discriminators regularize the representation learning of document z, original headline s_N, and generated headline $\widetilde{s_N}$
- Transfer discriminator $\boldsymbol{D_T}$: discriminate original data samples with generated data samples



Original clickbaits and generated non-clickbaits

$$\mathcal{L}_{D_T} = \boxed{\mathcal{L}_{D_T^{(1)}}(\theta_{D_T^{(1)}})} + \boxed{\mathcal{L}_{D_T^{(2)}}(\theta_{D_T^{(2)}})}$$

Original non-clickbaits and generated clickbaits

- Style discriminator $\boldsymbol{D_S}$: assign a correct label of styles for both original headlines and generated headlines

Original clickbaits and original non-clickbaits

$$\mathcal{L}_{D_S}(\mathbf{W}, \mathbf{b}) = \boxed{\mathcal{L}_{D_S}^{(1)}} + \boxed{\mathcal{L}_{D_S}^{(2)}}$$

Generated clickbaits and generated non-clickbaits

Discriminator Learning cont'd

- Pair discriminator D_p: ensure that the correspondences of documents and headlines are maintained

Proximity function $$p(h_i, x_j) = \frac{1}{1 + \exp(-\underbrace{\mathbf{s}^{(i)}}_{\text{Headline representation}}\mathbf{Q}\underbrace{\mathbf{z}^{(j)}}_{\text{Document representation}})}$$

- Maximizing the proximity of (document, headline) pairs with negative sampling

$$\mathcal{L}_{D_P} = -\log \sigma(\mathbf{s}^{(i)}\mathbf{Q}\mathbf{z}^{(i)}) - \sum_{k=1}^{K} \mathbb{E}_{x_k \sim P_n(x)}[\log \sigma(-\mathbf{s}^{(i)}\mathbf{Q}\mathbf{z}^{(k)})]$$

An optimization framework

- A mini-max game among generators and discriminators

$$\min_{\{\theta_G, \theta_{D_S}, \theta_{D_P}\}} \max_{\theta_{D_T}} \mathcal{L}_G + \alpha \mathcal{L}_{D_S} - \beta \mathcal{L}_{D_T} + \eta \mathcal{L}_{D_P}$$

Algorithm 1 The Learning Process of **SHG**

Input: The two set of training tuple $\mathcal{S}_0, \mathcal{S}_1$, α, β, λ, η, temperature γ, batch size k.

Output: The headline generator G conditioned on $(\mathbf{z}, \mathbf{y}^{\mathbf{U}})$, generated headlines $\mathcal{O}$

Pre-train the document generator A to obtain the content representations $\{\mathbf{z}_i\}_{i=1}^m = \{A(x_i)\}_{i=1}^m$

repeat

  for $l = 0, 1$ **do**

    Sample a mini-batch of k samples $\{(x^i, h^i)\}_{i=1}^k$ from the training tuples $\mathcal{S}_l$

    Unroll G from initial state $(\mathbf{z}_i, \mathbf{y}^L)$ by feeding $\mathbf{s}_i$, and get the last hidden state $\mathbf{s}_N$

    Unroll G from initial state $(\mathbf{z}_i, \mathbf{y}^U)$ by feeding $\tilde{\mathbf{s}}_i$, and get the last hidden state $\tilde{\mathbf{s}}_N$

  end for

  Training the discriminator D_T by gradient descent using Eqn 17

  Training the headline generator G through Eqn 9, and the style discriminator D_S through Eqn 14, and the discriminator D_P via Eqn 19.

until Convergence

Pre-train document autoencoder A

Obtain headline representations of the original headlines and the generated headlines

Train the transfer discriminator

Train the headline generator, style discriminator and pair discriminator

Experiments Setting

- Datasets
  - Professional writers (P): news reporters or editors who come up with clickbaits for the news pieces they publish
  - Social media users (M): clickbaits to lure people to click their posts on social media.

TABLE I: The statistics and descriptions of the datasets

Dataset	Source	# Clickbaits	# Non-clickbaits
P	Professional Writers	5,000	16,933
M	Social Media Users	4,883	16,150

- Baselines
  - SeqGAN [AAAI'17] : text generation using GAN with reinforcement learning
  - SVAE [CONLL'16]: sentence generation using Variational Auto-Encoder (VAE)
  - CrossA [NIPS'17]: generating sentences across different styles

Experiments - Evaluation questions

1. **Consistency**: are generated clickbaits/non-clickbaits consistent with the original datasets?
2. **Readability**, are generated headlines readable or not?
3. **Similarity**, are generated headlines semantically similar to original documents?
4. **Differentiability**: are generated clickbaits/non-clickbaits differentiable?
5. **Accuracy**: can generated clickbaits/non-clickbaits help improve the detection performance?

Data Quality (1–3)

Data Utility (4–5)

Experimental Results - Data Quality

- Consistency: CrossA and SHG have better consistency performance than other baselines.
- Readability: SHG can produce most readable headlines mostly

TABLE III: **EQ1**:The Data Consistency between Generated Headlines $\mathcal{O}$ and original headlines $\mathcal{H}$. Training data is $\mathcal{H}$ and test data is $\mathcal{O}$.

Data	Classifier	SeqGAN	SVAE	CrossA	SHG
P	LogReg	0.501	0.621	0.957	**0.963**
	DTree	0.504	0.603	**0.924**	0.920
	RForest	0.504	0.716	0.924	**0.976**
	XGBoost	0.500	0.722	0.938	**0.977**
	AdaBoost	0.500	0.714	0.957	**0.977**
	SVM	0.503	0.712	0.937	**0.963**
	GradBoost	0.501	0.722	0.933	**0.969**
M	LogReg	0.500	0.576	0.666	**0.763**
	DTree	0.53	0.623	0.687	**0.750**
	RForest	0.501	0.662	0.722	**0.825**
	XGBoost	0.501	0.610	0.660	**0.742**
	AdaBoost	0.500	0.671	0.691	**0.749**
	SVM	0.500	0.620	0.648	**0.725**
	GradBoost	0.501	0.674	0.692	**0.742**

TABLE IV: **EQ2**: The Readability score comparison of the generated headlines $\mathcal{O}$ on different generation methods.

Data	Methods	Clickbait	Non-Clickbait
P	SeqGAN	14.45	14.54
	SVAE	8.64	10.38
	CrossA	8.98	10.48
	SHG	**8.45**	**10.16**
M	SeqGAN	14.48	15.88
	SVAE	9.52	10.79
	CrossA	9.86	10.56
	SHG	**9.36**	**10.04**

Experimental Results - Data Quality

- Similarity: evaluate the semantic similarity of headlines and documents
  - Bilingual Evaluation Understudy (BLEU) score
  - Uni_sim: similarity of universal text embedding
- SHG can achieve better performances to preserve document content than CrossA

TABLE V: **EQ3**: The Average BLEU (BLEU-4) Score Comparison of Generated Headlines. $\mathcal{H}$ indicates original headlines, and $\mathcal{O}$ represents the generated headlines.

Data	Headlines	Methods	Clickbait	Non-Clickbait
P	$\mathcal{H}$		0.555	0.527
	$\mathcal{O}$	CrossA	0.407	0.432
		SHG	**0.453**	**0.446**
M	$\mathcal{H}$		0.541	0.534
	$\mathcal{O}$	CrossA	0.432	0.437
		SHG	**0.451**	**0.442**

TABLE VI: **EQ3**: The Average Uni_sim Value Comparison of Generated Headlines. $\mathcal{H}$ indicates original headlines, and $\mathcal{O}$ represents the generated headlines.

Data	Headlines	Methods	Clickbait	Non-Clickbait
P	$\mathcal{H}$		0.63	0.81
	$\mathcal{O}$	CrossA	0.20	0.22
		SHG	**0.37**	**0.40**
M	$\mathcal{H}$		0.64	0.81
	$\mathcal{O}$	CrossA	0.26	0.34
		SHG	**0.34**	**0.38**

Experimental Results - Data Utility

- Differentiability: we perform clickbait detection on generated datasets using various classifiers
- The synthetic headlines generated by SHG can consistently outperform other baselines for clickbait detection on AUC

TABLE VII: **EQ4**: The prediction performance of generated headlines $\mathcal{O}$ on AUC. The training data is $\mathcal{O}_{train}$ and test data is $\mathcal{O}_{test}$.

Data	Classifier	SeqGAN	SVAE	CrossA	SHG
P	LogReg	0.697	0.710	0.794	**0.864**
	DTree	0.701	0.766	0.791	**0.816**
	RForest	0.685	0.794	0.797	**0.849**
	XGBoost	0.701	0.795	0.795	**0.848**
	AdaBoost	0.694	0.800	0.793	**0.848**
	SVM	0.646	0.795	0.787	**0.847**
	GradBoost	0.702	0.797	0.792	**0.848**
M	LogReg	0.625	0.663	0.744	**0.855**
	DTree	0.612	0.712	0.771	**0.934**
	RForest	0.598	0.724	0.783	**0.883**
	XGBoost	0.616	0.651	0.697	**0.872**
	AdaBoost	0.642	0.667	0.708	**0.872**
	SVM	0.510	0.590	0.693	**0.890**
	GradBoost	0.624	0.654	0.708	**0.885**

Experimental Results - Data Utility

- Accuracy: performance improvement comparison of original headlines on AUC
  - The headlines generated by SVAE, CrossA, and SHG can increase the performance of clickbait detection to some extent
  - SHG can consistently outperform SVAE and CrossA on the performance improvement

Data	Classifier	Org	SeqGAN	SVAE	CrossA	SHG
P	LogReg	0.928	0.900 (↓ 3.02%)	0.933 (↑ 0.54%)	0.932 (↑ 0.64%)	**0.936** (↑ **0.86%**)
	DTree	0.894	0.882 (↓ 1.34%)	0.908 (↑ 1.57%)	0.900 (↑ 0.67%)	**0.910** (↑ **1.79%**)
	RForest	0.900	0.893 (↓ 0.78%)	0.912 (↑ 1.33%)	0.916 (↑ 1.78%)	**0.925** (↑ **2.78%**)
	XGBoost	0.919	0.914 (↓ 0.54%)	0.923 (↑ 0.43%)	0.926 (↑ 0.76%)	**0.928** (↑ **0.98%**)
	AdaBoost	0.917	0.896 (↓ 2.29%)	0.921 (↑ 0.44%)	0.921 (↑ 0.44%)	**0.931** (↑ **1.64%**)
	SVM	0.904	0.898 (↓ 0.66%)	0.917 (↑ 1.44%)	0.920 (↑ 1.77%)	**0.923** (↑ **2.10%**)
	GradBoost	0.921	0.914 (↓ 0.76%)	0.924 (↑ 0.33%)	0.926 (↑ 0.54%)	**0.928** (↑ **0.76%**)
M	LogReg	0.667	0.614 (↓ 7.95%)	0.680 (↑ 1.95%)	**0.685** (↑ **2.70%**)	0.684 (↑ 2.55%)
	DTree	0.618	0.612 (↓ 0.97%)	0.620 (↑ 0.32%)	0.622 (↑ 0.65%)	**0.632** (↑ **2.27%**)
	RForest	0.623	0.610 (↓ 2.09%)	0.634 (↑ 1.77%)	0.627 (↑ 0.64%)	**0.634** (↑ **3.93%**)
	XGBoost	0.655	0.643 (↓ 1.84%)	0.668 (↑ 1.98%)	0.671 (↑ 2.44%)	**0.681** (↑ **3.97%**)
	AdaBoost	0.654	0.639 (↓ 2.29%)	0.664 (↑ 0.65%)	0.671 (↑ 2.60%)	**0.680** (↑ **3.98%**)
	SVM	0.618	0.611 (↓ 1.13%)	0.651 (↑ 5.34%)	0.660 (↑ 6.80%)	**0.681** (↑ **10.19%**)
	GradBoost	0.657	0.643 (↓ 2.13%)	0.667 (↑ 1.52%)	0.671 (↑ 2.13%)	**0.682** (↑ **3.81%**)

Conclusion and Future Work

- We study the problem of generating clickbaits/non-clickbaits from original documents for clickbait detection
- We propose a novel deep generative model with adversarial learning
- Future work
  - Explore the generalization capacity of SHG on other styles such as positive-negative sentiment style and academic-news reporting style
  - Investigate the strategy of learning the disentangled representations of content and style

Data Mining & Machine Learning

Arizona State University

Deep Headline Generation for Clickbait Detection

Kai Shu, et al.

Acknowledgements

This work is supported by the ONR grant N00014-16-1-2257, N00014-17-1-2605, #1422215, #1663343, #1742702, #1820609, and AFRL FA8750-16-C-0108.