Number： 6624

Semantic Distillation and Structural Alignment Network for Fake news Detection

Shangdong Liu, Xiaofan Yue, Fei Wu, Jing Sun, Yujian Feng, Yimu Ji

Nanjing University of Posts and Telecommunications

1. Introduction

- In recent years, the rapid proliferation of multi-modal fake news has posed potential harm across various sectors of society, making the detection of multi-modal fake news crucial.
- Most existing methods can not effectively reduce the redundant information and preserve both semantic and structural information.
- To address these problems, this paper proposes a semantic distillation and structural alignment (SDSA) network. We design an semantic distillation module for modality-specific features to preserve task-relevant semantic information and eliminate redundant information.Then, we propose a triple similarity alignment module to preserve structural information.

2. Method

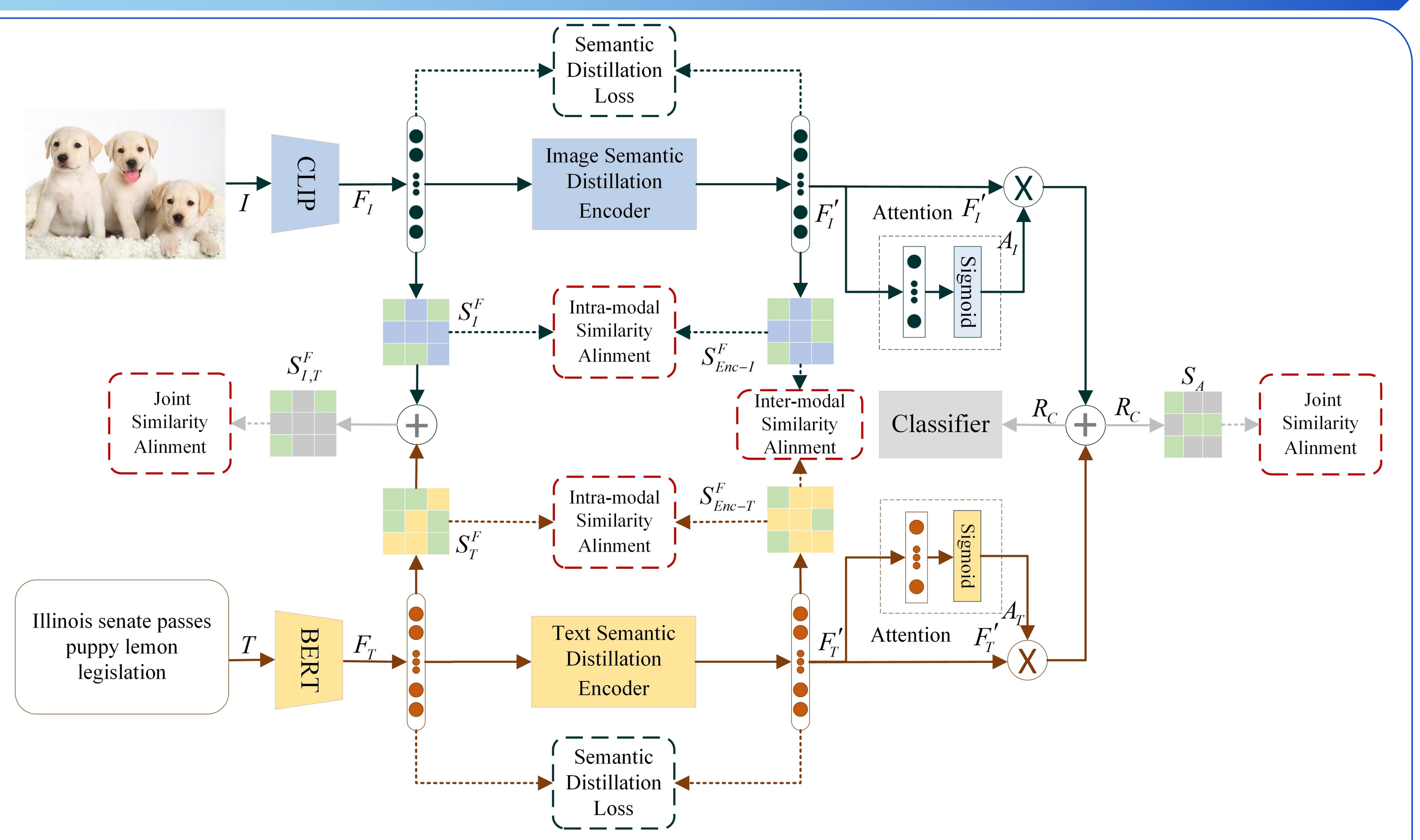


Fig. 1 The overall framework of our SDSA approach

Fig. 1 illustrates the overall architecture of our SDSA approach. SDSA consists of the following two modules: Semantic Distillation Module and Triple Similarity Alignment Module.

➢ *Semantic Distillation Module*

For the fake news detection task, semantic information is composed of two parts: task-relevant semantic information and task-irrelevant semantic information (redundant information). The semantic distillation module is used toto preserve task-relevant semantic information and eliminate redundant information.

We apply the information bottleneck theory to distill semantic information. Specifically, we minimize the difference in mutual information between the features before and after encoding and the labels. Due to the complexity of calculating mutual information, we resort to using KL divergence between the two as a proxy.

➢ *Triple Similarity Alignment Module*

To preserve the neighborhood structure from the original features within each modality. We align S_*^F with S_{Enc-*}^F, and calculate the mean squared error losses between S_*^F and S_{Enc-*}^F By minimizing these intra-modal mean squared error losses, we can retain the intra-modal neighborhood structures as much as possible.

To effectively reduce the differences between modalities, we align S_{Enc-I}^F and S_{Enc-T}^F, calculating the mean squared error loss between S_{Enc-I}^F and S_{Enc-T}^F. By minimizing this inter-modal mean squared error loss, we aim to bring the cross-modal feature neighborhood structures closer together, thus diminishing inter-modal disparities.

We utilize the complementary relationships among inter-modal neighborhood structures to construct a joint similarity matrix $S_{I,T}^F$. We align $S_{I,T}^F$ and S_A to enable fused features to preserve neighborhood structures from various modalities. Due to varying levels of semantic richness among different modalities, we introduce a hyperparameter α to adjust the weighting of different modality in the joint similarity matrix. This facilitates better utilization of neighborhood information from different modalities.

3. Experiment

➢ *Environment*

We conduct comprehensive experiments on two Weibo and Fakeddit datasets, to validate the effectiveness of our approach.The experiments are deployed on an NVIDIA GeForce 1080Ti GPU with PyTorch. The batch size for both datasets is set to 32.

➢ *Baseline*

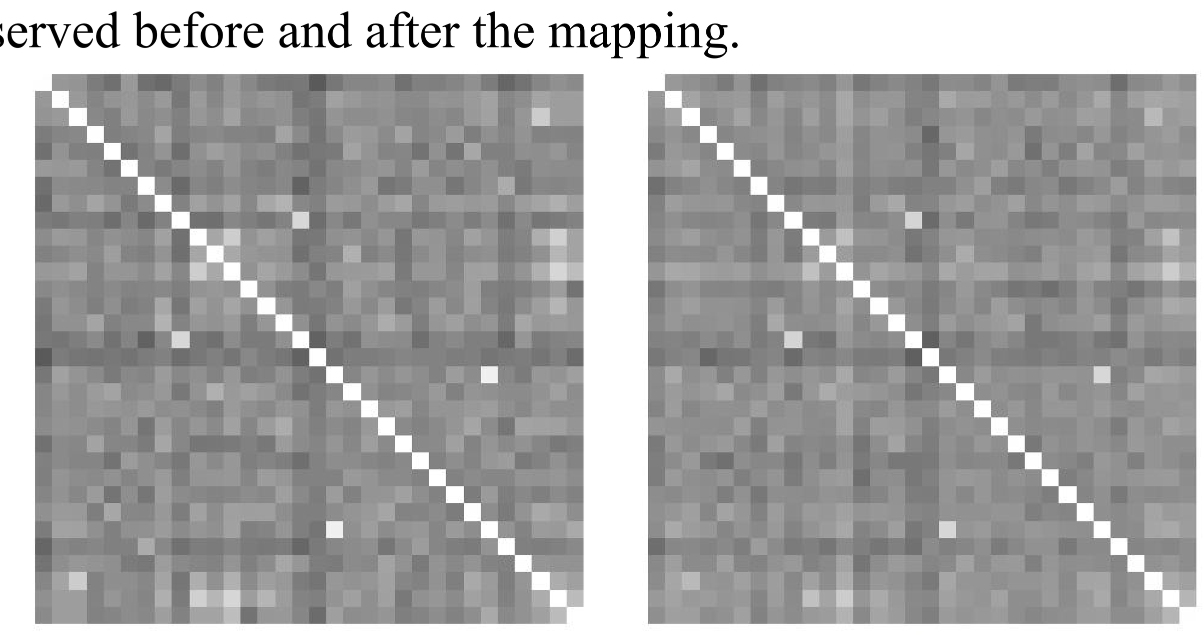
We compare our approach with state-of-the-art fake news detection methods, including EANN, SpotFake, BDANN, MVAE, MEAN, CAFE and MRML.

- The results of baselines and SPSSA are shown in Table 1. On the Weibo dataset, compared with MRML, the performance of our approach on Accuracy is 0.918, which significantly exceeds the MRML method by 2.1%. On the Fakeddit dataset, compared with CAFE [9], our approach leads by 4.1% on Accuracy. The leadership of the two data sets proves that our method can effectively reduce information redundancy and preserve both semantic and structural information.

Table 1: Comparison Results of Different Models on Weibo and Fakeddit Datasets

Method	Weibo							Fakeddit						
	Acc	Rumor			Non-rumor			Acc	Rumor			Non-rumor		
		P	R	F1	P	R	F1		P	R	F1	P	R	F1
EANN [KDD-2018]	0.782	0.827	0.697	0.756	0.752	0.863	0.804	0.724	0.727	0.719	0.723	0.722	0.729	0.726
SpotFake [BigMM-2019]	0.892	0.902	0.964	0.932	0.847	0.656	0.739	0.819	0.801	0.848	0.824	0.839	0.790	0.813
BDANN [IJCNN-2020]	0.842	0.830	0.870	0.850	0.850	0.820	0.830	0.812	0.836	0.776	0.805	0.791	0.847	0.818
HMCAN [SIGIR-2021]	0.885	0.920	0.845	0.881	0.856	0.926	0.890	0.881	0.880	0.882	0.881	0.882	0.880	0.881
MEAN [SPR-2022]	0.894	0.900	0.870	0.890	0.890	0.910	0.90	0.910	0.930	0.890	0.910	0.890	0.930	0.910
CAFE [WWW-2022]	0.840	0.855	0.830	0.842	0.825	0.851	0.837	0.912	0.946	0.886	**0.959**	0.878	0.942	0.909
MRML [ICASSP-2023]	0.897	0.898	0.887	0.892	**0.896**	0.905	0.901	0.840	0.819	0.874	0.846	0.865	0.807	0.835
SDSA*	0.898	0.916	0.888	0.902	0.880	0.909	0.894	0.941	0.946	0.936	0.941	0.936	0.947	0.942
SDSA*+S	0.905	0.919	0.899	0.909	0.891	0.912	0.901	0.944	0.952	0.935	0.943	0.936	0.953	0.944
SDSA*+I	0.906	0.924	0.895	0.909	0.887	0.919	0.903	0.950	0.946	**0.954**	0.950	**0.954**	0.946	0.950
SDSA	**0.918**	**0.939**	**0.902**	**0.920**	**0.896**	**0.935**	**0.915**	**0.953**	**0.965**	0.939	0.952	0.940	**0.966**	**0.953**

- To observe whether structural information is effectively preserved, we convert the similarity matrix of the dataset into grayscale images. The grayscale values represent the similarity between samples, with higher similarity resulting in brighte shades. The results are shown in Figure 2. By examining the distribution of the grayscale images, it is evident that the neighborhood structure for the samples have been well preserved before and after the mapping.



(a) (b)

Fig. 2: The grayscale images of the adjacency structures before and after mapping. (a) represents the grayscale image of the joint adjacency structure, (b) represents the grayscale image of the fused feature adjacency structure.

4. Conclusion

- In this paper, we introduced a semantic distillation and structural alignment network (SDSA) for fake news detection.
- We utilize the semantic distillation module to distill features, retaining meaningful semantic information while eliminating redundant information.
- A triple similarity alignment module is proposed to preserve both structural and semantic information. Compared with the current state-of-the-art methods, our method shows higher accuracy on two widely used fake news detection datasets.
- In future work, our goal is to explore more complex multimodal fake news detection problems, particularly in semi supervised scenarios.

Corresponding Author Shangdong Liu
lsd@njupt.edu.cn