Experiments

Evaluation Questions

- EQ1: Are Dual Emotion Features more effective than baseline features when used alone for fake news detection? How effective are the different types of features in Dual Emotion Features?
- EQ2: Can Dual Emotion Features help improve the performance of text-based fake news detectors?
- EQ3: How robust do the fake news detection models with Dual Emotion Features in real-world scenarios?
- EQ4: How effective are the components of Dual Emotion Features, including the publisher emotion, social emotion, and emotion gap?

Experiments

Ablation Study

- Observed that adding Dual Emotion Features into detectors all obtain the highest macro F1 scores.
- Exhibits that adopting Social Emotion or Emotion Gap improves the macro F1 scores more than Publisher Emotion on any models.
- Concludes that Social Emotion and Emotion Gap matter more when detecting fake news.

Models		R-19	W-16	W-20	W-20(t)
BiGRU+	Publisher Emotion	0.310	0.809	0.842	0.681
	Social Emotion	0.322	0.818	0.847	0.693
	Emotion Gap	0.336	0.811	0.849	0.693
	Dual Emotion Features	**0.340**	**0.826**	**0.855**	**0.701**
BERT+	Publisher Emotion	0.312	0.850	0.889	0.705
	Social Emotion	0.339	0.856	0.911	0.730
	Emotion Gap	0.338	0.858	0.906	0.731
	Dual Emotion Features	**0.346**	**0.867**	**0.915**	**0.734**
Nile TMRG+	Publisher Emotion	0.311	-	-	-
	Social Emotion	0.325	-	-	-
	Emotion Gap	0.337	-	-	-
	Dual Emotion Features	**0.342**	-	-	-
HSA-BLSTM+	Publisher Emotion	-	0.876	0.915	0.779
	Social Emotion	-	0.892	0.922	0.792
	Emotion Gap	-	0.901	0.926	0.800
	Dual Emotion Features	-	**0.908**	**0.932**	**0.805**

Table 8: Ablation study of the three components of *Dual Emotion Features*. The evaluation metric is macro F1 scores. R-19: RumourEval-19, W-16: Weibo-16, W-20: Weibo-20, and W-20(t): temporally split Weibo-20.