

Experiments

Performance Comparison: Weibo

Models	Weibo-16				Weibo-20			
	Macro F1 score	Accuracy	F1 score		Macro F1 score	Accuracy	F1 score	
			Fake	Real			Fake	Real
BiGRU	0.807	0.822	0.754	0.860	0.839	0.839	0.839	0.839
+ Emoratio	0.794	0.810	0.738	0.851	0.850	0.850	0.854	0.846
+ EmoCred	0.766	0.778	0.711	0.820	0.829	0.829	0.836	0.821
+ Dual Emotion Features	0.826	0.838	0.781	0.871	0.855	0.855	0.857	0.852
BERT	0.824	0.845	0.762	0.886	0.900	0.900	0.900	0.900
+ Emoratio	0.837	0.857	0.780	0.894	0.901	0.901	0.900	0.902
+ EmoCred	0.849	0.867	0.797	0.901	0.902	0.902	0.901	0.903
+ Dual Emotion Features	0.867	0.873	0.837	0.896	0.915	0.915	0.913	0.918
HSA-BLSTM	0.849	0.855	0.819	0.879	0.913	0.913	0.912	0.914
+ Emoratio	0.863	0.872	0.829	0.898	0.920	0.920	0.920	0.920
+ EmoCred	0.854	0.861	0.822	0.886	0.903	0.903	0.902	0.905
+ Dual Emotion Features	0.908	0.913	0.885	0.930	0.932	0.932	0.932	0.933

- Observed that Dual Emotion Features **outperforms** Emoratio & EmoCred on any models.
- However, when using Emoratio & EmoCred on BiGRU, sometimes the metrics even decrease, it reveals that Emoratio & EmoCred are more likely to be **overfitted**.

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?