Experiments....

Performance Comparison: Twitter Dataset

Method	Accuracy	Precision	Recall	F1
Text	0.532	0.598	0.541	0.568
Vis	0.596	0.695	0.518	0.593
VQA	0.631	0.765	0.509	0.611
NeuralTalk	0.610	0.728	0.504	0.595
att-RNN	0.664	0.749	0.615	0.676
EANN-	0.648	0.810	0.498	0.617
EANN	**0.715**	**0.822**	**0.638**	**0.719**

- **att-RNN** performs better than **VQA** and **NeuralTalk**
  - Shows that applying attention mechanism can improve
- **EANN-** tend to capture the event-specific features
  - Would lead failure of learning enough shared features among events
- **EANN** significantly improves the performance in terms of all the measures

Experiments.....

Performance Comparison: Weibo Dataset

Method	Accuracy	Precision	Recall	F1
Text	0.763	0.827	0.683	0.748
Vis	0.615	0.615	0.677	0.645
VQA	0.773	0.780	0.782	0.781
NeuralTalk	0.717	0.683	**0.843**	0.754
att-RNN	0.779	0.778	0.799	0.789
EANN-	0.795	0.806	0.795	0.800
EANN	**0.827**	**0.847**	0.812	**0.829**

- Similar result can be observed as those on Twitter dataset.
- However, we can see that **Text** is greatly higher than that of **Vis**
  - Because Weibo dataset doesn't have imbalanced issue, and with sufficient diversity, useful linguistic patterns can be extracted.
  - Images of Weibo dataset are more complicated in semantic meaning than Twitter.
  - **Vis** can't learn meaningful representations, although use VGG19 (But MVNN can?)