

Model Details

Graph Attention Network (GAT)

Velikovi et al. (2018) introduced graph attention networks to address various shortcomings of GCNs. Most importantly, it enables nodes to attend over their neighborhoods features without depending on the graph structure upfront. The key idea is to compute the hidden representations of each node in the graph, by attending over its neighbors, following a self-attention (Vaswani et al., 2017) strategy. Let $Z^l = \{z_1, z_2, \dots, z_n\}$ be the input node features and $Z^{l+1} = \{z'_1, z'_2, \dots, z'_n\}$ be the output node features. Here n is the number of nodes in the graph and $z_i \in \mathcal{R}^F$, $z'_i \in \mathcal{R}^{F'}$, where F is the dimension of the input feature and F' is the dimension of the output feature. The attention coefficients are computed by,

$$e_{ij} = f(W'[Wh_i || Wh_j]) \quad (1)$$

Here, W is a shared linear transformation $\in \mathcal{R}^{F \times F'}$. f is the activation function, W' is another linear layer $\in \mathcal{R}^{F' \times 1}$. Since we have a fully connected graph, we normalize the attention coefficients across all the other nodes,

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1, k \neq i}^n \exp(e_{ik})} \quad (2)$$

The final output features are computed as follows,

$$h'_i = \sigma \left(\sum_j \alpha_{ij} Wh_j \right) \quad (3)$$

For the *GAT + 2 Attn Heads* model,

$$h'_i = \sigma \left(\sum_j \alpha_{ij} W_1 h_j \right) \parallel \sigma \left(\sum_j \alpha_{ij} W_2 h_j \right)$$

We can see that in the entire formulation, the adjacency matrix is not used, and hence for our fully connected graph, *GAT* is essentially equivalent to *GAT + SS*.

Qualitative Results

We closely inspect the attention maps generated by the GAT model for the four way classification, as shown in Figure 1.

We see that the model is learning to attend to specific sentences for trusted news articles and is attending to all the sentences for satirical articles.

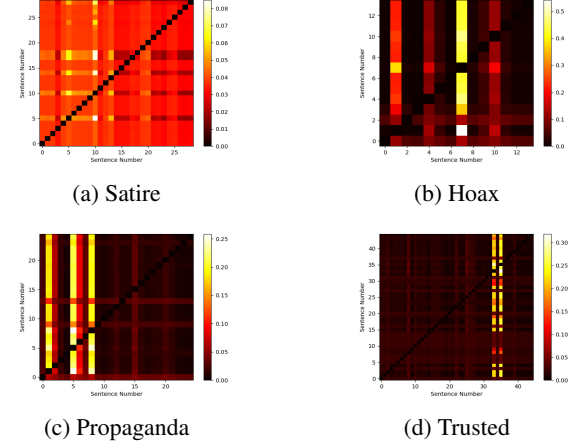


Figure 1: Attention heatmaps generated by GAT for different types of news articles in 4-way classification.

For the satirical articles, since there are no factual jumps, the model can't easily assign higher weights to particular sentence representation. This results in a heatmap with almost equal weights assigned to all the sentences. However, there is no obvious difference in the kind of attention maps generated for the hoax and propaganda articles. From the confusion matrix shown in Figure 2, we see that the model is highly confused between hoax and propaganda.

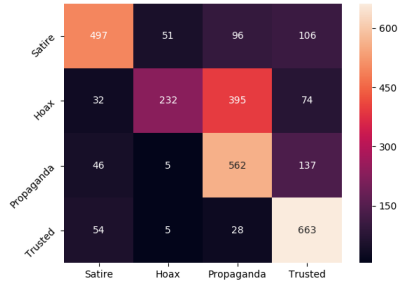


Figure 2: Confusion matrix for the GAT model in a 4-way classification setting on LUN-test.

This confusion explains the similar nature of attention maps generated for hoax and propaganda articles. This behaviour is as per our expectation, because we faced the most confusion in classifying propaganda and hoax, given the similar nature of these kinds of articles.

References

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all

you need. In *Advances in neural information processing systems*, pages 5998–6008.

Petar Velickovi, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Li, and Yoshua Bengio. 2018. [Graph attention networks](#). In *International Conference on Learning Representations*.