GraphMERT

I. Embedding Layer

(A) No H-GAT (no leaves)



(B) With H-GAT on leaves

Fused Node Feature t_i' = H-GAT(t_i , a_r , W_r , h_j), j = 1...m

Relation embedding

matrix W_r

Semantic token t_i

Relation a_r

Syntactic token h_m

All head tokens

II. Transformer Layers

