

Utterance1

Utterance2

 $f_{id}$   $T' \times \text{num}_f$  $f_{id}$  $f_{raw}$   $T' \times 512$  $f_{raw}$ *Self-attention* $T' \times \text{num}_f/2$  $T' \times \text{num}_f$  $\text{num}_f$  $T' \times \text{num}_f$  $T' \times \text{num}_f$  $T' \times \text{num}_f$  $\text{num}_f$  $f_{self}$  $f_{self}$ *Mutual-attention* $T' \times \text{num}_f/2$  $T' \times \text{num}_f$  $T' \times \text{num}_f$  $T' \times \text{num}_f$  $T' \times \text{num}_f$  $\text{num}_f$  $f_{mutual}$  $f_{mutua}$ 

FC

Sum

Avg pool

Softmax

⊗ Element-wise product