

- (a)
- (b)
- (c)
- (d)
- (e)
- (f)
- (g) In attention computation, enc_{masks} is used to remove attention from padded tokens. If we do in this way, the attention from padded tokens will attend in the prediction, which disobeys intuition and may cause unknown affects.