

Sentiment Intensity Prediction

3.1

Multimodal Fusion

\mathbf{x}_1^V

\mathbf{x}_2^V

...

$\mathbf{x}_{T_v}^V$

\mathbf{x}_1^T

\mathbf{x}_2^T

\mathbf{x}_3^T

\mathbf{x}_4^T

\mathbf{x}_5^T

\mathbf{x}_6^T

\mathbf{x}_7^T

\mathbf{x}_1^A

\mathbf{x}_2^A

...

$\mathbf{x}_{T_A}^A$

Pretrained Language Model

[CLS]

The

food

is

very

[MASK]

!

[CLS]

The

food

is

very

great

!

Sentiment Word Masking