

000 A Appendix

001 A.1 Reproducibility of the paper

002 We implement experiments on GTX 1080Ti. The
 003 main hyperparameters include audio hidden dimension,
 004 video hidden dimension, text hidden dimension,
 005 audio dropout rate, video dropout rate, text
 006 dropout rate, learning rate, weight decay, rank1
 007 and rank2. Grid search is employed to find the
 008 appropriate combination of parameters. For each
 009 method, we randomly try 2000 combinations, since
 010 the model is small and the running time is short as
 011 shown in Table 1. The feature extraction method
 012 and the division of training and test sets follow
 013 (Zadeh et al., 2018). If the paper is accepted,
 014 we promise to open the source code and the best-
 015 performing hyperparameters.

016 Table 1: The model size and execution time of our
 017 methods.

Method	Size(MB)	Time(s)
TFN	1163	19.3
LMF	627	11.7
FT-LMF	1097	17.9
Dual-LMF	721	13.0

024 A.2 Derivations for Eqns. 5 and 6 in the 025 paper

026 $W_h^i \cdot \tilde{V}$ can be rewritten as:

$$027 W_h^i \cdot \tilde{V} = \left[\sum_{r=1}^R \bigotimes_{m=1}^M (W_h^i)_{m,r} \right] \cdot \left[\bigotimes_{m=1}^M v_m \right] \quad (1)$$

028 where “.” denotes linear operation for $\bigotimes_{m=1}^M v_m$.
 029 Since $\sum_{r=1}^R \bigotimes_{m=1}^M (W_h^i)_{m,r}$ and $\bigotimes_{m=1}^M v_m$ have
 030 the same size $R^{\prod_{m=1}^M d_m}$, we can rewrite the linear
 031 operation as the combination of element-wise mul-
 032 tiplication and summation. The two formations are
 033 equivalent.

$$039 W_h^i \cdot \tilde{V} = \sum_{r=1}^R \left[\sum_{m=1}^M \left[(W_h^i)_{m,r} \circ \bigotimes_{m=1}^M v_m \right] \right] \\ 040 = \sum_{r=1}^R \left[\sum_{m=1}^M \left[(W_h^i)_{m,r} \circ \bigotimes_{m=1}^M v_m \right] \right] \quad (2)$$

045 where $\bigotimes_{m=1}^M (W_h^i)_{m,r} \circ \bigotimes_{m=1}^M v_m$ can be rewr-
 046 itten as another formation, $\bigotimes_{m=1}^M \left[(W_h^i)_{m,r} \circ v_m \right]$.
 047 The equivalence can be proven by element-wise
 048 comparison:

049 Proposition 1.

$$\bigotimes_{m=1}^M (W_h^i)_{m,r} \circ \bigotimes_{m=1}^M v_m = \bigotimes_{m=1}^M \left[(W_h^i)_{m,r} \circ v_m \right] \quad (3)$$

050 *Proof.*

$$051 \begin{aligned} & \left[\bigotimes_{m=1}^M (W_h^i)_{m,r} \circ \bigotimes_{m=1}^M v_m \right]_{c_1, c_2, \dots, c_M} \\ &= \left[\bigotimes_{m=1}^M (W_h^i)_{m,r} \right]_{c_1, c_2, \dots, c_M} \circ \left[\bigotimes_{m=1}^M v_m \right]_{c_1, c_2, \dots, c_M} \\ &= \left[(W_h^i)_{1,r} \right]_{c_1} \circ \dots \circ \left[(W_h^i)_{M,r} \right]_{c_M} \circ (v_1)_{c_1} \circ \dots \circ (v_M)_{c_M} \\ &= \left\{ \left[(W_h^i)_{1,r} \right]_{c_1} \circ (v_1)_{c_1} \right\} \circ \dots \circ \left\{ \left[(W_h^i)_{M,r} \right]_{c_M} \circ (v_M)_{c_M} \right\} \\ &= \left[(W_h^i)_{1,r} \circ v_1 \right]_{c_1} \circ \dots \circ \left[(W_h^i)_{M,r} \circ v_M \right]_{c_M} \\ &= \left[\bigotimes_{m=1}^M \left[(W_h^i)_{m,r} \circ v_m \right] \right]_{c_1, c_2, \dots, c_M} \end{aligned} \quad (4)$$

052 where $c_1, c_2, \dots, c_M (c_m \in [1, 2, \dots, d_m])$ denotes
 053 the index of the elements in high-order tensor. \square

054 $W_h^i \cdot \tilde{V}$ can be rewritten as follows:

$$055 W_h^i \cdot \tilde{V} = \sum_{r=1}^R \left[\sum_{m=1}^M \left[(W_h^i)_{m,r} \circ v_m \right] \right] \quad (5)$$

056 where $\sum_{m=1}^M \left[(W_h^i)_{m,r} \circ v_m \right]$ can be rewritten
 057 as another formation, $\prod_{m=1}^M \left[(W_h^i)_{m,r}^\top v_m \right]$. The
 058 equivalence can be proven as follows:

059 Proposition 2.

$$060 \sum_{m=1}^M \left[(W_h^i)_{m,r} \circ v_m \right] = \prod_{m=1}^M \left[(W_h^i)_{m,r}^\top v_m \right] \quad (6)$$

061 *Proof.*

$$062 \prod_{m=1}^M \left[(W_h^i)_{m,r}^\top v_m \right] = \prod_{m=1}^M \sum \left[(W_h^i)_{m,r} \circ v_m \right] \quad (7)$$

063 Following the simple transformation like $(a + b)(c + d) = ac + ad + bc + bd$, we can
 064 easily transform $\prod_{m=1}^M \sum \left[(W_h^i)_{m,r} \circ v_m \right]$ to
 065

100 $\sum \bigotimes_{m=1}^M \left[(W_h^i)_{m,r} \circ v_m \right]$. These two formations
 101 are equal, just with different operation orders. The
 102 former utilizes summation(\sum) first, while the later
 103 uses multiplication(\bigotimes) between different elements
 104 first. \square
 105

106 Therefore, we obtain the final formation of
 107 $W_h^i \tilde{V}$:
 108

$$109 W_h^i \cdot \tilde{V} = \sum_{r=1}^R \prod_{m=1}^M \left[(W_h^i)_{m,r}^T v_m \right] \quad (8)$$

112 A.3 Derivations for Eqn. 17 in the paper

113 $W_k^j \cdot H_i$ can be rewritten as:
 114

$$115 W_k^j \cdot H_i = \left[\sum_{r_2=1}^{R_2} \bigotimes_{m=1}^M (W_k^j)_{m,r_2} \right] \cdot \left[\sum_{r_1=1}^{R_1} \bigotimes_{m=1}^M \left[(W_h^i)_{m,r_1}^T V_m \right] \right] \quad (9)$$

116 similar to Eqns. 2, 5, and 8, we obtain the final
 117 formation of $W_k^j \cdot H_i$,
 118

$$122 W_k^j \cdot H_i \\ 123 = \sum \left[\sum_{r_2=1}^{R_2} \sum_{r_1=1}^{R_1} \left[\bigotimes_{m=1}^M (W_k^j)_{m,r_2} \circ \bigotimes_{m=1}^M [(W_h^i)_{m,r_1}^T V_m] \right] \right] \\ 124 = \sum \left[\sum_{r_2=1}^{R_2} \sum_{r_1=1}^{R_1} \bigotimes_{m=1}^M \left[(W_k^j)_{m,r_2} \circ [(W_h^i)_{m,r_1}^T V_m] \right] \right] \\ 125 = \sum_{r_2=1}^{R_2} \sum_{r_1=1}^{R_1} \left[\sum_{m=1}^M \left[(W_k^j)_{m,r_2} \circ [(W_h^i)_{m,r_1}^T V_m] \right] \right] \\ 126 = \sum_{r_2=1}^{R_2} \sum_{r_1=1}^{R_1} \prod_{m=1}^M \left[(W_h^i)_{m,r_1}^T V_m (W_k^j)_{m,r_2} \right] \\ 127 \quad (10)$$

137 References

138 Amir Zadeh, Paul Pu Liang, Soujanya Poria, Pra-
 139 teek Vij, Erik Cambria, and Louis-Philippe Morency.
 140 2018. Multi-attention recurrent network for human
 141 communication comprehension. In *AAAI*.

100	$\sum \bigotimes_{m=1}^M \left[(W_h^i)_{m,r} \circ v_m \right]$. These two formations	150
101	are equal, just with different operation orders. The	151
102	former utilizes summation(\sum) first, while the later	152
103	uses multiplication(\bigotimes) between different elements	153
104	first. \square	154
105		155
106	Therefore, we obtain the final formation of	156
107	$W_h^i \tilde{V}$:	157
108		158
109	$W_h^i \cdot \tilde{V} = \sum_{r=1}^R \prod_{m=1}^M \left[(W_h^i)_{m,r}^T v_m \right]$	159
110		160
111		161
112	A.3 Derivations for Eqn. 17 in the paper	162
113	$W_k^j \cdot H_i$ can be rewritten as:	163
114		164
115	$W_k^j \cdot H_i = \left[\sum_{r_2=1}^{R_2} \bigotimes_{m=1}^M (W_k^j)_{m,r_2} \right] \cdot \left[\sum_{r_1=1}^{R_1} \bigotimes_{m=1}^M \left[(W_h^i)_{m,r_1}^T V_m \right] \right]$	165
116		166
117		167
118		168
119	similar to Eqns. 2, 5, and 8, we obtain the final	169
120	formation of $W_k^j \cdot H_i$,	170
121		171
122	$W_k^j \cdot H_i$	172
123	$= \sum \left[\sum_{r_2=1}^{R_2} \sum_{r_1=1}^{R_1} \left[\bigotimes_{m=1}^M (W_k^j)_{m,r_2} \circ \bigotimes_{m=1}^M [(W_h^i)_{m,r_1}^T V_m] \right] \right]$	173
124		174
125		175
126		176
127	$= \sum \left[\sum_{r_2=1}^{R_2} \sum_{r_1=1}^{R_1} \bigotimes_{m=1}^M \left[(W_k^j)_{m,r_2} \circ [(W_h^i)_{m,r_1}^T V_m] \right] \right]$	177
128		178
129	$= \sum_{r_2=1}^{R_2} \sum_{r_1=1}^{R_1} \left[\sum_{m=1}^M \left[(W_k^j)_{m,r_2} \circ [(W_h^i)_{m,r_1}^T V_m] \right] \right]$	179
130		180
131		181
132	$= \sum_{r_2=1}^{R_2} \sum_{r_1=1}^{R_1} \prod_{m=1}^M \left[(W_h^i)_{m,r_1}^T V_m (W_k^j)_{m,r_2} \right]$	182
133		183
134		184
135		185
136		186
137	References	187
138	Amir Zadeh, Paul Pu Liang, Soujanya Poria, Pra-	188
139	teek Vij, Erik Cambria, and Louis-Philippe Morency.	189
140	2018. Multi-attention recurrent network for human	190
141	communication comprehension. In <i>AAAI</i> .	191
142		192
143		193
144		194
145		195
146		196
147		197
148		198
149		199