# A Appendices

## A.1 Ablation study

We also conduct additional ablation studies in terms of the following aspects:

• The number of hops: The proposed framework consists of a multi-hop Dual Memory Interaction (DMI) to gradually refine the global memory representations and correlation vectors for each word. The multiple hops are necessary to distill complicated latent relations among aspect and opinion words as transferable knowledge, and also provide a more confident aspect attention distribution as a learnable alignment weight for each word. To validate that, we conduct experiments with different number of hops as shown in Table 1. As we can see, increasing the number of hops from 1 to 2 shows significant improvements for the AD and ADS tasks with the best performance obtained using 2 hops. However, when further increasing the number of hops to 3, the performance will be degraded. A possible reason for the degradation with more hops can be that it brings in more noises with over-extraction and provides very sharp alignment weights for the fine-grained adaptation.

| Transfer Pair | 1 hop | | 2 hops | | 3 hops | |
|---|---|---|---|---|---|---|
| | AD | ADS | AD | ADS | AD | ADS |
| $\mathbb{S}\rightarrow\mathbb{R}$ | 45.45 | 29.05 | **52.05** | **41.03** | 49.68 | 34.10 |
| $\mathbb{L}\rightarrow\mathbb{R}$ | 48.64 | 38.33 | **56.12** | **43.04** | 51.76 | 40.48 |
| $\mathbb{D}\rightarrow\mathbb{R}$ | 44.09 | 34.41 | **51.55** | **41.01** | 49.49 | 40.95 |
| $\mathbb{R}\rightarrow\mathbb{S}$ | 34.37 | 26.44 | **39.02** | **28.01** | 27.57 | 19.64 |
| $\mathbb{L}\rightarrow\mathbb{S}$ | 35.45 | 26.83 | **38.26** | **27.20** | 36.98 | 27.59 |
| $\mathbb{D}\rightarrow\mathbb{S}$ | 28.12 | 18.58 | **36.11** | **26.62** | 32.22 | 22.88 |
| $\mathbb{R}\rightarrow\mathbb{L}$ | 40.69 | 29.02 | 45.01 | **34.13** | **46.80** | 33.99 |
| $\mathbb{S}\rightarrow\mathbb{L}$ | 27.48 | 18.62 | **35.99** | **27.04** | 29.74 | 20.67 |
| $\mathbb{R}\rightarrow\mathbb{D}$ | 43.40 | 33.26 | **43.76** | **35.44** | 42.72 | 34.79 |
| $\mathbb{S}\rightarrow\mathbb{D}$ | 33.23 | 24.24 | **41.21** | **33.56** | 31.47 | 23.78 |
| Average<br>($\Delta$) | 38.09<br>(5.82) | 27.88<br>(5.83) | **43.91**<br>- | **33.71**<br>- | 39.84<br>(4.07) | 29.89<br>(3.82) |

Table 1: Ablation results (%). Comparisons with different number of hops. $\Delta$ refers to the improvements of the model with 2 hops over that with other number of hops.

• Alternating training v.s. Joint training: We investigate the effect of different training strategies, i.e., alternating training and joint training. As shown in Table 2, the model with alternating training outperforms that with joint training by 1.78% and 2.63% Micro-F1 on average for the AD and ADS tasks, respectively. This shows that when optimizing many objective functions simultaneously, the alternating training strategy can achieve better performances by separating the word representation learning into two stages, which could make the optimization easier.

| Transfer Pair | Joint training | | Alternating training | |
|---|---|---|---|---|
| | AD | ADS | AD | ADS |
| $\mathbb{S}\rightarrow\mathbb{R}$ | 48.06 | 33.28 | **52.05** | **41.03** |
| $\mathbb{L}\rightarrow\mathbb{R}$ | 55.04 | 43.02 | **56.12** | **43.04** |
| $\mathbb{D}\rightarrow\mathbb{R}$ | 51.06 | 40.95 | **51.55** | **41.01** |
| $\mathbb{R}\rightarrow\mathbb{S}$ | **39.07** | 25.89 | 39.02 | **28.01** |
| $\mathbb{L}\rightarrow\mathbb{S}$ | **40.74** | **30.74** | 38.26 | 27.20 |
| $\mathbb{D}\rightarrow\mathbb{S}$ | 35.13 | 24.13 | **36.11** | **26.62** |
| $\mathbb{R}\rightarrow\mathbb{L}$ | **46.18** | 32.09 | 45.01 | **34.13** |
| $\mathbb{S}\rightarrow\mathbb{L}$ | 29.98 | 21.32 | **35.99** | **27.04** |
| $\mathbb{R}\rightarrow\mathbb{D}$ | 42.10 | 35.38 | **43.76** | **35.44** |
| $\mathbb{S}\rightarrow\mathbb{D}$ | 33.97 | 24.03 | **41.21** | **33.56** |
| Average<br>($\Delta$) | 42.13<br>(1.78) | 31.08<br>(2.63) | **43.91**<br>- | **33.71**<br>- |

Table 2: Ablation results (%). Alternating training v.s. Joint training. $\Delta$ refers to the improvements of the model with alternating training over that with joint training.