TransESC: Smoothing Emotional Support Conversation via Turn-Level State Transition

Weixiang Zhao, Yanyan Zhao* , Shilong Wang, Bing Qin

Research Center for Social Computing and Information Retrieval

Harbin Institute of Technology, China

{wxzhao, yyzhao}@ir.hit.edu.cn

code: https://github.com/circle-hit/TransESC
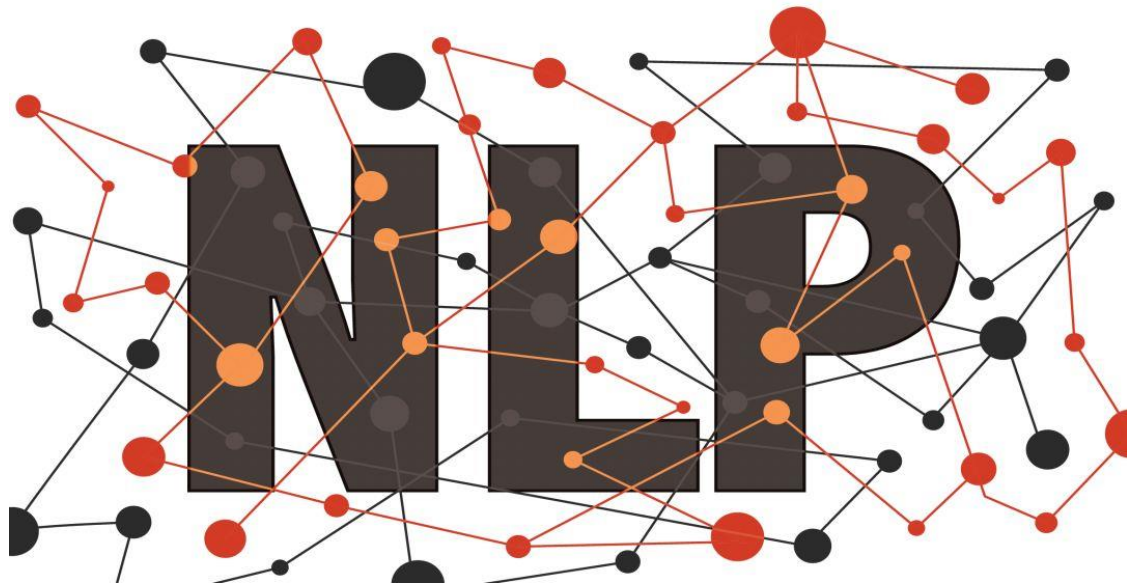
2023. 5. 18   •   ChongQing

2023_ACL

Reported by Junhao Cao
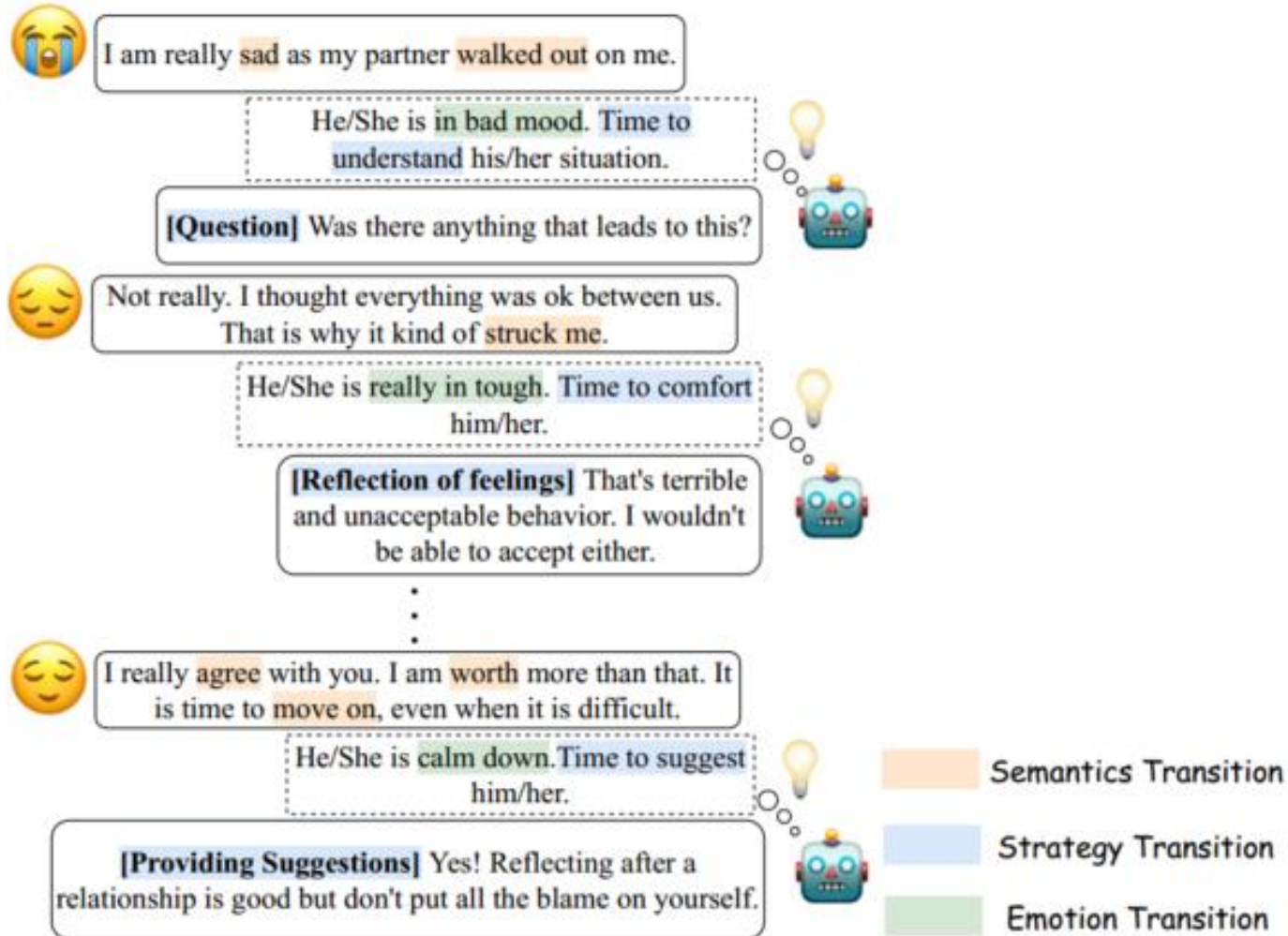
NATURAL LANGUAGE PROCESSING
NLP

1.Introduction

2.Method

3.Experiments

Introduction

Previous works ignoring to grasp the fine-grained transition information at each dialogue turn.
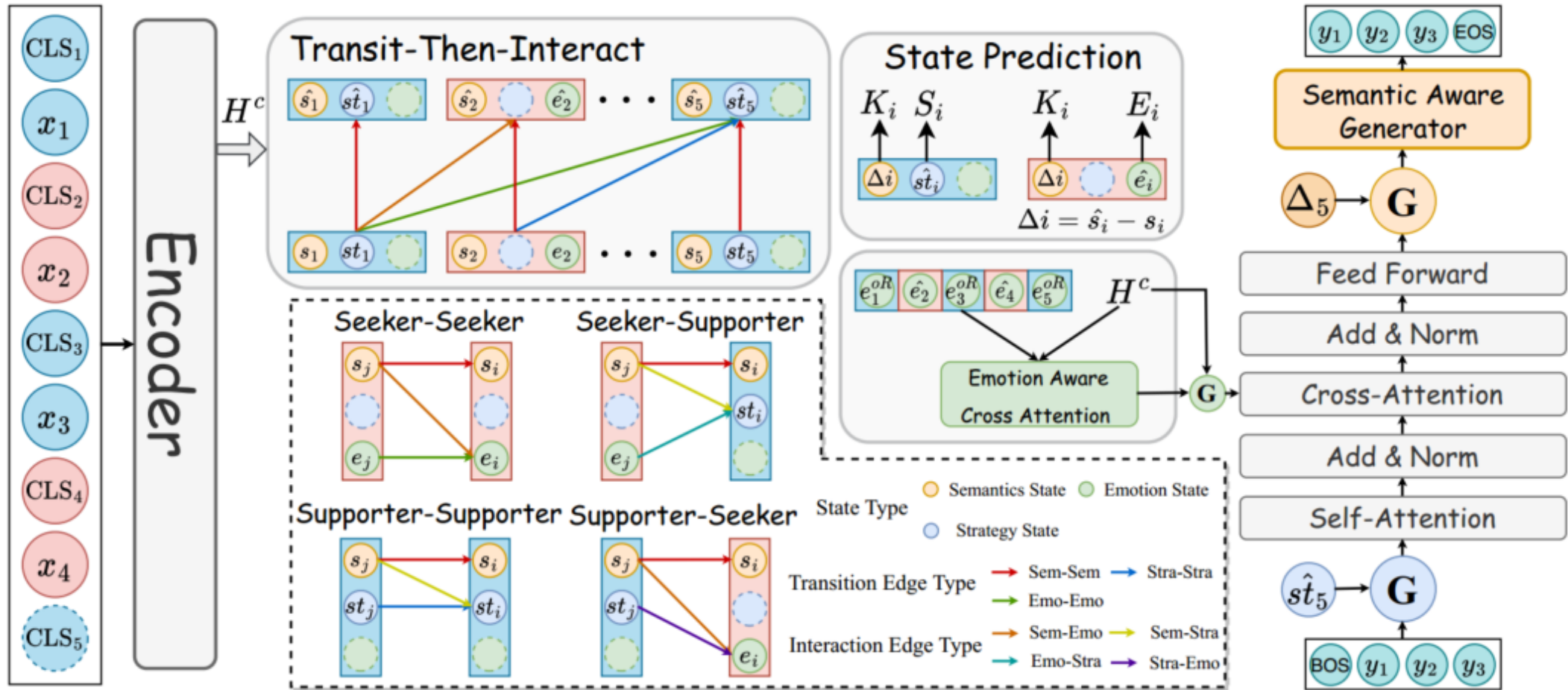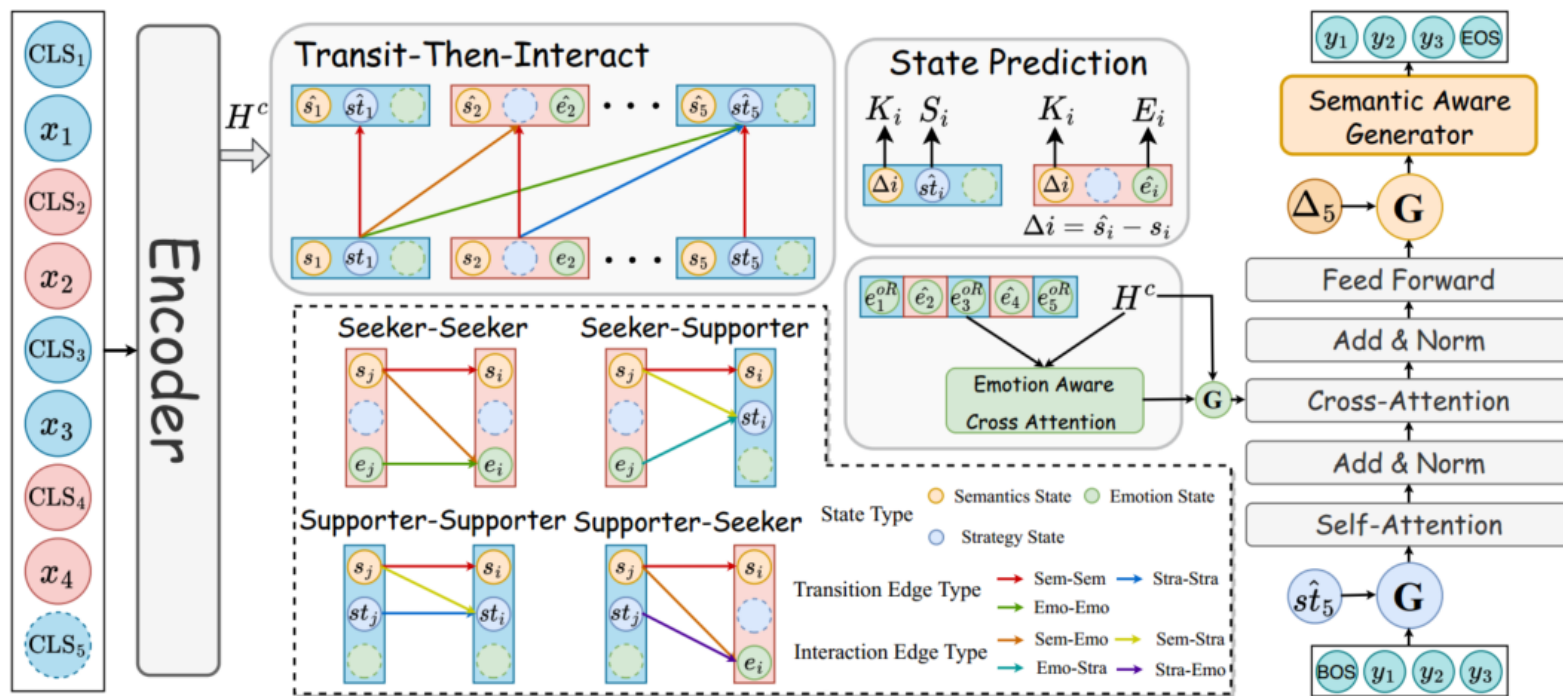
Method

Figure 2: The overall architecture of our proposed TransESC model, which mainly consists of three modules: Context Encoder, Turn-Level State Transition Module and Transition-Aware Decoder.

Method

$$D = [X_1, X_2, \cdots, X_N]$$

$$X_i = [w_1^i, w_2^i \cdots, w_m^i]$$

$$K_i = [k_1^i, k_2^i \cdots, k_k^i]$$

support strategy S_i of the supporter

emotional state label E_i of the seeker

Turn-Level State Transition

$$\hat{v}_i = \underset{j \in \mathcal{N}}{\mathrm{MHA}}(q_i, k_j, v_j), \qquad (1)$$

$$s'_i = \underset{e_{ij} \in \mathcal{T}}{\text{R-MHA}}(s_i + r_{ij}, s_j + r_{ij}, s_j), \qquad (2)$$

$$s''_i = \underset{e_{ij} \in \mathcal{I}}{\text{R-MHA}}(s'_i + r_{ij}, s'_j + r_{ij}, s'_j), \qquad (3)$$

$$\hat{s}_i = g^{tti} \odot s'_i + (1 - g^{tti}) \odot s''_i$$
$$g^{tti} = \sigma([s'_i; s''_i]W^{tti} + b^{tti}) \qquad (4)$$

Method
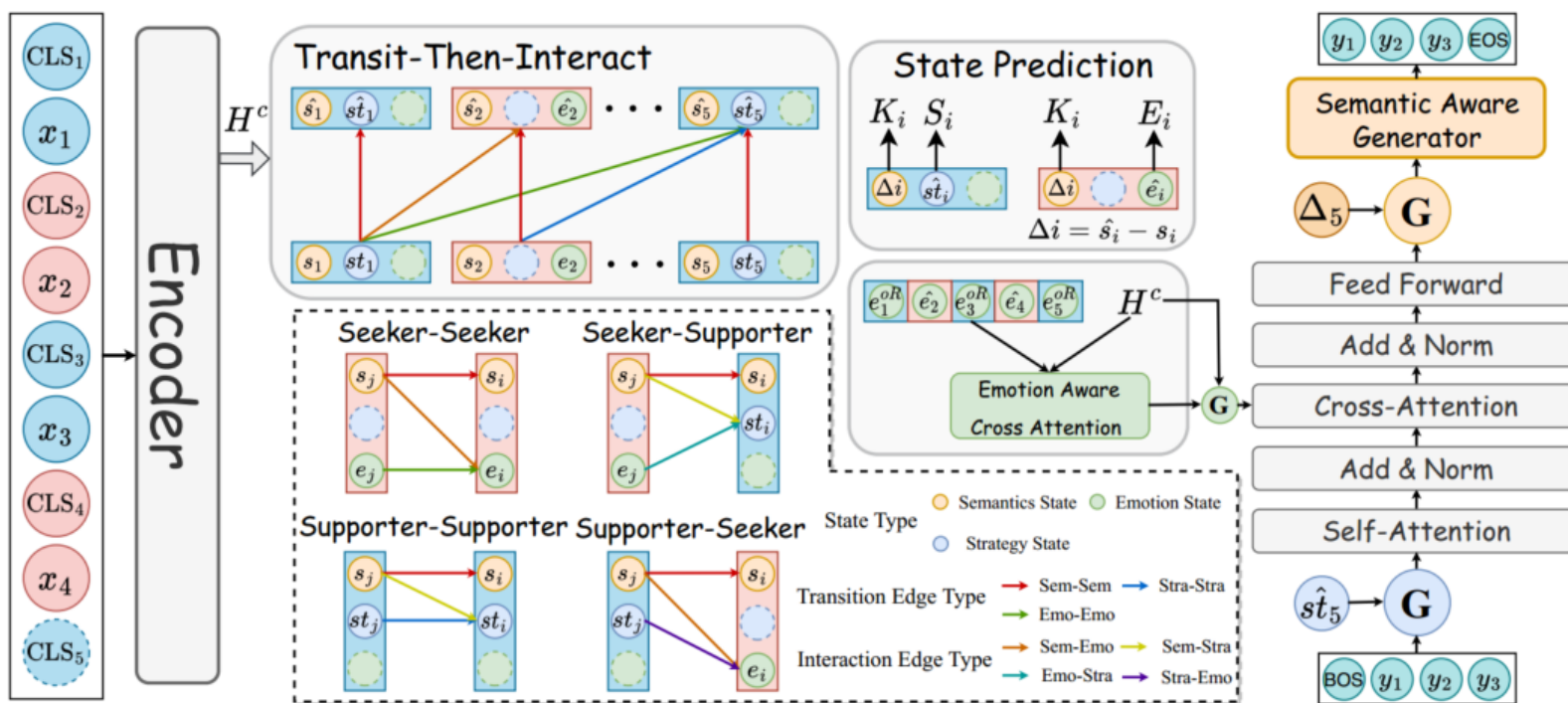
State Prediction

Semantic Keyword Prediction

$$\Delta_i = \hat{s}_i - s_i$$

$$\mathcal{L}_{SEM} = -\sum_{i=1}^{N}\sum_{j=1}^{k}\log p(k_j^i|\Delta_i)$$

$$= -\sum_{i=1}^{N}\sum_{j=1}^{k}\log f_{k_j^i} \qquad (5)$$

$$f = \text{softmax}(W^{sem}\Delta_i + b^{sem}) \qquad (6)$$

Supporter Strategy Prediction

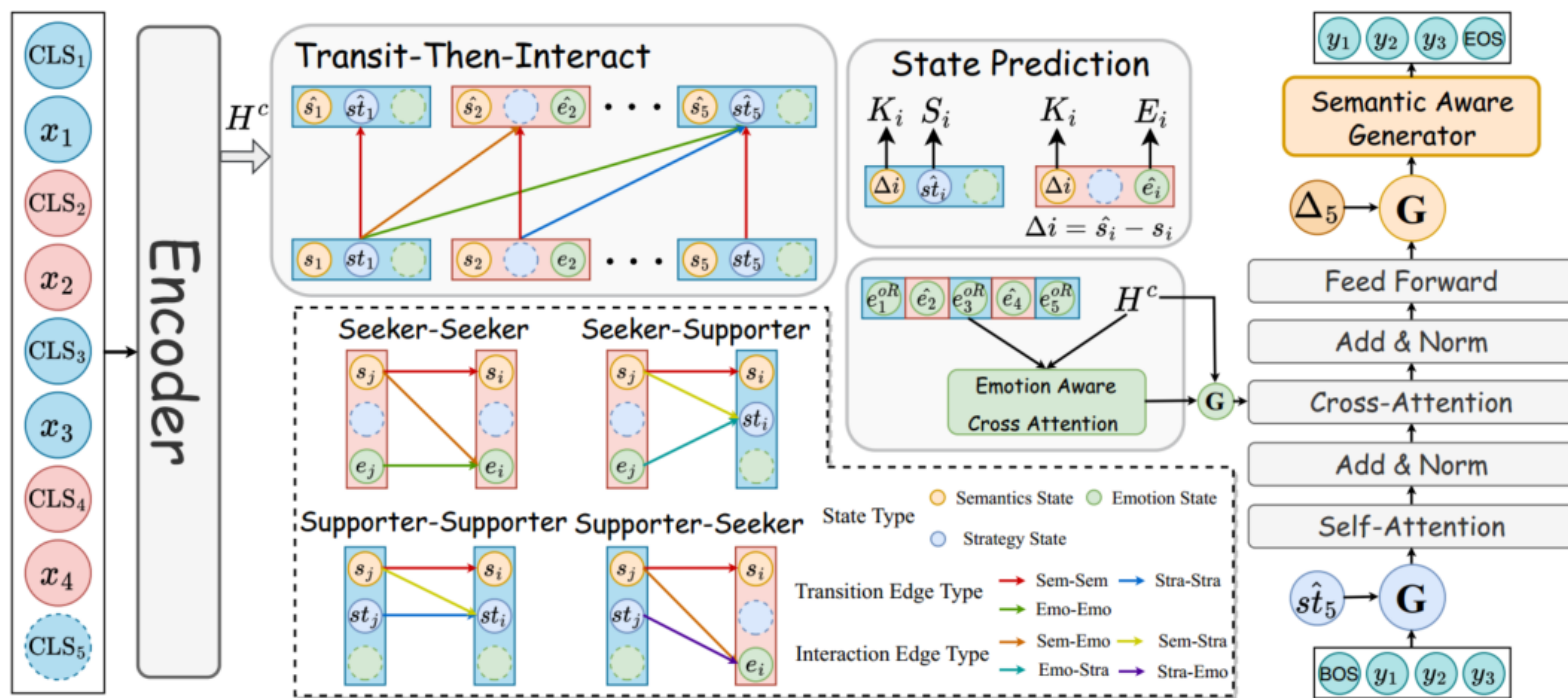$$\hat{y}_{str} = \text{softmax}(W^{str}\hat{st}_i + b^{str}) \qquad (7)$$

$$\mathcal{L}_{STR} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{n_s}\hat{y}_{str,i}^j \cdot log(y_{str,i}^j) \qquad (8)$$

Seeker Emotion Prediction

$$\hat{y}_{emo} = \text{softmax}(W^{emo}\hat{e}_i + b^{emo}) \qquad (9)$$

$$\mathcal{L}_{EMO} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{n_e}\hat{y}_{emo,i}^j \cdot log(y_{emo,i}^j) \qquad (10)$$

Method

Transition-Aware Decoder

$$\hat{E}_i = g^{str} \odot E_i + (1 - g^{str}) \odot \hat{st}$$
$$g^{str} = \sigma([E_i; \hat{st}]W^1 + b^1) \quad (11)$$

$$\hat{H} = g^{emo} \odot H^c + (1 - g^{emo}) \odot \hat{H}^{emo}$$
$$\hat{H}^{emo} = \text{Cross-Att}(H^c, H^{emo}) \quad (12)$$
$$g^{emo} = \sigma([H^c; \hat{H}^{emo}]W^2 + b^2)$$

$$h_t = \text{Decoder}(\hat{E}_{y<t}, \hat{H}) \quad (13)$$

$$\hat{h} = g^{sem} \odot h_t + (1 - g^{sem}) \odot \Delta_i$$
$$g^{sem} = \sigma([h_t; \Delta_i]W^{sem} + b^{sem}) \quad (14)$$

$$P(y_t \mid y_{<t}, D) = \text{softmax}(W\hat{h} + b) \quad (15)$$

$$L_{gen} = -\sum_{t=1}^{M} \log P(y_t \mid D, y_{<t}). \quad (16)$$

$$\mathcal{L} = \gamma_1 \mathcal{L}_{GEN} + \gamma_2 \mathcal{L}_{SEM} + \gamma_3 \mathcal{L}_{STR} + \gamma_4 \mathcal{L}_{EMO} \quad (17)$$

Experiment

Model	Acc	PPL	D-1	D-2	B-1	B-2	B-3	B-4	R-L
Transformer	-	89.61	1.29	6.91	-	6.53	-	1.37	15.17
Multi-TRS	-	89.52	1.28	7.12	-	6.58	-	1.47	14.75
MoEL	-	133.13	2.33	15.26	-	5.93	-	1.22	14.65
MIME	-	47.51	2.11	10.94	-	5.23	-	1.17	14.74
BlenderBot-Joint	17.69	17.39	2.96	17.87	18.78	7.02	3.20	1.63	14.92
GLHG	-	**15.67**	3.50	**21.61**	**19.66**	7.57	3.74	2.13	16.37
MISC	31.67	16.27	4.62	20.17	16.31	6.57	3.26	1.83	17.24
TransESC (Ours)	**34.71**	15.85	**4.73**	20.48	17.92	**7.64**	**4.01**	**2.43**	**17.51**

Table 1: Comparison of our model against state-of-the-art baselines in terms of the automatic evaluation. The best results among all models are highlighted in bold.

Experiment

TransESC vs.	BlenderBot-Joint			MISC		
	Win	Lose	Tie	Win	Lose	Tie
Fluency	**54.7**‡	18.0	27.3	**65.7**‡	10.7	23.7
Identification	**37.3**‡	16.0	46.7	**32.0**	19.3	48.7
Empathy	**39.3**‡	7.0	53.7	**48.0**‡	5.7	46.3
Suggestion	**37.0**	27.7	35.3	**46.7**†	17.3	36.0
Overall	**51.7**‡	26.0	22.3	**64.0**‡	17.7	18.3

Table 2: The results of the human interaction evaluation (%). TransESC performs better than all other models (sign test, $\ddagger$ / $\dagger$ represent p-value < 0.05 / 0.1).

Experiment

Model	Dist-1	B-2	B-4	R-L
TransESC	4.73	**7.64**	**2.43**	**17.51**
w/o Sem. Trans	4.55	7.04	2.13	17.37
w/o Stra. Trans	4.29	6.68	2.01	17.15
w/o Emo. Trans	**4.82**	7.14	2.22	17.45
w/o T-L. Trans	4.19	6.35	1.94	16.88

Table 3: Results of ablation study. Sem./Stra./Emo./T-L. Trans refer to the semantics/strategy/emotion/all three types of turn-level transition, respectively.

Experiment

Situation	There is no hope, I am struggling with the pandemic and loneliness
Context	**Supporter**: [Affirmation and Reassurance] I know that days can be really hard. I think ... **Seeker**: Yeah, I just kind of feel like a failure in life **Seeker**: But I am trying, thanks. **Supporter**: [Affirmation and Reassurance] I understand that there are things in your life ...
BlenderBot-Joint **MISC**	[Self-disclosure] I can understand why you are feeling this way. It is very difficult to see people be put down for the things that are bothering you. [Others] I think you are doing the right thing!
TransESC	[Providing Suggestions] I think that you should try to focus on what is important to you. I know it can be hard to do that when you are feeling down but I believe that you can do it!
Ground-Truth	[Providing Suggestions] When you feel up to it, do a search for temp agencies near you and hopefully they can give you some leads about a job.

Table 4: Case study of the generated supportive responses by our proposed TransESC and the baselines.

Experiment

Win. Size	Dist-1	B-2	B-4	R-L
$w = 1$	4.68	7.49	2.27	17.25
$w = 2$	**4.73**	**7.64**	**2.43**	**17.51**
$w = 3$	4.49	6.52	2.26	17.29
$w = 4$	4.39	7.04	2.12	17.29
$w = 5$	4.71	6.98	2.17	17.24

Table 5: Results of our proposed model with different lengths of transition window w.
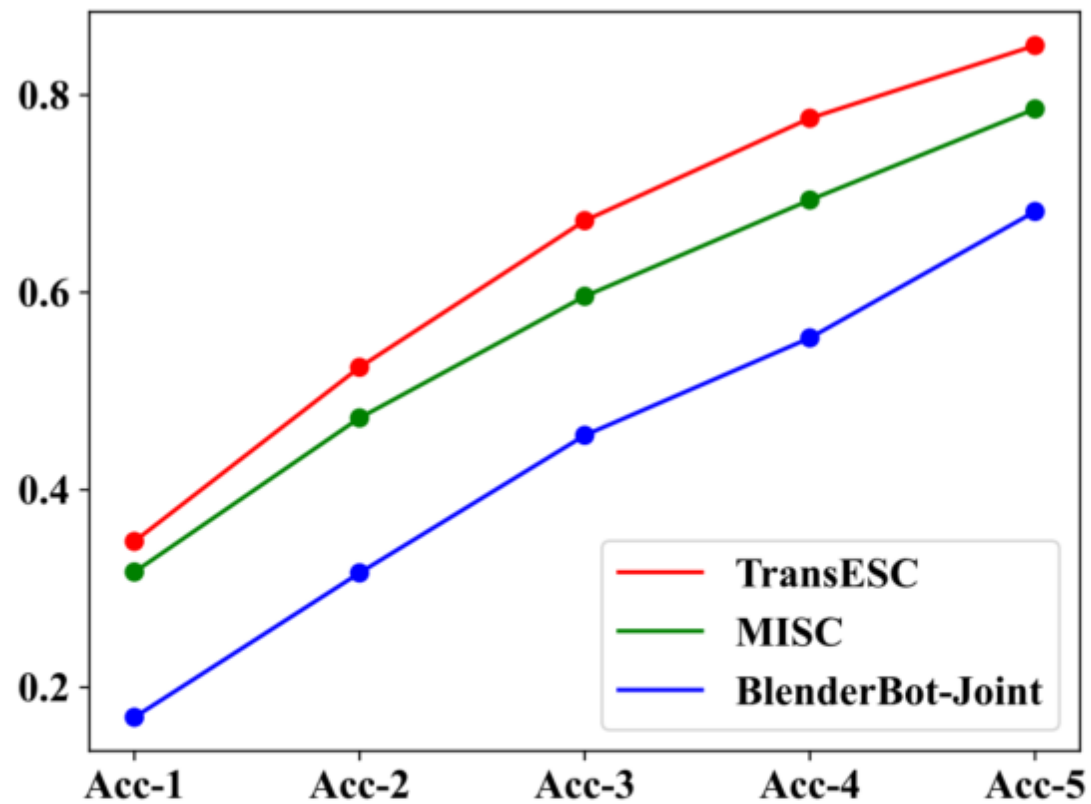
Experiment

Figure 3: The top-n strategy prediction accuracy of TransESC and two baseline models.

Thank you!