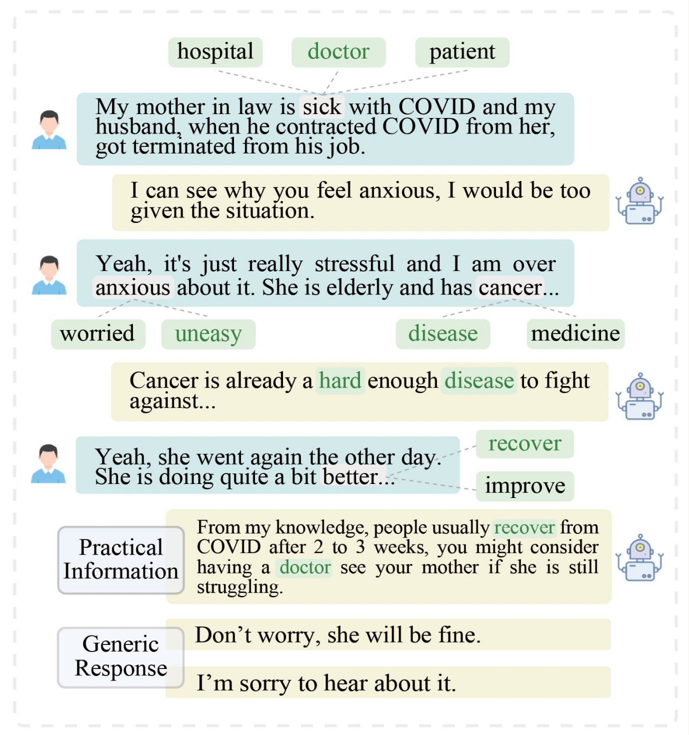
# Knowledge-enhanced Memory Model for Emotional Support Conversation

*Mengzhao Jia, Qianglong Chen, Liqiang Jing, Dawei Fu and Renyu Li.*

# Background

- Mental disorders are known for their high burden, with more than 50% of adults experiencing a mental illness or disorder at some point in their lives; yet despite its high prevalence, only one in five patients receive professional treatment.

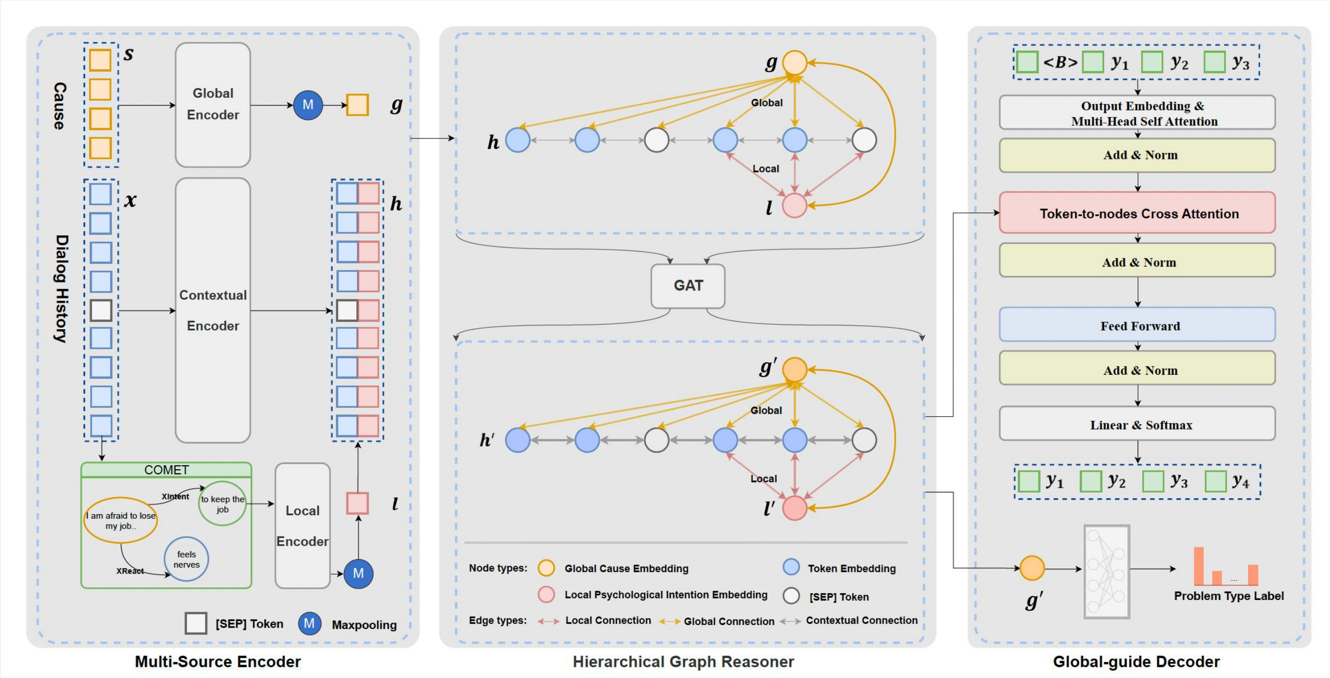
- Emotional Support Conversations (ESConv) has garnered substantial attention in recent years. They have emerged as a promising alternative strategy for mental health intervention.



# Prior research

Prior research primarily concentrated on two aspects:

- Enhance the model’s comprehension of the contextual semantics in the conversations



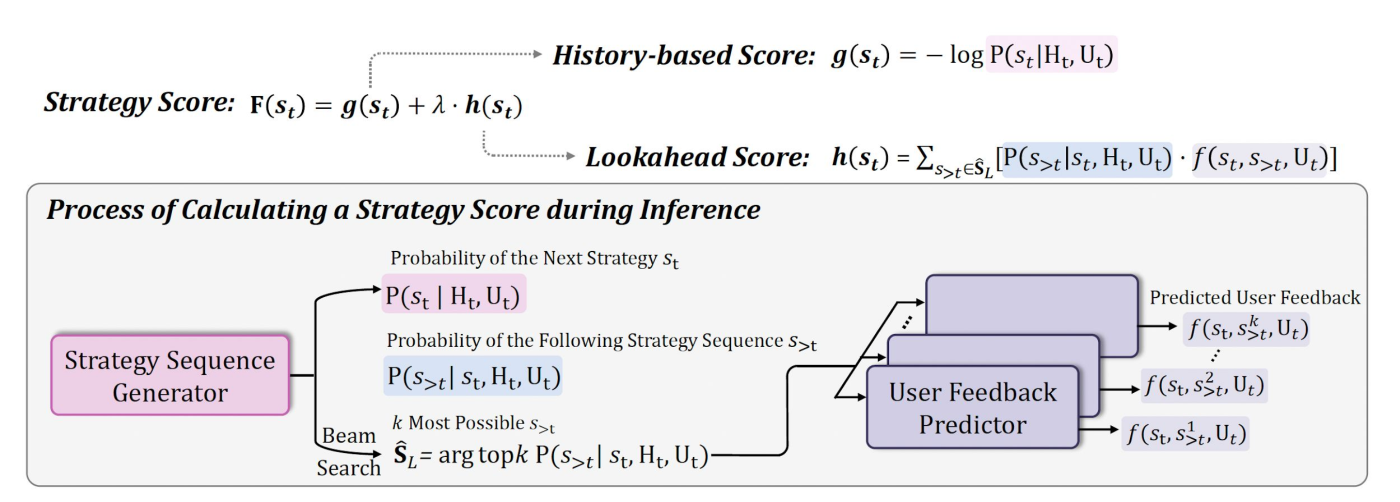

Designed a hierarchical graph network to capture the overall emotional problem cause and specific user intentions.

*Control Globally, Understand Locally: A Global-to-Local Hierarchical Graph Network for Emotional Support Conversation*

# Prior research

Prior research primarily concentrated on two aspects:

- Predict the dialogue strategy accurately and respond based on the predicted strategy category




Employed a lookahead heuristics for dialogue strategy planning and selection

*Improving Multi-turn Emotional Support Dialogue Generation with Lookahead Strategy Planning*

# Challenges

**Variability of emotions.**

As the conversation goes, the user’s emotional state evolves subtly and constantly.
How to model the dynamic emotional change during the dialogueprocess is the first challenge.

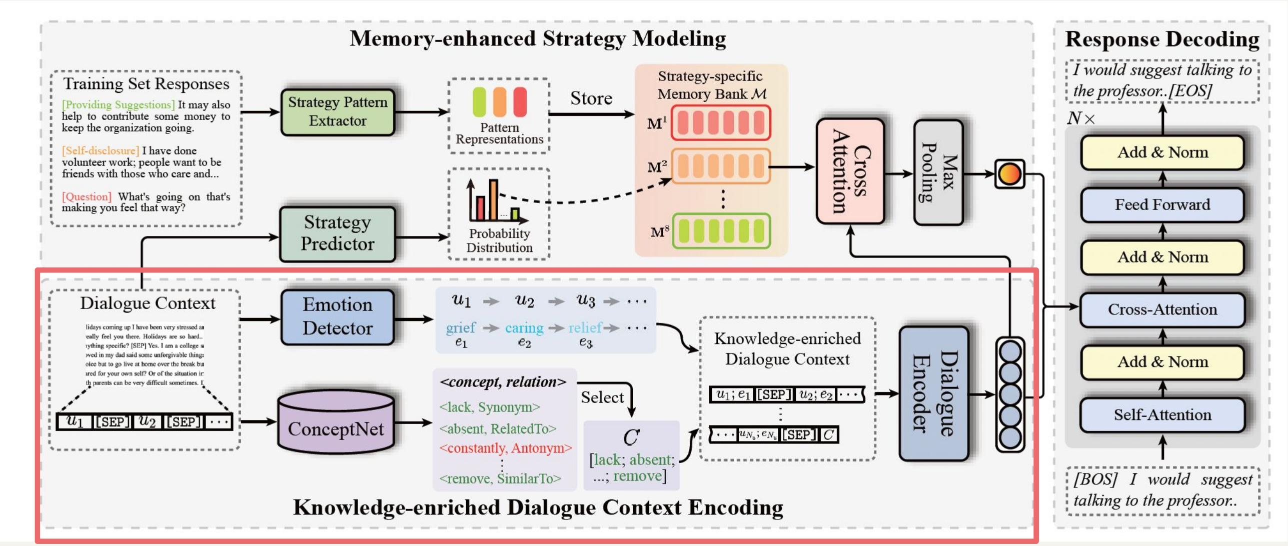
**Practicality of the response**

Dialogue systems are inclined to make generic responses, which are deficient to provide personalized and suitable suggestions.
The integration of appropriate concepts poses a non-trivial challenge in generating practical responses.

**Intricate strategy modeling.**

Dialogue strategy has been reported as a highly complex concept encompassing various intricate linguistic features.
How to model strategy information sufficiently is the third challenge.

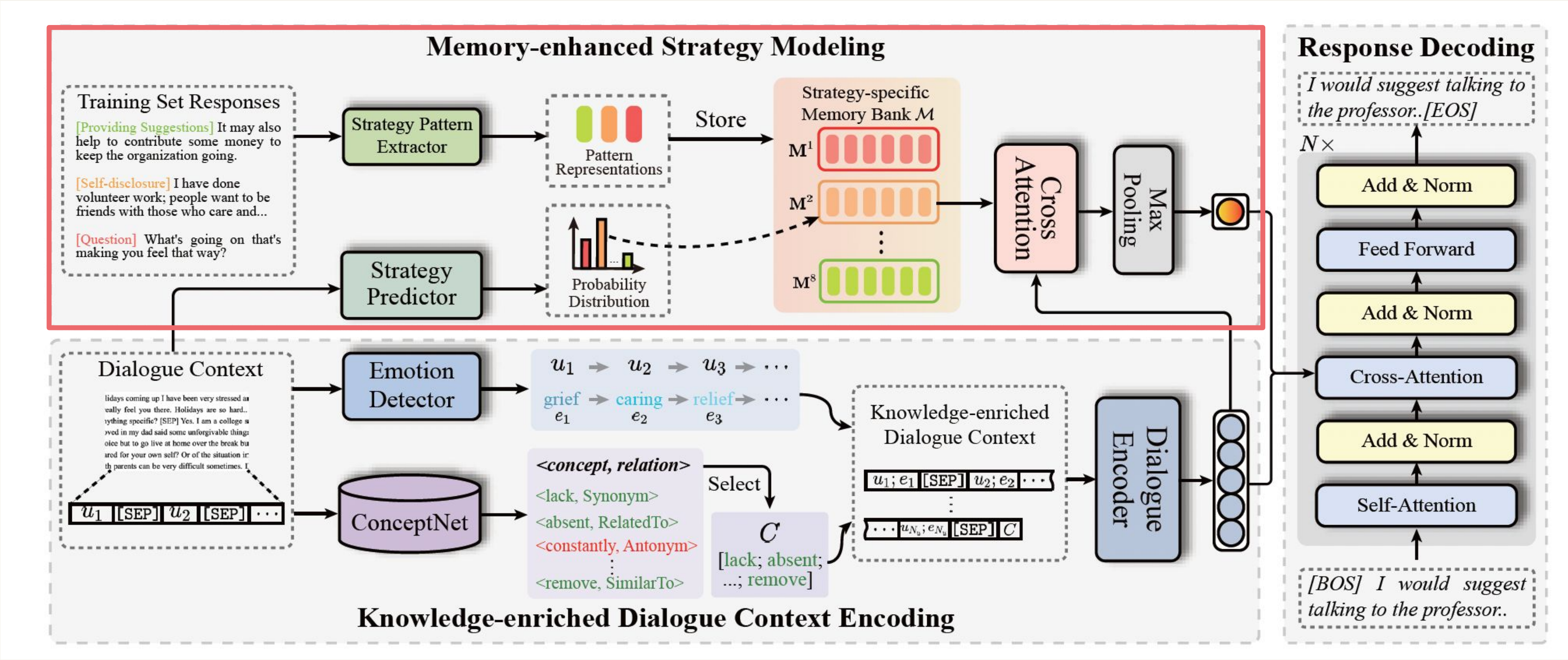
# Method -- MODERN




To capture the emotional change as the conversation progresses, the knowledge-enriched dialogue context encoding module detects the emotions for all utterances and explicitly injects them into the dialogue context as a kind of emotional knowledge.

This module also introduces the concepts reasoning and selection component to acquire valid context-related concepts from a knowledge graph and incorporate them into the dialogue context to fulfill meaningful and practical suggestion generation.

# Method -- MODERN




The memory-enhanced strategy modeling module learns strategy patterns by a strategy-specific memory bank. In this way, it can detect and mimic the intricate patterns in human emotional support strategies.

# Method -- Memory-enhanced strategy modeling module

**Strategy Pattern Modeling**
We first acquire strategy pattern representations of each responses

**Strategy-specific Memory Bank**
we devise a memory bank mechanism to store multiple pattern representations according to their strategy categories.

**Strategy Prediction**
we leverage a strategy predictor, which aims to capture information relevant to strategy decisions in the context

**Memory-enhanced Encoding**
We integrate the aforementioned corresponding memory bank and the context representation, so as to fully exploit the abundant pattern information of he particular strategy.

---

Algorithm 1: Training Procedure.

---

**Input:** training set $\mathcal{P}$ for optimizing the model $\mathcal{F}$, hyperparameters $\{\lambda_1, \lambda_2\}$.
**Output:** Parameters $\mathbf{\Theta}$.

Initialize parameters: $\mathbf{\Theta}$
Initialize memory bank: $\mathcal{M}$
**repeat**
    Randomly sample a batch from $\mathcal{P}$.
    **for** each sample $(\mathcal{D}, R, g, s)$ **do**
        Add strategy pattern representation into the memory bank $\mathcal{M}$ by Eqn. (6).
        Update the memory bank $\mathcal{M}$ by Eqn. (7).
    **end for**
    Update $\mathbf{\Theta}$ by optimizing the loss function in Eqn. (13).
**until** $\mathcal{F}$ converges.

# Experiment Setup – Dataset & Evaluation

**ESConv Dataset.**

- The dataset contains 1,300 long conversations and overall 38,350 utterances, with an average of 29.5 utterances in each dialogue.

- For the comprehensive evaluation, we conducted both automatic and human evaluations.

**Automatic evaluation**

- Metrics: PPL (Perplexity); BLEU-{1,2,3,4}; ROUGE-L; METEOR; and CIDEr.

**Human evaluation**

- Fluency (which response is more fluent, correct, and coherent in grammar and syntax)
- Relevance (which response talks more relevantly regarding current dialogue context)
- Empathy (which response is better to react with appropriate emotion according to the user's emotional state)
- Information (which response provides more suggestive information to help solve the problem)

# Experiment Results – Automatic Evaluations

Table 1: Performance comparison under automatic evaluations. The best results are highlighted in bold.

| Model | PPL | B-1 | B-2 | B-3 | B-4 | R-L | MT | CIDEr |
|---|---|---|---|---|---|---|---|---|
| MoEL | 264.11 | 19.04 | 6.47 | 2.91 | 1.51 | 15.95 | 7.96 | 10.95 |
| MIME | 69.28 | 15.24 | 5.56 | 2.64 | 1.50 | 16.12 | 6.43 | 10.66 |
| DialoGPT-Joint | 15.71 | 17.39 | 5.59 | 2.03 | 1.18 | 16.93 | 7.55 | 11.86 |
| BlenderBot-Joint | 16.79 | 17.62 | 6.91 | 2.81 | 1.66 | 17.94 | 7.54 | 18.04 |
| MISC | 16.16 | - | 7.31 | - | 2.20 | 17.91 | **11.05** | - |
| GLHG | 15.67 | 19.66 | 7.57 | 3.74 | 2.13 | 16.37 | - | - |
| FADO | 15.72 | - | 8.00 | 4.00 | 2.32 | 17.53 | - | - |
| PoKE | 15.84 | 18.41 | 6.79 | 3.24 | 1.78 | 15.84 | - | - |
| MultiESC | 15.41 | 21.65 | 9.18 | 4.99 | 3.09 | 20.41 | 8.84 | 29.98 |
| **MODERN** | **14.99** | **23.19** | **10.13** | **5.53** | **3.39** | **20.86** | 9.26 | **30.08** |

1) MODERN outperforms the baselines in most metrics, which is a powerful proof of the effectiveness of the proposed method.
2) Our MODERN consistently exceeds MultiESC with the same BART backbone. This suggests that BART model with large-scale pretraining still requires strategy pattern information and knowledge.

# Experiment Results – Human Evaluations

Table 2: The human evaluation results in four dimensions.

| Comparisons | Aspects | Win | Tie | Lose |
|---|---|---|---|---|
| vs. FADO | Fluency | 37.0 | 24.0 | 9.0 |
| | Relevance | 35.5 | 20.5 | 14.0 |
| | Empathy | 34.0 | 16.5 | 19.5 |
| | Information | 37.5 | 17.0 | 15.5 |
| vs. MultiESC | Fluency | 30.5 | 24.5 | 15.0 |
| | Relevance | 32.5 | 20.5 | 17.0 |
| | Empathy | 35.0 | 19.5 | 15.5 |
| | Information | 37.5 | 21.0 | 11.5 |

1) MODERN outperforms all baselines across different evaluation metrics,
2) In addition, the number of “Win” cases is the largest for the Information metric compared with other metrics, which demonstrates that integrating context-related concepts can supply meaningful information for emotional support.

# Conclusion & Contributions

1) We first analyze the current challenges of the ESConv task, and according to that propose a novel knowledge-enhanced Memory mODEl for emotional suppoRt coNversation, named MODERN, which can model complex supportive strategy as well as utilize emotional knowledge and context-related concepts to perceive the variability of emotions and provide practical support advice.

2) We propose a memory-enhanced strategy modeling module, where a unique memory bank is designed to model intricate strategy patterns, and an auxiliary strategy classification task is introduced to distill the strategy pattern information.

3) We present a thorough validation and evaluation of our model, providing an in-depth analysis of the results and a comparison with other models. The extensive experiments on the ESConv dataset (Liu et al. 2021a) demonstrate that MODERN achieves state-of-the-art performance under both automatic and human evaluations.

# Thank you
# Q&A