Empathetic Response Generation through Graph-based Multi-hop Reasoning on Emotional Causality

Jiashuo Wang [*], Wenjie Li, Peiqin Lin, Feiteng Mu

Department of Computing, The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

ABSTRACT

Empathetic response generation aims to comprehend the user emotion and then respond to it appropriately. Most existing works merely focus on what the emotion is and ignore how the emotion is evoked, thus weakening the capacity of the model to understand the emotional experience of the user for generating empathetic responses. To tackle this problem, we consider the emotional causality, namely, what feelings the user expresses (i.e., emotion) and why the user has such feelings (i.e., cause). Then, we propose a novel graph-based model with multi-hop reasoning to model the emotional causality of the empathetic conversation. Finally, we demonstrate the effectiveness of our model on EMPATHETICDIALOGUES in comparison with several competitive models.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Empathetic responding, which aims to understand the emotional experience of the user and then reply to acknowledge it appropriately, has benefited a wide range of downstream applications, such as medical dialogue systems [1], counselling conversation [2] and social chatbots [3]. Recent years have witnessed the emergence of empathetic dialogue systems, and some methods [4–6] have been proposed for the task of empathetic response generation and achieved promising results. Specifically, to better understand the emotional experience of the user, MoEL and MIME [4,5] utilize the emotional labels, and EmpDG [6] uses both coarse-grained labels annotated by human and fine-grained emotional words identified by the sentiment lexicons. However, these methods only focus on what the emotion is when encoding the conversation history, which is insufficient to fully understand the emotional experience of the user.

From the aspect of psychology, emotional cause is recognized as an important feature in the analysis of human emotion [7], and itself should be an integral component of emotional experience [8]. Therefore, to understand the emotional experience for generating empathetic responses, it is essential to figure out the forming process (i.e., cause) of the present feelings (i.e, emotion).

To this end, we propose a novel model, named **G**raph-based multi-hop **R**easoning on **E**motional **C**ausality (*GREC* for short), which introduces emotional causality analysis to help understand the emergence of the user's emotion and facilitate empathetic response generation with such knowledge. Specifically, to represent the user's emotional experience embedded in the conversational history, we first construct a series of emotional causality graphs through multi-hop reasoning over the external commonsense graph. Then, we adopt the multi-layer Graph Convolutional Networks (GCN) [9] to model the emotional causality graphs. The encoded graphs are used to serve for the response generation in two ways. On the one hand, the fused graph-aware representation is fed into the decoder to generate a generic vocabulary distribution for the empathetic response. On the other hand, to explicitly exploit the information of emotional causality, multi-hop reasoning on the emotional causality graphs is adopted to obtain the vocabulary distribution over the nodes of the graphs at each step of generation, which affects the final vocabulary distribution.

The main contribution of this paper can be summarized as follows:

(1) It integrates emotional causality reasoning into empathetic response generation, providing an inner and deeper understanding of the user's emotional experience. To the best of our knowledge, this is the first work on the topic of empathetic conversation.

(2) It develops a novel multi-hop graph-based reasoning and generation approach to capture the emotional causal relations among utterances and model the process of emotion forming. Reasoning on graphs improves the explainability of the model to some extent.

(3) It verifies the effectiveness of the proposed model as well as the contribution of each component of the model. The

* Corresponding author.
E-mail addresses: jessiejs.wang@connect.polyu.hk (J. Wang), cswjli@comp.polyu.edu.hk (W. Li), lpq29743@gmail.com (P. Lin), feitengmu.mu@connect.polyu.hk (F. Mu).

experimental results demonstrate that *GREC* with emotional causality reasoning enhances the empathetic expression of replies compared to baselines.

2. Related work

2.1. Empathetic response generation

Recent years have witnessed the emergence of empathetic dialogue systems [10]. Similar to the task of emotional response generation, empathetic response generation also focuses on the emotional expression of users. However, there is a remarkable difference between these two tasks. Emotional response generation aims to generate responses with specific emotions [11–15], while empathetic response generation emphasizes response generation by understanding the emotional experience of the user. Several works have already attempted to make chatbots more empathetic. Specifically, Rashkin et al. have proposed a new dataset and benchmarked several existing models on it [16]. With a transformer-based encoder–decoder architecture, MoEL [4] softly combines different outputs from several emotional decoders and generates the final response. MIME [5] assumes that empathetic responses often mimic the emotions of the speakers to some degree, and its consideration of polarity-based emotion clusters and emotional mimicry shows to improve empathy and contextual relevance of responses. EmpDG [6] exploits both dialogue-level and token-level emotion to capture the nuances of the speaker's emotion. Moreover, it also involves the speaker's feedback to achieve better expression of empathy with an interactive adversarial learning framework. Based on GPT, CAiRE fine-tunes on empathetic dialogue corpus after pretraining a language model with large corpus and multiple objectives [17]. Besides, it is also a promising research direction to involve empathy in various dialogue systems to increase the user experience. For instance, audio-based dialogue systems consider voice tone, intonation, and content of speakers to generate empathetic responses [1,18]. Fung et al. investigated empathetic dialogue for booking systems with a latent variable to store the dialog state [19]. In this paper, we focus on empathetic response generation for text-based chit chat.

2.2. Emotion cause analysis

With deeper exploration of human emotions and feelings, more attention is turned to inner emotional information (e.g., experiences, causes, and consequences of an emotion) [20] from apparent emotional information (e.g., emotion categories) [21, 22]. The stimuli and cause of emotion have been explored in recent ten years in the field of natural language processing. Emotion cause analysis was first proposed as a new task [23] along with a linguistic rule-based method and a manually annotated small scale dataset [24]. After several works [25,26] refines the concept of emotion cause, neural based models were come up with. Gui et al. [27] extracted the emotion cause by considering emotion cause extraction as a reading comprehension task in question-answering. Xia et al. [26] proposed a two-step approach, which first performs individual emotion extraction and cause extraction via multi-task learning, and then conducts emotion-cause pairing and filtering. A RNN-transformer based hierarchical network, named RTHN, was also proposed to extract emotion cause at the level of clause [28]. However, the emotion-cause relation has not been addressed in the existing methods for empathetic response generation. In this paper, we adopt the aforementioned RTHN [28] to discover the emotion cause relations from conversations and leverage this inner emotional evidence to generate empathetic responses.

2.3. Dialogue systems with multi-hop reasoning over commonsense knowledge graphs

Multi-hop reasoning over external knowledge graphs has been proved an effective approach to involve rich semantic information and enable the explainability of models. On the one hand, informative commonsense observations and relations in the structured knowledge boost studies on knowledge fusion between semantic and symbolic space for reasoning [29,30]. On the other hand, paths and flows provide evidences for knowledge selection and deduction, making systems explainable. Therefore, multi-hop reasoning is popular in various downstream tasks. Specifically, KagNet [31], a model based on graph convolutional networks and LSTMs with a hierarchical path-based attention mechanism, is proposed for the task of commonsense question answering. For reading comprehension based QA, a multi-hop pointer-generator model was proposed, along with a selection algorithm to extract knowledge from ConceptNet and an attention mechanism to generate answer [32]. For conversation generation, Liu et al. presented an augmented knowledge graph based open-domain chatting machine [33]. Moreover, multi-hop reasoning has also been used in long text generation, such as story completion, review generation, and description generation [34,35]. Our work preforms multi-hop reasoning over external commonsense knowledge to imitate the processes of emotion formation for better understanding human emotion.

3. Methodology

The overview architecture of our proposed model is shown in Fig. 1. Based on a typical transformer-based Seq2Seq model, we use the emotional causality to augment both the encoder and the decoder with a graph-based approach.

For the encoder, we construct a series of emotional causality graphs for the conversation through multi-hop reasoning over ConceptNet (Section 3.1.1), and then apply a multi-layer GCN encoder to model these emotional causality graphs (Section 3.1.2). Besides, we use a typical transformer encoder for context understanding (Section 3.1.3).

For the decoder, we combine both the sequential information obtained by the transformed-based conversation context encoder and the structural knowledge obtained by the graph-based emotional causality encoder to computes the generic vocabulary distribution (Section 3.2.1). Additionally, a graph-based decoder with multi-hop reasoning computes the distribution over the concepts of the graph to enhance the generic vocabulary distribution for better generating the response (Section 3.2.2).

3.1. Emotional causality augmented encoder

3.1.1. Emotional causality graph construction

To understand the emotional causality of the user's emotional experience in the conversation context, we construct a series of emotional causality graphs based on the given context. Specifically, we first identify the emotion concepts and the cause concepts in the dialogue context, and then reason out the relations between emotion concepts and cause concepts from the external commonsense knowledge graph, namely, Concept-Net, through a multi-hop strategy for constructing the emotional causality graphs.

To obtain the emotion concepts and cause concepts, we need to first identify the emotion clauses and cause clauses of the conversation context. For the emotion clauses, we directly regard the clauses of the immediate utterance $\mathcal{Q}$ as the emotion clauses, since we aim to generate the specific response with empathy for the immediate utterance. Here, we construct multiple emotional
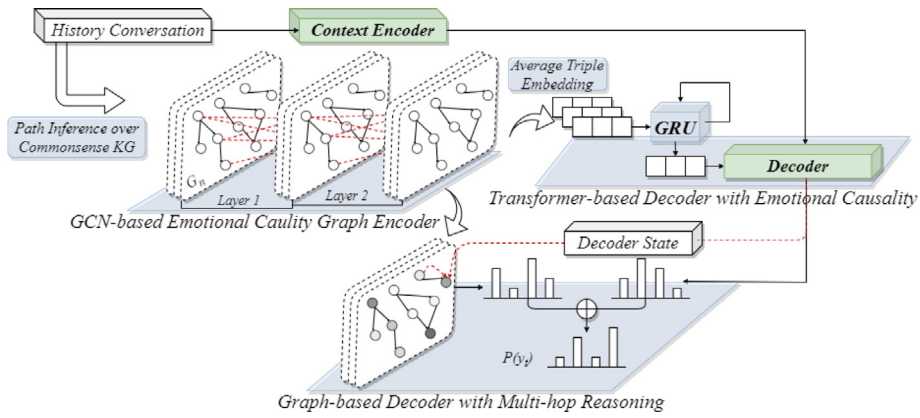
Fig. 1. This is an overview architecture of *GREC*. Apart from the transformer context encoder to encode the conversation context, a GCN encoder encodes the emotional causality graphs after the construction through multi-hop reasoning over the commonsense knowledge graph. At the decoder side, a transformer-based decoder augmented with emotional causality and a graph-based decoder with multi-hop reasoning work together to the response generation.

causality graphs for different clauses of the immediate utterance rather than a single graph for the immediate utterance because there may be different causes of the immediate utterance's emotional clauses. Then, for each emotion clause, we extract its cause clauses from the conversation context by applying an emotion cause extraction model, namely, RTHN [28], which can identify whether a clause is the emotion cause of another clause in a given context. The cause clauses are simply flatted as the cause context of each emotion clause. Given the emotion clauses, we select verbs, nouns, adjectives, and adverbs in both emotion clauses and cause clauses as the emotion concepts and cause concepts, respectively.

To construct the emotional causality graphs based on the emotion concepts and cause concepts, we adopt multi-hop reasoning over ConceptNet to infer semantic paths from emotion concepts to cause concepts, as shown in Fig. 2. The detailed process of deriving intermediate concepts and relations which have the potential to help connect cause concepts with emotion concepts is shown in Algorithm 1. Specifically, we use a concept stack S to collect all candidate concepts to be in the emotional causality graph. The concept stack S is initialized with all cause concepts in $C_{cas}^{(i)}$. At each hop h, we pop concepts from S, and collect their neighbour concepts that have not been passed (line 6–10). Then, we select the top K concepts from all neighbours concepts according to semantic similarity and push them into the concept stack S (line 11–17). Here, we regard the cosine similarity between the word embeddings as the semantic similarity of the two concepts. If there is an emotion concept v_e in the concept stack S, we pop v_e directly without collecting neighbouring concepts of v_e. After H hops, we obtain the emotional causality graph. The intermediate concepts and edges are kept in the emotional causality graph to simulate the behaviour that humans think of various terms in their brains to construct a sentence when chatting.

3.1.2. Graph-based emotional causality encoder

Graph neural networks have been proved effective for producing graph-aware representations of nodes in the graph via integrating neighbours' information of each node [36]. Therefore, to obtain the representation of concepts and relations in the emotional causality graphs, we apply multi-layer GCN [9] encoders to encode the emotional causality graphs. Moreover, following the idea of the TransE model [37], we update a concept embedding with the subtraction between each neighbour concept embedding and the corresponding relation embedding to obtain the relation representation.

Fig. 2. This is an example that illustrates the process to deduce causal paths over an external commonsense graph starting from the cause concept *experience*. *Underlined terms in red* correspond to the cause concepts in the cause context, while *underlined terms in blue* correspond to the emotion concepts in the emotional clause. The *green space* represents a segment of the external commonsense knowledge graph. All the available concepts during this process are maintained to involve the rich semantic information. We preform multi-hop emotional causality inference on clause pair with a cause-emotion relation via a series of similar deduction processes from all cause concepts.

For each emotion clause, there is a constructed emotional causality graph $G = (V, R)$, where V denotes concepts and R denotes relations. The concepts in V are initialized by pretrained word embeddings, and the relations in R are initialized with randomly initialized relation-type embeddings. For each concept v_i, we update its embedding at the $(l + 1)^{th}$ layer by aggregating its neighbours $\mathcal{N}_i$ including pairs of the concept and the relation liking to v_i:

$$h_i^{(l+1)} = \sigma \left(W_s^{(l)} h_i^{(l)} + \sum_{(j,r) \in N_i} \frac{1}{|\mathcal{N}_i|} W_n^{(l)} \left(h_j^{(l)} - h_r^{(l)} \right) \right), \tag{1}$$

Algorithm 1: Framework of Emotional Causality Graph Construction by Multi-hop Reasoning over *ConceptNet*

Input: *ConceptNet*, G_C;
The sets of cause concepts and emotion concepts:
$V_{cas} = \{v_1{}^{cas}, \ldots, v_M{}^{cas}\}$, $V_{emo} = \{v_1{}^{emo}, \ldots, v_N{}^{emo}\}$.
Output: An emotional causality graph, G.

---

1   Initialize an emotional causality graph $G = (v_{cas}, \emptyset)$;
2   Initialize a concept stack S with cause concepts;
3   Initialize an empty temporal triple list T;
4   **for** *h from 1 to H* **do**
5   **while** *S is not empty* **do**
6   pop v_i from S;
7        $\mathcal{N}_i$: the neighbouring concepts of v_i in *ConceptNet*;
8   **foreach** v_j *in* $\mathcal{N}_i$ **do**
9   append (v_i, r, v_j) into T;
10   **end**
11   select the top K tail concepts which is most similar to the concepts in V_{emo} from T;
12   **foreach** (v_i, r, v_j) *in* T **do**
13   add v_j and r into G;
14   **if** v_j *not in* V_{emo} **then**
15   push v_j into S
16   **end**
17   **end**
18   **end**
19   **end**

---

where $h_i^{(l)}$, $h_j^{(l)}$ and $h_r^{(l)}$ are the embeddings of node v_i, node v_j, and the relation between v_i and v_j at layer $(l)^{th}$; $W_s^{(l)}$ and $W_n^{(l)}$ are the two trainable parameter matrices specific to the layer l^{th}; and σ is a non-linear active function. The relation embedding is also updated at the $(l+1)^{th}$ layer via a linear active function:

$$h_r^{(l+1)} = W_R^{(l)} h_r^{(l)}. \tag{2}$$

After L iterations, we are able to obtain $\left\{ h_{v_1}^{(L)}, \ldots, h_{v_{|V|}}^{(L)} \right\}$, a set of concept representations, and $\left\{ h_{r_1}^{(L)}, \ldots, h_{r_{|R|}}^{(L)} \right\}$, a set of relation representations.

3.1.3. Transformer-based conversation context encoder

We utilize a transformer encoder [38,39] for conversation context understanding. Same as previous works [4–6], we flatten all conversation context in C, and then add a special token *CLS* at the beginning of the input, denoting the global memory of the whole sequence:

$$\overline{C} = [CLS; C]. \tag{3}$$

After that, the input tokens are converted to word embeddings. To embody the influence of the emotional cause effect, we incorporate fixed causal embeddings into the input to show the importance of the clause in terms of their frequency as an emotion cause. In addition, we adopt dialogue state embeddings to distinguish situation text, the speaker's utterance and the listener's utterance. The representation of the encoder input is as follows:

$$E^C = emb\left(\overline{C}\right) + emb_{cas}\left(\overline{C}\right) + emb_{utt}\left(\overline{C}\right). \tag{4}$$

Here, emb, emb_{cas}, and emb_{utt} represent the token embedding layer, causal embedding layer, and the dialogue state embedding layer, respectively. Finally, we use Transformer [38] to encode the input.

$$S^C = trs_{enc}\left(E^C\right), \tag{5}$$

where $S^C \in \mathcal{R}^{L_c \times d_{model}}$, and L_C is the length of the input sequence.

To better learn the context representation, we introduce an auxiliary task of emotion prediction, which is also adopted in previous methods [4–6]. Specifically, we use the global memory of the whole sequence S_0^C to predict the emotion label q_e as follows:

$$q = S_0^C, \tag{6}$$

$$q_e \sim P(e|C) = \sigma(W_e q). \tag{7}$$

3.2. Emotional causality augmented decoder

The emotional causality augmented decoder for empathetic response generation consists of two components, namely, a transformer-based decoder involving the overall emotional causality information and a graph-based decoder with multi-hop reasoning on the emotional causality graphs. These two decoders compute a generic vocabulary distribution and a graph-related concept distribution, respectively. And the final vocabulary distribution is a gated combination of these two vocabulary distributions. Under this setting, we can enhance the impact of emotional causality in both implicit and explicit ways.

3.2.1. Transformer-based decoder with emotional causality

To integrate emotional causality with the original transformer-based decoder, we first apply GRU [40] to obtain an emotional causality representation of the overall emotional causality information, and then introduce the emotional causality to the process of generating the generic vocabulary distribution.

For each conversation, there is a sequence of encoded emotional causality graphs. We obtain the representation of each emotional causality graph through average pooling of triple vectors in the graph $(v_i, r, v_j) \in \mathcal{G}_n$, which is

$$\mathcal{G}_n = \sum_{(v_i, r, v_j) \in \mathcal{G}_n} \frac{1}{|\mathcal{G}_n|} \left[h_i^{(L)}; h_r^{(L)}; h_j^{(L)} \right], \tag{8}$$

where $[;]$ represents the concatenation operation. With this sequence of emotional causality graph representations $\{\mathcal{G}_1, \ldots, \mathcal{G}_{N_q}\}$, we use a bidirectional *Gate Recurrent Unit* (GRU) [40] to encode them. That is:

$$\begin{aligned} \overrightarrow{s_n} &= GRU\left(\overrightarrow{s_{n-1}}, \mathcal{G}_n\right), \\ \overleftarrow{s_n} &= GRU\left(\overleftarrow{s_{n-1}}, \mathcal{G}_{N_q-n+1}\right). \end{aligned} \tag{9}$$

Then, we concatenate the forward last hidden state and the backward last hidden state of the GRU cells to combine the forward and backward representations of the emotional causality:

$$s_{N_q} = \left[\overrightarrow{s_{N_q}}; \overleftarrow{s_{N_q}}\right]. \tag{10}$$

Finally, we obtain the emotional causality vector, representing the overall emotional causality, through a linear transformation:

$$\mathcal{H}_{\mathcal{Q}} = \sigma(W_g s_{N_q}), \tag{11}$$

where W_g is the trainable parameter matrix.

Then we integrate the emotional causality vector to the transformer decoder. Similar to the context encoder, we use a special token *SOS* together to represent the universal information of the decoder input $y_{<t}$. Then we append the emotional causality vector $\mathcal{H}_{\mathcal{Q}}$ to this special vector. That is

$$E_t = emb\left([y_{<t}]\right), \tag{12}$$

$$s_t = trs_{dec}([emb(SOS) + \mathcal{H}_{\mathcal{Q}}; E_t], S^C). \tag{13}$$

Finally, we compute the generic vocabulary distribution:

$$P\left(\mathcal{V} \mid \mathbf{s}_t, \mathcal{H}_{\mathcal{Q}}\right) = softmax_{w \in \mathcal{V}}\left(W_{voc} s_t + b\right). \tag{14}$$

3.2.2. Graph-based decoder with multi-hop reasoning

Inspired by the idea of *multi-hop reasoning flow* [34] and *pointer-generator decoder* [32], we adopt a graph-based decoder, which computes the concept distribution with multi-hop reasoning on the emotional causality graphs. At each decoding step, we adjust the weight of each concept in the graph by combining its neighbouring evidences and the current decoder state. Initially, scores of all cause concepts are set to 1, while other concepts are assigned with a score of 0. Then, we broadcast information of scored concepts on the graph to update the unvisited concepts. For an unvisited concept $v_i \in V$, $score(v)$ is computed by aggregating the evidences from its visited neighbouring concepts $\mathcal{N}_i^{in}$:

$$score(v_i) = \sum_{(v_j, r) \in \mathcal{N}_i^{in}} \frac{1}{|\mathcal{N}_i^{in}|} \left(\gamma \cdot score(v_j) + R\left(v_j, r, v_i\right) \right). \qquad (15)$$

Here, γ is a discount factor and $R\left(v_j, r, v_i\right)$ is the triple relevance under the current decoder output, which can be computed as below:

$$h_{(j,r,i)} = \left[h_j^{(L)}; h_r^{(L)}; h_i^{(L)} \right], \qquad (16)$$

$$R\left(v_j, r, v_i\right) = \sigma\left(h_{(j,r,i)} W_{rel} s_t\right). \qquad (17)$$

After L-hop interactions, the distribution over the concepts is as follows:

$$P\left(V \mid \mathbf{s}_t, \mathcal{G}\right) = softmax_{v \in V}\left(score(v)\right). \qquad (18)$$

The final generation distribution conjoins the distribution over the concepts in the emotional causality graphs V and the distribution over the standard vocabulary $\mathcal{V}$ with a soft gate g_t, which decides whether or not to refer to the emotional causality graphs:

$$g_t = \sigma\left(W_g s_t\right), \qquad (19)$$

$$o_t \sim P\left(y_t \mid y_{<t}, C, \mathcal{G}, \mathcal{H}_\mathcal{Q}\right) \begin{aligned} &= g_t \cdot P\left(V \mid \mathbf{s}_t, \mathcal{G}\right) \\ &+ (1 - g_t) \cdot P\left(\mathcal{V} \mid \mathbf{s}_t, \mathcal{H}_\mathcal{Q}\right) \end{aligned}. \qquad (20)$$

3.3. Training

To enhance the empathy of the model, we train the model via multi-task learning with an auxiliary task to predict the emotion label of the speaker. We compute the multi-task loss with the prediction results $q_e \sim P\left(e|C\right)$ which is compute in Eq. (7).

$$L^e = -\sum_{e=1}^{|E|} p_e log\left(q_e\right), \qquad (21)$$

where p_e is the real label probability of the speaker's emotion (1 for the true emotion label, and 0 for the others), and the q_e is the predicted one (Section 3.1.3).

The loss of response generation task is:

$$L^g = -\sum_{t=1}^{|Y|} \sum_{v=1}^{|\mathcal{V}|} p_t^{\,v} log\left(o_t^{\,v}\right), \qquad (22)$$

where Y is the target output, $p_t^{\,v}$ is the true probability of the token w at the position t in Y, and $o_t^{\,v}$ is the predicted one.

The total loss of the model is a weighted sum of these two components through a coefficient λ:

$$Loss = L^g + \lambda L^e. \qquad (23)$$

4. Experiments

4.1. Dataset

In this paper, we utilize EMPATHETICDIALOGUES [16], a dataset of conversations grounded in specific emotional situations to conduct experiments. Table 1 shows an example from EMPATHETICDIALOGUES. This dataset consists of 25k crowdsourced one-on-one conversations between a speaker and a listener. Each dialogue is developed based on an emotion label and a given situation paragraph. There are 32 emotion categories and they are distributed in a balanced way. The speakers talk about their situations and the listeners attempt to understand the speakers' feelings and reply accordingly. At the training time, the emotional labels of the conversations are given in terms of the degree of empathy to facilitate multi-task learning, while the labels are not provided when testing the model and evaluating the generation performance.

4.2. Comparison models

We compare our *GREC* with the following models for empathetic response generation:

Transformer with Multitask (Multi-TRS). This is a universal transformer encoder–decoder structure model [38]. We train this model in a multi-task learning setting, where the loss is calculated by Eq. (21)–(23).

MoEL. The mixture of Empathetic Listeners (MoEL) [4] is a transformer-based model consisting of multiple emotion-specific decoders and a meta decoder. Each emotion-specific decoder is optimized to react to a certain emotion, while the meta decoder integrates evidences from different decoders and generates the final empathetic reply. We implement it using the same parameters and the values reported in [4].

MIME. MIME [5] designs an empathetic response generation model with the consideration of polarity-based emotion clusters and emotional mimicry. It attempts to achieve an appropriate balance of emotions in positive and negative emotion groups. Based on MoEL [4], MIME introduces stochasticity into the emotion mixture for various empathetic responses.

EmpDG. This model [6] exploits both dialogue-level and token-level emotion to capture the nuances of the speaker's emotion. It also involves the speaker's feedback to achieve better expression of empathy with an interactive adversarial learning framework.

4.3. Experiment setup

We use *Pytorch*[1] to implement the proposed model. The word embedding is initialized with pretrained Glove vectors[2] of dimension 300 [41] and shared by the transformer encoder, GCN encoder, and the projection layer of the decoder. For the transformer encoder and decoder, the layer number is set to 6, and the number of attention heads is set to 8. The embedding dimensions of the query, key, and value are set to 40. We replace the position wise feedforward sub-layer with a 1D convolution with 50 filters of width 3. The causal embeddings to denote the importance of the clauses in the conversation context are randomly initialized and fixed during training, and the embedding dimension is set to 300 empirically. For RTHN, which is used to extract the emotional causal relationships from the conversation, we follow the original paper [28] and pretrain an emotion causality detection model based on an annotated relational emotion dataset, namely REMAN[3] [42]. We finally use a *RTHN* model that has F1 of 0.9481. We use Adam for optimization [43]. The batch size is 32 and the initial learning rate is 0.1. The learning rate is updated every 500 steps with a decay rate of 0.1 until the learning rate is 1e-5. The

---

Table 1
An example from EMPATHETICDIALOGUES.

Label: Excited
Situation: My wife and I took a 3 week trip to New Zealand last year.
Conversation:
Speaker: My wife and I decided to go on an international trip last year, and it was my first trip out of the US. It was an amazing experience, and one of the best and most memorable things I have ever done.
Listener: That is awesome! How helpful where the other people so that you could enjoy your trip?
Speaker: My wife was amazing. She booked the whole trip for us, and planned out some things for us to do while we were there.
Listener: Must have been super fun.

beam size is set to 5. For statistical significance tests, we run the model 10 times.

For a fair comparison, the parameter settings and training processes of all state-of-the-art models are the same as suggested in their original papers and implementations.

4.4. Evaluation metrics

We evaluate our models from three aspects, i.e., automatic evaluation, human ratings, and human A/B test.

Automatic Evaluation. We use per-word perplexity (PPL) [44] and BLEU [45] to measure the performance of response generation. PPL measures the confidence to generate the responses. BLEU evaluates the coherency of the generated responses with golden responses [46]. Lower PPL and higher BLEU are expected. We use *multi-bleu.perl*[4] [47] to compute the BLEU scores.

Human Ratings. For qualitative evaluation, we use manual ratings to compare *GREC* and other models. We first randomly sample 120 instances from the test dataset. Then, we score these responses from the following aspects: **empathy** (whether the response shows an understanding of the feelings of the speaker), **relevance** (whether the response is relevant to the conversation) and **fluency** (whether the response is fluent and natural). The scores range from 1 to 5, where 1 is the lowest and 5 is the highest. We evaluate each metric independently.

Human A/B Test. In order to compare the overall performances of different models, we adopt human A/B task. Given two models A and B – *GREC vs. Multi-TRS, MoEL, MIME* in our case, we ask the annotator to pick the model with the better responses. The annotator can select a *Tie* if the responses are both good or both bad.

5. Results and analysis

5.1. Experimental results

The results of the automatic evaluation and human ratings are shown in Table 2. We find that our model outperforms other models in most of the evaluation metrics. Our model achieves the lowest PPL and the highest BLEU among all models, which indicates that responses generated by our model are more human-like. For the human rating results, **fluency** of different models are marginal. However, it is obvious that *GREC* obtains the highest scores in terms of **empathy** and **relevance**. We think this is achieved because of the incorporation of emotional causality. Although the models MoEl, MIME and EmpDG also employ emotion wisely, they merely utilize emotion labels or emotional words to represent the emotion of the speaker. However, emotional causality helps the model to understand how the emotion of the speaker emerges. Emotional causality provides rich semantic relations between the cause and the emotion, which facilitates the

---

Table 2
Results of automatic evaluation and manual evaluation.

Methods	Automatic evaluation		Manual evaluation		
	PPL↓	BLEU↑	Empathy↑	Relevance↑	Fluency↑
Multi-TRS	33.00	2.95	2.90	2.94	3.70
MoEL	35.35	2.64	2.91	2.91	3.77
MIME	33.05	2.72	3.06	3.13	**3.79**
EmpDG	56.00	1.70	3.05	2.87	3.68
GREC (Ours)	**32.66**	**3.16****	**3.21***	**3.32***	3.78

Asterisk symbols represent statistical significance according to student t-test the best comparison model, with $p < 0.05(*)$ and $p < 0.01(**)$, respectively.

Table 3
Human A/B test results for *GREC vs. Multi-TRS, MoEL and MIME*.

GREC vs.	GREC wins	GREC loses	Tie
Multi-TRS	50.0%	22.5%	27.5%
MoEL	48.3%	20.8%	30.8%
MIME	35.0%	27.5%	37.5%
EmpDG	62.7%	11.7%	25.6%

understanding of the speaker's emotion. The multi-hop reasoning also emphasizes relevant information in the emotional causality graphs. Therefore responses generated by *GREC* are more empathetic responses and relevant.

From Table 3, responses generated by *GREC* are more often preferred by human judges in general than the responses from *Multi-TRS*, *MoEL*, and *MIME*. This also indicates that our model is able to generate more empathetic and more relevant responses.

5.2. Analysis and discussion

5.2.1. Ablation study

To verify the effect and the contribution of each component of *GREC*, we conduct the ablation study. We consider the following three settings.

(1) **w/o graph**. We remove the graph structure of emotional causality, and involve emotional causality in the format of text. Hence, we delete the emotional causality construction (Section 3.1.1). For emotional causality understanding, we use a sequence of emotional cause clause contexts extracted from the conversation context, and apply RNNs to encode them instead of the GCN encoder in Section 3.1.2. We retain the implicit and explicit use of emotional causality when decoding. Instead of the concept distribution computation in Eq. (18), we compute the weight of the emotion concepts and the cause concepts $\mathcal{V}$ by

$$P(V \mid \boldsymbol{s}_t, \mathcal{V}) = softmax_{w \in \mathcal{V}}(emb(w)s_t + b). \tag{24}$$

(2) **w/o implicit decoding**. We ablate the implicit use of the emotional causality by deleting the involvement of the emotional causality in the transformer decoder in Section 3.2.1 (i.e., *w/o encoder*). We replace Eq. (13) by

$$s_t = trs_{dec}([emb(SOS); E_t], S^C). \tag{25}$$

Table 4
Results of automatic evaluation and manual evaluation for *Hierarchical Emotional Causal RNN*.

Methods	Automatic evaluation		Manual evaluation		
	PPL↓	BLEU↑	Empathy↑	Relevance↑	Fluency↑
GREC	32.66	3.16	3.21	3.32	3.78
w/o graph	33.93	2.20	2.92	2.94	3.75
w/o implicit decoding	33.50	3.05	3.13	3.17	3.67
w/o explicit reasoning	32.94	2.81	3.02	3.02	3.85

(3) **w/o explicit reasoning**. We ablate the explicit use of the emotional causality by removing the graph-based decoder with multi-hop reasoning in Section 3.2.2. The final vocabulary distribution is calculated by Eq. (14) only.

The results in Table 4 show that all components contribute to *GREC* because removing any of them decreases the overall performance. Moreover, each component contributes to *GREC* from different aspects. Based on the observation and the design of our model, we have the following analysis.

(1). Emotional causality graphs contribute most to our model. The performance drops in both **empathy** and **relevance** significantly by 9.03% and 11.45% respectively in the absence of the graph structure. Graph structures with external commonsense knowledge can provide a better representation of the emotional causality. Although the leverage of emotional cause clause can also provide evidence for emotion understanding, which outperforms Multi-TRS and MoEL in the aspect of **empathy** and **relevance**, the semantic relationship between the cause and emotion is more abundant in the emotional causality graphs. Besides, emotional causality graphs involve more information for the decoder with multi-hop reasoning.

(2). Decoding with the emotional causality in an implicit way also contributes to our model. Compared to the other two ablation models, the performance is decreased less in most of the evaluation metrics, however, the **fluency** decreases significantly. On the one hand, the transformer decoder with emotional causality is an implicit mechanism to use emotional causality, therefore it cannot change the performance remarkably. On the other hand, it provides a chance where the emotional causality can influence the computation of generic vocabulary distribution, making the involvement of emotional causality more natural.

(3). Graph-based decoding with multi-hop reasoning also has a great contribution to *GREC*. This component reasons out important information from the emotional causality graphs, which increases empathetic and relevant expression. However, the graph-based decoder with multi-hop reasoning has a negative influence on **fluency**. One reasonable explanation is that emotional causality graphs contain not only useful information, but also noises, which decrease the **fluency** of the responses. In future work, it is a meaningful direction to attend more on the useful evidences, and reduce the influence of noises in the graphs.

5.2.2. Case study
We provide some cases to show the effectiveness of *GREC*. Table 5 shows two instances from the test data. From the results, *GREC* can capture the emotion of the speaker more accurately, and respond in a more natural way, compared with other models. In both these two conversations, *GREC* generates responses with appropriate emotion and expression. For example, it can use the verb "get it back" and "complain" as the response to a disgusting attitude. Such expression is pertinent and specific in this conversation, showing a deeper understanding of the speaker's emotion. However, other models prefer to using emotional words to express the emotion, which is not natural and can be used in a wide range of scenarios. It can also capture the concepts that

bring the emotion of the speaker. In the second conversation, *GREC* talks about "shadow", which is the core content of this conversation. Besides, *GREC* usually performs better in or after the second turn of a conversation compared to other models, which is also reflected in these two cases. We believe that the emotional causality graphs carry important information in the conversation context, and allow such information to have an impact on the current turn.

To figure out how the information in the emotional causality helps the empathetic response generation, we display the scores of the concepts in the emotional causality graph when generating a sentence. Fig. 3 shows an emotional causality graph for the second utterance of the speaker. The related concepts, which are suitable for this conversation, are inferred and selected. Moreover, the attention weights of the concepts varying over the decoding steps are reflected in this figure. The attention weights for the trivial terms, e.g., "war" and "storm", decrease, while weights for necessary terms, such as "terrified" and "darkness", increase over steps. In addition, *"terrified"* is even selected in the response. This indicates that the model can reason out important information from the emotional causality graphs for empathetic response generation through a multi-hop strategy. Besides, the concepts with higher scores are deduced with flows from the cause concepts. This indeed enhances model transparency and explainability.

6. Conclusions

In this work, we incorporate the emotional causality into empathetic response generation. To implement this idea, we propose a novel graph-based multi-hop reasoning model, called *GREC*. Instead of simple employment of the superficial emotional information, such as emotion labels and emotional words, *GREC* explores how the emotion emerges and provides a deeper understanding of the speaker's emotion. With a typical transformer Seq2Seq model as the skeleton, we augment the encoder and decoder with emotional causality. For emotional causality representation and understanding, we construct emotional causality graphs through multi-hop reasoning over ConceptNet, and apply a multi-layer GCN to encode them. For empathetic response generation, we involve the emotional causality information into the transformer decoder and design a graph-based decoder with multi-hop reasoning on the emotional causality graphs. We conduct experiments on a crowdsourced dataset, i.e., EMPATHETICDIALOGUES, and empirically prove that *GREC* outperforms competitive models including those state-of-the-arts. Additionally, we show that graph-based reasoning provides better model transparency and explainability in the ablation study and the case study.

CRediT authorship contribution statement

Jiashuo Wang: Conceptualization, Methodology, Software, Writing – original draft. **Wenjie Li:** Supervision, Writing – review & editing. **Peiqin Lin:** Writing – review & editing. **Feiteng Mu:** Performance evaluation, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table 5
Responses generated by different models. These utterances are from continuous dialogues. The gold responses are also displayed. Terms showing empathy are labelled in *green*.

Conversation 1	
Label: DISGUSTED	
Situation: I once went to a pizza place and found a hair on my pizza. I was grossed out.	
Speaker	I had pizza at a restaurant a couple days ago and found a hair on my pizza!
Gold	It is *disgusting*. What did you do?
Transformer	What did you do?
MoEL	What did you do?
MIME	What did you eat?
EmpDG	That is a little.
GREC	That is so *gross*! Did you get it back?
Speaker	I told the waiter and demanded a refund! It was gross.
Gold	You did good. You should have called the manager to tell him know.
Transformer	Wow, that is *gross*. I hope you were able to fix it .
MoEL	Oh wow, that must have been *disgusting*!
MIME	That is a good idea.
EmpDG	Oh no! I would have been too! I love it!
GREC	It is so *gross*. Did you *complain* to the manager?
Conversation 2	
Label: TERRIFIED	
Situation: One night my children and I came home, and I saw a shadow inside the home when I came inside. My heart jumped into my throat and you could hear my scream in the next town for fear!	
Speaker	One night when my children and I came home, I saw a shadow inside! I was so fearful and afraid!
Gold	Oh that can be *scary*. I do not watch horror movies because I know that I will do just that!
Transformer	*Terrified*, oh no! I am so sorry to hear that.
MoEL	Oh wow, that must have been *scary*!
MIME	Oh no! That is *scary*! I am glad you are ok!
EmpDG	Oh no, did you go to?
GREC	Oh no! I *hate* when that happens to me. What was it?
Speaker	It was so scary! Luckily, it was actually my shadow! The light made my shadow show up, but it first appeared to be an intruder!
Gold	Well I am glad you were ok! Sometimes it is not just a shadow.
Transformer	*Terrified*, I am sure it was a *scary movie*.
MoEL	I would have been so *scared*.
MIME	Oh man! That is *scary*! I am glad you are ok!
EmpDG	Oh no. I would have been too. I would have been too. I would have been too.
GREC	I can imagine. I would have been *terrified*.

I can imagine. I would   generating the 7th token "have"   generating the 8th token "been"   generating the 9th token "terrified"   .<EOS>
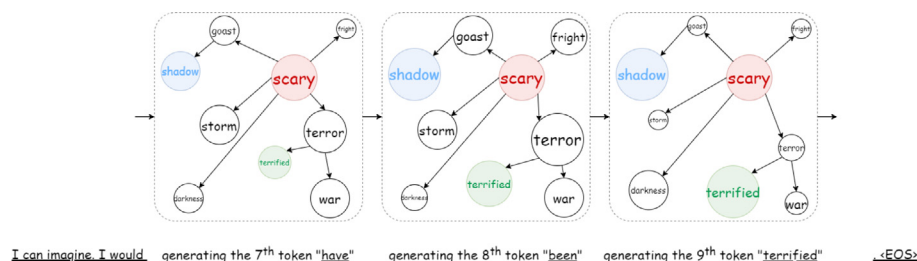
Fig. 3. This is a piece from the emotional causality graph for "*It was so scary …an intruder!*", which shows the attention weights of the concepts vary over the decoding steps. *Red circles* represent the cause concept, *blue circles* represent the emotion concept, and *green circles* represent the concept showing in the final generated response. The size of the circle corresponds to the attention weight for the concept in the current decoding step.

References

[1] H. Eide, R. Frankel, A.C.B. Haaversen, K.A. Vaupel, P.K. Graugaard, A. Finset, Listening for feelings: identifying and coding empathic and potential empathic opportunities in medical dialogues, Patient Educ. Couns. 54 (3) (2004) 291–297.

[2] V. Pérez-Rosas, R. Mihalcea, K. Resnicow, S. Singh, L. An, Understanding and predicting empathic behavior in counseling therapy, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Vancouver, Canada, 2017, pp. 1426–1435, http://dx.doi.org/10.18653/v1/P17-1131.

[3] L. Zhou, J. Gao, D. Li, H.-Y. Shum, The design and implementation of XiaoIce, an empathetic social chatbot, Comput. Linguist. 46 (1) (2020) 53–93, http://dx.doi.org/10.1162/coli_a_00368.

[4] Z. Lin, A. Madotto, J. Shin, P. Xu, P. Fung, MoEL: Mixture of empathetic listeners, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP, Association for Computational

Linguistics, Hong Kong, China, 2019, pp. 121–132, http://dx.doi.org/10.18653/v1/D19-1012.

[5] N. Majumder, P. Hong, S. Peng, J. Lu, D. Ghosal, A. Gelbukh, R. Mihalcea, S. Poria, MIME: MIMicking emotions for empathetic response generation, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, Association for Computational Linguistics, Online, 2020, pp. 8968–8979, http://dx.doi.org/10.18653/v1/2020.emnlp-main.721.

[6] Q. Li, H. Chen, Z. Ren, P. Ren, Z. Tu, Z. Chen, EmpDG: Multi-resolution interactive empathetic dialogue generation, in: Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 4454–4466, http://dx.doi.org/10.18653/v1/2020.coling-main.394.

[7] A. Wierzbicka, Emotions Across Languages and Cultures: Diversity and Universals, Cambridge University Press, 1999.

[8] W. James, What is an emotion? Mind os-IX (34) (1884) 188–205, http://dx.doi.org/10.1093/mind/os-IX.34.188.

[9] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings, OpenReview.net, 2017.

[10] Y. Ma, K.L. Nguyen, F.Z. Xing, E. Cambria, A survey on empathetic dialogue systems, Inf. Fusion 64 (2020) 50–70.

[11] W. Shi, Z. Yu, Sentiment adaptive end-to-end dialog systems, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 1509–1519, http://dx.doi.org/10.18653/v1/P18-1140.

[12] P. Zhong, D. Wang, C. Miao, An affect-rich neural conversational model with biased attention and weighted cross-entropy loss, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 7492–7500.

[13] H. Zhou, M. Huang, T. Zhang, X. Zhu, B. Liu, Emotional chatting machine: Emotional conversation generation with internal and external memory, in: S.A. McIlraith, K.Q. Weinberger (Eds.), Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th Innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2–7, 2018, AAAI Press, 2018, pp. 730–739.

[14] Z. Song, X. Zheng, L. Liu, M. Xu, X. Huang, Generating responses with a specific emotion in dialog, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 3685–3695, http://dx.doi.org/10.18653/v1/P19-1359.

[15] X. Zhou, W.Y. Wang, MojiTalk: Generating emotional responses at scale, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 1128–1137, http://dx.doi.org/10.18653/v1/P18-1104.

[16] H. Rashkin, E.M. Smith, M. Li, Y.-L. Boureau, Towards empathetic open-domain conversation models: A new benchmark and dataset, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 5370–5381, http://dx.doi.org/10.18653/v1/P19-1534.

[17] Z. Lin, P. Xu, G.I. Winata, F.B. Siddique, Z. Liu, J. Shin, P. Fung, CAiRE: An empathetic neural chatbot, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 13622–13623.

[18] M.I. Torres, J.M. Olaso, N. Glackin, R. Justo, G. Chollet, A spoken dialogue system for the empathic virtual coach, in: 9th International Workshop on Spoken Dialogue System Technology, Springer, 2019, pp. 259–265.

[19] P. Fung, D. Bertero, P. Xu, J.H. Park, C.-S. Wu, A. Madotto, Empathetic dialog systems, in: The International Conference on Language Resources and Evaluation. European Language Resources Association, 2018.

[20] C.O. Alm, D. Roth, R. Sproat, Emotions from text: Machine learning for text-based emotion prediction, in: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Vancouver, British Columbia, Canada, 2005, pp. 579–586.

[21] S. Aman, S. Szpakowicz, Identifying expressions of emotion in text, in: International Conference on Text, Speech and Dialogue, Springer, 2007, pp. 196–205.

[22] R.K. Bakshi, N. Kaur, R. Kaur, G. Kaur, Opinion mining and sentiment analysis, in: 2016 3rd International Conference on Computing for Sustainable Global Development, IEEE, 2016, pp. 452–455.

[23] S.Y.M. Lee, Y. Chen, C.-R. Huang, A text-driven rule-based system for emotion cause detection, in: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Association for Computational Linguistics, Los Angeles, CA, 2010, pp. 45–53.

[24] S.Y.M. Lee, Y. Chen, S. Li, C.-R. Huang, Emotion cause events: Corpus construction and analysis, in: Proceedings of the Seventh International Conference on Language Resources and Evaluation, European Language Resources Association (ELRA), Valletta, Malta, 2010.

[25] L. Gui, D. Wu, R. Xu, Q. Lu, Y. Zhou, Event-driven emotion cause extraction with corpus construction, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Austin, Texas, 2016, pp. 1639–1649, http://dx.doi.org/10.18653/v1/D16-1170.

[26] R. Xia, Z. Ding, Emotion-cause pair extraction: A new task to emotion analysis in texts, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 1003–1012, http://dx.doi.org/10.18653/v1/P19-1096.

[27] L. Gui, J. Hu, Y. He, R. Xu, Q. Lu, J. Du, A question answering approach to emotion cause extraction, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2017, pp. 1593–1602.

[28] R. Xia, M. Zhang, Z. Ding, RTHN: A RNN-transformer hierarchical network for emotion cause extraction, in: S. Kraus (Ed.), Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10–16, 2019, ijcai.org, 2019, pp. 5285–5291, http://dx.doi.org/10.24963/ijcai.2019/734.

[29] S. Ji, S. Pan, E. Cambria, P. Marttinen, P.S. Yu, A survey on knowledge graphs: Representation, acquisition, and applications, IEEE Trans. Neural Netw. Learn. Syst. (2021) 1–21, http://dx.doi.org/10.1109/TNNLS.2021.3070843.

[30] H. Zhou, T. Young, M. Huang, H. Zhao, J. Xu, X. Zhu, Commonsense knowledge aware conversation generation with graph attention, in: J. Lang (Ed.), Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13–19, 2018, Stockholm, Sweden, ijcai.org, 2018, pp. 4623–4629, http://dx.doi.org/10.24963/ijcai.2018/643.

[31] B.Y. Lin, X. Chen, J. Chen, X. Ren, KagNet: Knowledge-aware graph networks for commonsense reasoning, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP, Association for Computational Linguistics, Hong Kong, China, 2019, pp. 2829–2839, http://dx.doi.org/10.18653/v1/D19-1282.

[32] L. Bauer, Y. Wang, M. Bansal, Commonsense for generative multi-hop question answering tasks, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 4220–4230, http://dx.doi.org/10.18653/v1/D18-1454.

[33] Z. Liu, Z.-Y. Niu, H. Wu, H. Wang, Knowledge aware conversation generation with explainable reasoning over augmented graphs, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP, Association for Computational Linguistics, Hong Kong, China, 2019, pp. 1782–1792, http://dx.doi.org/10.18653/v1/D19-1187.

[34] H. Ji, P. Ke, S. Huang, F. Wei, X. Zhu, M. Huang, Language generation with multi-hop reasoning on commonsense knowledge graph, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP, Association for Computational Linguistics, Online, 2020, pp. 725–736, http://dx.doi.org/10.18653/v1/2020.emnlp-main.54.

[35] L. Zhao, J. Xu, J. Lin, Y. Zhang, H. Yang, X. Sun, Graph-based multi-hop reasoning for long text generation, 2020, CoRR abs/2009.13282, arXiv:2009.13282.

[36] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun, Graph neural networks: A review of methods and applications, AI Open 1 (2020) 57–81.

[37] A. Bordes, N. Usunier, A. García-Durán, J. Weston, O. Yakhnenko, Translating embeddings for modeling multi-relational data, in: C.J.C. Burges, L. Bottou, Z. Ghahramani, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a Meeting Held December 5–8, 2013, Lake Tahoe, Nevada, United States, 2013, pp. 2787–2795.

[38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U. von Luxburg, S. Bengio, H.M. Wallach, R. Fergus, S.V.N. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4–9, 2017, Long Beach, CA, USA, 2017, pp. 5998–6008.

[39] M. Dehghani, S. Gouws, O. Vinyals, J. Uszkoreit, L. Kaiser, Universal transformers, in: 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6–9, 2019, OpenReview.net, 2019.

[40] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder–decoder for statistical machine translation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1724–1734, http://dx.doi.org/10.3115/v1/D14-1179.

[41] J. Pennington, R. Socher, C. Manning, GloVe: Global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1532–1543, http://dx.doi.org/10.3115/v1/D14-1162.

[42] E. Kim, R. Klinger, Who feels what and why? Annotation of a literature corpus with semantic roles of emotions, in: Proceedings of the 27th International Conference on Computational Linguistics, Association for Computational Linguistics, Santa Fe, New Mexico, USA, 2018, pp. 1345–1359.

[43] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: Y. Bengio, Y. LeCun (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings, 2015.

[44] L.R. Bahl, F. Jelinek, R.L. Mercer, A maximum likelihood approach to continuous speech recognition, IEEE Trans. Pattern Anal. Mach. Intell. (2) (1983) 179–190.

[45] K. Papineni, S. Roukos, T. Ward, W.-J. Zhu, Bleu: a method for automatic evaluation of machine translation, in: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, 2002, pp. 311–318, http://dx.doi.org/10.3115/1073083.1073135.

[46] C.-W. Liu, R. Lowe, I. Serban, M. Noseworthy, L. Charlin, J. Pineau, How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Austin, Texas, 2016, pp. 2122–2132, http://dx.doi.org/10.18653/v1/D16-1230.

[47] D. Britz, A. Goldie, M.-T. Luong, Q. Le, Massive exploration of neural machine translation architectures, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Copenhagen, Denmark, 2017, pp. 1442–1451, http://dx.doi.org/10.18653/v1/D17-1151.