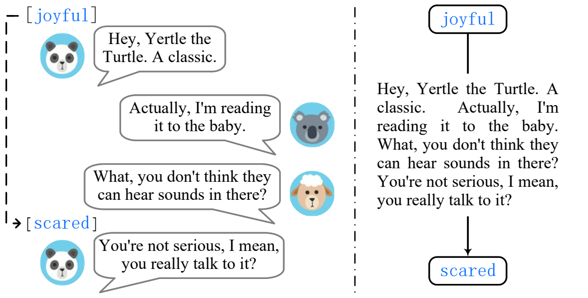
EmoTransKG: An Innovative Emotion Knowledge Graph to Reveal Emotion Transformation

Huan Zhao, Xupeng Zha*, Zixing Zhang*
College of Computer Science and Electronic Engineering, Hunan University, China
zhaxupeng@hnu.edu.cn

Abstract

This paper introduces EmoTransKG, an innovative Emotion Knowledge Graph (EKG) that establishes connections and transformations between emotions across diverse open-textual events. Compared to existing EKGs, which primarily focus on linking emotion keywords to related terms or on assigning sentiment dimension ratings to emotion words by humans, EmoTransKG aims to represent the general knowledge involved in emotion transformations. Specifically, in conversations, successive emotions expressed by a single speaker are temporally considered as the head and tail entities, with open-text utterances (events) occurring between them representing the relation. To explore the knowledge of emotion transformations described in EmoTransKG, we develop a Transformer-based translational model called EmoTransNet, which predictively trains tail entities by interpreting the relation as an operation that transforms the source emotion into the target emotion. Particularly, our designed EmoTransNet serves as a plug-in module that seamlessly integrates with any conversational emotion recognition (CER) models for emotion retrofitting. Experimental results on two CER datasets demonstrate that the incorporation of EmoTransNet with baseline models results in substantial improvements, and the qualitative visualization of entities and relations clearly clarify their unique roles in emotion transformations. These experiments confirm the quality and effectiveness of EmoTransKG.[1]



(a) A segment of conversation (b) An emotion triple graph

Figure 1: An authentic example of a conversation segment from EmoryNLP dataset (Zahiri and Choi, 2018) and its corresponding emotion triple: the head entity is `joyful`, the relation is "Hey, Yertle the Turtle. A classic. Actually, I'm reading it to the baby. What, you don't think they can hear sounds in there? You're not serious, I mean, you really talk to it?" and the tail entity is `scared`. Both `joyful` and `scared` are conveyed by the same speaker.

1 Introduction

In recent years, Knowledge Graphs (KGs)[2], such as WordNet (Fellbaum, 1998), Google KG (Singhal, 2012), or ConceptNet 5.5 (Speer et al., 2017), have emerged as influential repositories and conveyors of knowledge, with a growing expectation to endow computers with human-like cognitive abilities by providing a wealth of world knowledge, including general commonsense knowledge (Liu and Singh, 2004; Matuszek et al., 2006) and emotional knowledge (Bradley and Lang, 1999; Stevenson et al., 2007). Unlike commonsense knowledge, emotions are abstract concepts that resist direct quantification, characterized by fuzzy boundaries and significant variation in personal expression and experience, making the construction of the emotion knowledge graph (EKG) an inherent challenge.

Emotions and sentiments are pervasive in text, which refers to written language and transcriptions of communication. Currently, the development of EKGs relies heavily on these textual sources. In written language, where words serve as the smallest utterances of meaning in a language, many projects strive to rate word lexicons[3] (Bradley and Lang,

*Corresponding authors.

[1]Available at https://github.com/XP-ZHA/EmoTransKG

[2]In 2012, Google introduced the "Knowledge Graph", emphasizing its structure around vertices and edges. In the context of knowledge graphs, any knowledge resource that can be converted into such a structure is viewed as a knowledge graph.

[3]The word lexicon can be converted into an EKG by linking words and their annotations through the use of the `RelatedTo` relation.

Findings of the Association for Computational Linguistics: ACL 2024, pages 12098–12110
August 11-16, 2024 ©2024 Association for Computational Linguistics

1999; Warriner et al., 2013; Mohammad, 2018a) in a fine-grained dimensional space of emotion, such as the three-dimensional VAD (Valence, Arousal, and Dominance) space (Osgood et al., 1957; Russell, 2003), arguing that individual emotions can be effectively conveyed through word meanings. However, word-based EKGs have limitations, including inconsistent annotations from different annotators, challenges associated with fixed granularity, and rating scales that may inadequately capture the intended sentiment. Additionally, there are text-based EKGs (Gill et al., 2008) that involve semantic analysis of context, going beyond simple word mapping. One such example is COMET (Bosselut et al., 2019), which expresses individual intentions and reactions toward textual events. Transcriptions of communication are mainly used for tasks related to emotion recognition or analysis, such as Conversational Emotion Recognition (CER) (Zhang et al., 2019), however, the study of how to make them work in EKGs is still open. Furthermore, these efforts concentrate on the expression of emotions rather than their transformations or connections.

Emotion transformation, as outlined by Greenberg (2002), is a process of *changing emotions with emotions* only through new experiences. Natural language emotion transformation tasks aim to rephrase input sentences to satisfy a given affective label. Some approaches (Helbig et al., 2020) use word-emotion lexicons, like WordNet (Fellbaum, 1998), to reverse the sentiment polarity of the text through word substitution techniques. Recent studies (Luo et al., 2019; Shen et al., 2017; Li et al., 2018) suggest that deep learning approaches, especially encoder-decoder models, have the potential to generate fluent and natural language. Nonetheless, the resulting text often lacks contextual consistency due to the absence of parallel data for supervision. Essentially, these approaches focus on altering the emotions expressed in the text, rather than exploring emotion transformation and its fundamental patterns and regularities. Access to emotion transformation holds benefits for a number of fields, including psychology (e.g., developing interventions for emotion regulation), natural language processing (e.g., understanding the interplay of language and emotion to enhance human-computer interactions), and neuroscience (e.g., understanding the dynamic mapping of neural circuitry underlying emotion experience).

In this paper, we describe how we construct and evaluate EmoTransKG, an emotion knowledge graph designed to address the aforementioned gap in emotion transformation using communication transcripts. Specifically, we collect real-world conversation segments in which a speaker delivers initial and final utterances (referred to as "events") that are manually annotated with emotion keywords (e.g., happy, sad, and angry). For each segment, we define the emotions annotated in the one utterance and its successive one as head and tail entities respectively, and all textual utterances (a.k.a. emotional events (Scherer, 1993)) within them as a relation, resulting in the creation of an emotion knowledge triple. This construction is inspired by the cognitive theory proposed by Ortony et al. (1988), which views emotions as reactions to events. An example of converting a conversation segment into an EmoTransKG triple is illustrated in Figure 1. To evaluate EmoTransKG and analyze the underlying patterns and regularities in emotion transformation, we introduce EmoTransNet, a Transformer-based translational model that learns distributional representations of emotion knowledge triples. We then employ EmoTransNet for emotion retrofitting in CER to validate the effectiveness of EmoTransKG and the learned representations for visual analysis to uncover the impact of the head entity and the relation on the tail entity.

The primary contributions of this work can be summarized as follows:

- We propose a brand new emotion knowledge graph called EmoTransKG, which establishes connections among emotion entities and by defining unstructured, open-text event sequences as their relations. To the best of our knowledge, we are the first to construct emotion relations in such a knowledge graph framework.
- We introduce EmoTransNet, a Transformer-based translational model that captures the complex knowledge of emotion transformations embedded in the triples of EmoTransKG and the relatedness among relations.
- We apply EmoTransNet to retrofit the predicted emotions in CER baseline models and to visually analyze emotion knowledge triples. Experiments conducted on two benchmarking collections demonstrate that emotion retrofitting consistently enhances the performance of the models, and that EmoTransKG successfully establishes meaningful transformations and connections between emotions.

2 Related Works

2.1 Emotion Knowledge Graph

In the current literature, EKGs typically rely on two different representations of emotion: one represents emotions as points in a multidimensional appraisal space, and the other relies on predefined emotion keywords (Calvo and D'Mello, 2010). The former approach involves developing lexicons (Bradley and Lang, 1999; Warriner et al., 2013; Mohammad, 2018a) that map words onto a psychologically-based affect space. However, these emotion words are rated in a context-free environment, whereas in reality they are gathered in a context-dependent textual setting. In contrast, EmoTransKG constructs emotion knowledge that is not explicitly stated in the text and requires consideration of the event context. The latter approach focuses on text-based EKGs (Scherer, 1993; Gill et al., 2008) that attempt to establish connections between text and emotion keywords. However, the success of existing text-based EKGs (Gill et al., 2008) is limited to texts conveying fear and joy. To circumvent this limitation, some commonsense knowledge graphs (Bosselut et al., 2019; Zhang et al., 2020) utilize well-defined logical relations instead of strict matching relations. In contrast to their definition of texts as entities and without addressing the connections between emotions, EmoTransKG considers texts as relations between emotions, representing general knowledge of emotion transformations. This approach lowers our expectations for text awareness. In addition, the EmoTransKG structure exhibits a unique relation pattern of multi-edge, while also displaying common relation patterns—antisymmetry, inversion, and composition (Sun et al., 2019)—observed in existing KGs.

2.2 KG Representation Learning

In recent years, there have been significant advances in learning KG representations (Wang et al., 2017) for KG inference, with the translational approach gaining considerable popularity. This approach interprets relations as transformations from head entities to tail entities. TransE (Bordes et al., 2013), a seminal model in this family, assesses the plausibility of the triple by computing the semantic correlations of the triple items in a continuous feature space. To capture more complex structures in KGs, several variants of TransE, such as TransH (Wang et al., 2014), TransR (Lin et al., 2015), and TransD (Ji et al., 2015), have been developed. We also note that some works employ deep learning models to learn representations for KGs and formulate non-linear scoring functions for plausibility assessment. Notable works in this area include ConvKB (Nguyen et al., 2018) and HypER (Balazevic et al., 2019), which use convolutional neural networks, KEGCN (Yu et al., 2021), which utilizes a graph neural network, and both COMET (Bosselut et al., 2019) and KAN (Dun et al., 2021) build on the Transformer framework. In light of these advancements, we propose a Transformer-based translational model called EmoTransNet in this work. Our model not only effectively captures emotion knowledge encoded in event contexts (relations), but also identifies potential correlations among items within the triple.

2.3 CER Incorporating KG

CER has emerged as a crucial task in emotion recognition and interaction, with the goal of accurately predicting the emotion expressed by a speaker with each utterance or event in a conversation. Recent research in CER has focused on integrating external knowledge to enhance the understanding of conversational context. For instance, COSMIC (Ghosal et al., 2020), SKAIG (Li et al., 2021), and CauAIN (Zhao et al., 2022) integrate the commonsense knowledge graph COMET to strengthen the logic of contextual interactions. Similarly, KET (Zhong et al., 2019) introduces the emotion lexicon NRC VAD (Mohammad, 2018b) to enrich event representations. These efforts aim to enhance the perception of language and comprehension of event interactions, distinguishing them from the function of our EmoTransKG in CER; EmoTransKG aims to improve the transformation of the speaker's emotions during a conversation.

3 Methodology

In this section, we first outline the construction process of EmoTransKG and its characteristics (§3.1). Next, we present the EmoTransNet model for training the EmoTransKG base (§3.2). Finally, we apply the EmoTransNet for emotion retrofitting in CER (§3.3).

3.1 Construction of EmoTransKG

EmoTransKG is an EKG designed to construct emotion transformations and connections by extracting segments from communication transcripts. In particular, in a conversation consisting of a sequence of N utterances/events $U =$

$\{u_1, u_2, \ldots, u_N\}$ paired with corresponding emotion labels $\{y_1, y_2, \ldots, y_N\} \in \mathcal{S}$, where $\mathcal{S}$ is a predefined set of basic emotions consisting of emotion keywords, a segment denoted as $U_{i-j} = \{u_i, u_{i+1}, \ldots, u_j\}$ is extracted, such that $i < j$, along with two emotion labels $\{y_i, y_j\}$. It is essential to ensure consistent speaker identities for u_i and u_j, while distinguishing them from the rest of the events $U_{(i+1)-(j-1)}$. This one-hop link structure within the same speaker facilitates stronger connections between emotions, as the psychological interactions between events are locally effective, as claimed by Shen et al. (2021) and Li et al. (2021). Consequently, an EmoTransKG triple is created, where the head entity is y_i, the relation is U_{i-j}, and the tail entity is y_j, or succinctly expressed as $(h = y_i, r = U_{i-j}, t = y_j)$ (Ref. Figure 1).

Notation. We define $X^h \in \mathcal{S}$ as the head entity, a sequence of textual events $X^r = \{x_1^r, x_2^r, \ldots, x_{|r|}^r\}$ as the relation, and $X^t \in \mathcal{S}$ as the tail entity in the EmoTransKG triple. Clearly, in this definition, the EmoTransKG triple is instantiated with two symbolic entities and a natural language relation.

Relation Pattern. EmoTransKG possesses a unique graph structure characterized by a closed class of entities and an open class of relations, setting it apart from other KGs. The relations in EmoTransKG are temporally directed and linguistically irreversible, align with certain logical relations, such as `Causes` and `MadeOf` in COMET (Bosselut et al., 2019), while differing from certain semantic relations, such as `RelatedTo` and `SimilarTo` in ConceptNet 5.5 (Speer et al., 2017). Therefore, it is essential to discuss the relation patterns in EmoTransKG:

- **Antisymmetry:** As the relation in EmoTransKG is directed and irreversible, for any given triples (h, r, t) and (t, r, h), it holds that:
$$(h, r, t) \wedge (t, r, h) \Rightarrow h = t$$
- **Inversion:** EmoTransKG forms a strongly connected graph, where for any given triple (h, r_1, t), there exist a relation r_2 such that:
$$(h, r_1, t) \Rightarrow (t, r_2, h)$$
- **Composition:** EmoTransKG is a strongly connected graph, with each node forming a closed loop with its edges. For any given triples (h, r_1, t_1) and (t_1, r_2, t_2), there exist a relation r_3 such that:
$$(h, r_1, t_1) \wedge (t_1, r_2, t_2) \Rightarrow (h, r_3, t_2)$$

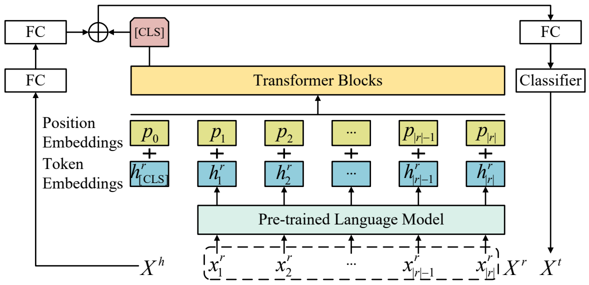


Figure 2: The proposed EmoTransNet model for EmoTransKG training.

- **Multi-edge:** In graph theory, multi-edge refers to two or more edges incident to the same pair of vertices. In EmoTransKG, for any given entities h and t, we have
$$(h, r_1, t) \wedge (h, r_2, t) \Rightarrow r_1 \neq r_2$$
A structure with such form is a multi-edge relation pattern, and its definition is first proposed and formalized in the domain of KGs.

3.2 EmoTransNet: Training EmoTransKG

Having created EmoTransKG, our goal is to study the influence of initial emotion (head entity) and events (relation) on future emotion (tail entity). To achieve this goal, we introduce EmoTransNet, a Transformer-based translational model designed to take the head entity X^h and its associated relation X^r as input and learn to generate the tail entity X^t as output, i.e., predicting t in $(h, r, ?)$. The proposed EmoTransNet model is depicted in Figure 2.

Head Entity Representation. We represent symbolic head entities using one-hot vectors in the space $\mathbb{R}^{|\mathcal{S}|}$. In order to enhance their expressive power, we define and use two fully connected layers to project the entity vectors into a higher dimensional feature space as follows:

$$h^h = W_2^h(W_1^h X^h + b_1^h) + b_2^h. \tag{1}$$

Here, $W_1^h \in \mathbb{R}^{|r| \times d_h}$, $W_2^h \in \mathbb{R}^{d_h \times d}$, $b_1^h \in \mathbb{R}^{d_h}$, and $b_1^h \in \mathbb{R}^d$ are trainable parameters. The resulting output, $h^h \in \mathbb{R}^d$, is considered to be the representation of the head entity.

Relation Representation. To reduce computation in extracting relation representations, we employ a two-step feature extraction scheme to account for factors such as the language length of the relation. Initially, we extract the event-level feature representation $h_i^r \in \mathbb{R}^d$ for each event x_i^r using a pretrained language model. Although the event itself is independent of the language model, we use the RoBERTa model (Liu et al., 2019) in this work.

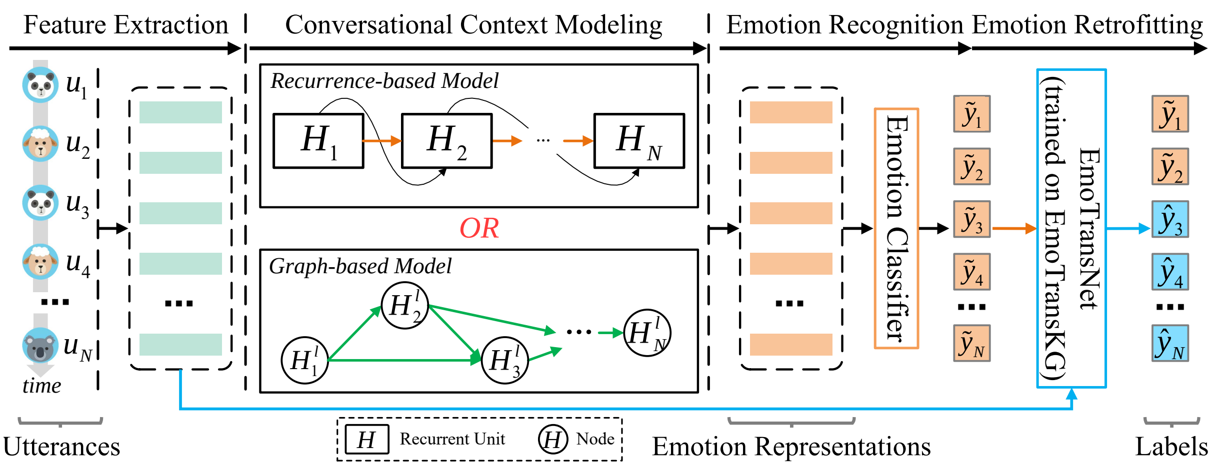


Figure 3: The architecture for integrating EmoTransNet, trained on EmoTransKG, with the CER model, incorporating three fundamental modules of the CER model: Feature Extraction, Conversational Context Modeling, and Emotion Recognition, alongside a novel module introduced by EmoTransNet: Emotion Retrofitting.

Upon obtaining the event representations $\{h_1^r, h_2^r, \ldots, h_{|r|}^r\}$ from RoBERTa, we proceed to capture the relation semantics involving these events. To achieve this, we prepend a special token, [CLS], which serves as an aggregator for the relation information, at the beginning of the event representations. This results in the creation of the input sequence $h^r = \{h_{[\mathrm{CLS}]}^r, h_1^r, h_2^r, ..., h_{|r|}^r\} \in \mathbb{R}^{(|r|+1)\times d}$. In order to account for the directed nature of the relation, we introduce position embeddings $p_t \in \mathbb{R}^{(|r|+1)\times d}$ encoded by sine and cosine functions into the input sequence:

$$h^{r,0} = h^r + p_t. \tag{2}$$

Next, we employ a standard Transformer architecture (Vaswani et al., 2017) with L layers to encode the input sequence $h^{r,0} \in \mathbb{R}^{(|r|+1)\times d}$. The l-th Transformer block can be formulated as follows:

$$\tilde{g}^{r,l} = \textsc{MultiAttn}(h^{r,l-1}), \tag{3}$$
$$g^{r,l} = \textsc{LayerNorm}(\tilde{g}^{r,l} + h^{r,l-1}), \tag{4}$$
$$\tilde{h}^{r,l} = \mathrm{FFN}(g^{r,l}), \tag{5}$$
$$h^{r,l} = \textsc{LayerNorm}(\tilde{h}^{r,l} + g^{r,l}), \tag{6}$$

where MultiAttn is a multi-head self-attention mechanism, LayerNorm is a layer normalization operation, and FFN is a two-layer feed-forward neural network with ReLU activation. After propagating through L layers, the hidden state $h_{[\mathrm{CLS}]}^{r,L} \in \mathbb{R}^d$ of the [CLS] token from the last layer is adopted as the representation of the relation.

Tail Entity Representation. The head entity representation h^h and the relation representation $h_{[\mathrm{CLS}]}^{r,L}$ are concatenated and fed into a fully connected layer to obtain the representation of the tail entity:

$$h^t = W_1^t[h^h \| h_{[\mathrm{CLS}]}^{r,L}] + b_1^t. \tag{7}$$

Here, $\|$ denotes the concatenation operation. $W_1^t \in \mathbb{R}^{2d\times d}$ and $b_1^t \in \mathbb{R}^d$ are learnable parameters.

Loss Function. Finally, we utilize a fully connected layer with softmax activation as the classifier to predict the tail entity:

$$\bar{Y} = \mathrm{softmax}(W_2^t h^t + b_2^t), \tag{8}$$

where $W_2^t \in \mathbb{R}^{d\times|\mathcal{S}|}$ and $b_2^t \in \mathbb{R}^{|\mathcal{S}|}$. The cross-entropy loss utilized to train the EmoTransNet model is calculated on all tail entities by:

$$L = -\frac{1}{M}\sum_{i=1}^{M}\sum_{e=1}^{|\mathcal{S}|} y_i^e \log(\bar{Y}_i^e), \tag{9}$$

where M is the number of triples in EmoTransKG, y_i^e is the one-hot vector denoting the tail entity of triple i, and e is the dimension of each entity.

3.3 CER with EmoTransNet

One way to assess the plausibility and effectiveness of KGs is to examine the correlation and consistency between entities and relations. The CER task serves as a benchmark for evaluating emotion propagation and transformation, testing whether a method can accurately capture emotion interactions in a conversational context. In this study, we employ the CER task as an automatic evaluation metric for EmoTransKG through EmoTransNet.

Retrofitting, as redefined by Faruqui et al. (2015), is a process of adjusting word embeddings using a knowledge graph. Generalizing this concept, we utilize EmoTransKG to adjust the predicted emotion confidence vectors produced by the CER model, which we refer to as "emotion retrofitting." Figure 3 illustrates the integration of EmoTransNet with the CER model, wherein EmoTransNet's parameters are frozen. The CER framework comprises three fundamental modules: Feature Extraction (FE) for encoding the semantic embeddings of each utterance, Conversational Context Modeling (CCM) for capturing interactions between utterance units or nodes using a recurrent- or graph-based model, and Emotion Recognition (ER) for predicting the emotion label of each utterance. This can be expressed as follows:

$$\tilde{Y} = \text{ER}(\text{CCM}(\text{FE}(U))), \tag{10}$$

where $\tilde{Y}_i$ represents the predicted emotion confidence vector for utterance u_i.

The emotion retrofitting function can be represented as follows:-

$$\hat{Y}_j = \gamma \tilde{Y}_j + (1-\gamma)\text{EmoTransNet}(Y_i, U_{i-j}), \tag{11}$$

where γ is a hyper-parameter that controls the preservation of emotion confidence in the predicted vector space. Note that, (1) the turns i and j correspond to the same speaker; (2) $Y_i = \text{One-Hot}(\tilde{Y}_i)$, where One-Hot$(\cdot)$ is a one-hot assignment operation: the predicted emotion confidence vector of the first utterance of each speaker is not adjusted; and (3) $Y_i = \text{One-Hot}(\hat{Y}_i)$: the adjusted emotion confidence vector $\hat{Y}_i$ will replace the original $\tilde{Y}_i$ as the input of EmoTransNet. This replacement is made because $\hat{Y}_i$ is more robust than $\tilde{Y}_i$, even though the effect of this robustness on future emotions decays exponentially with sequence turns. The emotion retrofitting described above is a plug-in module; it can refine emotion confidence vectors predicted by any CER model, as the updates in Eq. 11 are agnostic to the CER model.

Another task of entity prediction in EmoTransKG is to predict h in $(?, r, t)$. This facilitates the development of a bi-directional emotion retrofitting scheme similar to Bi-LSTM, although it is not currently operational. It is evident that our designed EmoTransNet is conceived as a single approach for extracting knowledge concerning emotion transformation within EmoTransKG.

Dataset	# Conversations Train	# Conversations Valid	# Conversations Test	# Utterances Train	# Utterances Valid	# Utterances Test
IEMOCAP	100	20	31	4,810	1,000	1,623
EmoryNLP	713	99	85	9,934	1,344	1,328

Dataset	# Classes / # Entities	# Relations Train	# Relations Valid	# Relations Test	Avg. U. PR.
IEMOCAP	6	4,947	623	1,561	2.97
EmoryNLP	7	5,347	689	739	3.52

Table 1: Dataset and derived EmoTransKG statistics. "Avg. U. PR." is the average number of utterances/events per relation.

4 Experiments

4.1 Experimental Setup

Dataset and Evaluation Metrics. EmoTransKG uses segments from conversation transcripts to create emotion knowledge triples. In this work, we use the IEMOCAP (Busso et al., 2008) and EmoryNLP (Zahiri and Choi, 2018) resources as conversation transcripts, but other conversation resources could have been used as well, since the ambiguity surrounding the definition of basic emotions. IEMOCAP is a collection of dyadic conversations that defines basic emotions as `happy`, `sad`, `neutral`, `angry`, `excited`, and `frustrated`, based on conversational content, facial expressions, and hand movements. EmoryNLP consists of textual transcripts from multi-speaker conversations in the TV show Friends and defines basic emotions as follows: positive: {`joyful`, `peaceful`, `powerful`}, negative: {`scared`, `sad`, `mad`}, neutral: {`neutral`}.

Table 1 presents the statistics of all datasets and their corresponding EKGs. Following recent CER works (Ghosal et al., 2020; Li et al., 2021), we use only the textual data from the above datasets and choose weighted-F1 as the evaluation metric for CER tasks. For both the creation of EmoTransKG and the training of EmoTransNet, the training set of each dataset is used. Additionally, to test the statistical significance of the performance improvement, we perform a paired t-test (Kim, 2015).

Baseline Models. Given that emotion retrofitting is a model-agnostic plug-in that facilitates the integration of various CER models, we select ten CER models to test whether EmoTransKG can improve their performance. Please refer to Appendix A for a comprehensive list of the baseline models and to Appendix B for implementation details.

Method	IEMOCAP	EmoryNLP
Recurrence-based methods		
DialogueRNN (Majumder et al., 2019)	66.03	38.25
DialogueRNN+EmoTransKG	67.19 (↑ 1.16)	38.90 (↑ 0.65)
COSMIC* (Ghosal et al., 2020)	67.35	38.59
COSMIC*+EmoTransKG	68.39 (↑ 1.04)	39.06 (↑ 0.47)
BiERU (Li et al., 2022)	64.59	36.72
BiERU+EmoTransKG	66.25 (↑ 1.66)	37.27 (↑ 0.55)
Graph-based methods		
DialogueGCN (Ghosal et al., 2019)	65.88	38.13
DialogueGCN+EmoTransKG	67.11 (↑ 1.23)	38.80 (↑ 0.67)
RGAT (Ishiwatari et al., 2020)	65.59	37.84
RGAT+EmoTransKG	66.79 (↑ 1.20)	39.01 (↑ 1.17)
SKAIG* (Li et al., 2021)	66.71	38.40
SKAIG*+EmoTransKG	67.59 (↑ 0.88)	38.93 (↑ 0.53)
DAG-ERC (Shen et al., 2021)	67.74	38.84
DAG-ERC+EmoTransKG	68.32 (↑ 0.58)	39.48 (↑ 0.64)
CauAIN* (Zhao et al., 2022)	65.20	38.31
CauAIN*+EmoTransKG	66.38 (↑ 1.18)	38.89 (↑ 0.58)
SUNET (Song et al., 2023)	66.98	38.79
SUNET+EmoTransKG	67.52 (↑ 0.54)	39.32 (↑ 0.53)
DualGATs (Zhang et al., 2023)	65.22	37.79
DualGATs+EmoTransKG	65.89 (↑ 0.67)	38.28 (↑ 0.49)

Table 2: The overall results on different methods on two CER datasets. The marker "*" indicates the incorporation of external knowledge. All "+EmoTransKG" methods show significant test p-value < 0.05 compared to their corresponding baseline models.

4.2 Evaluation Results and Analysis

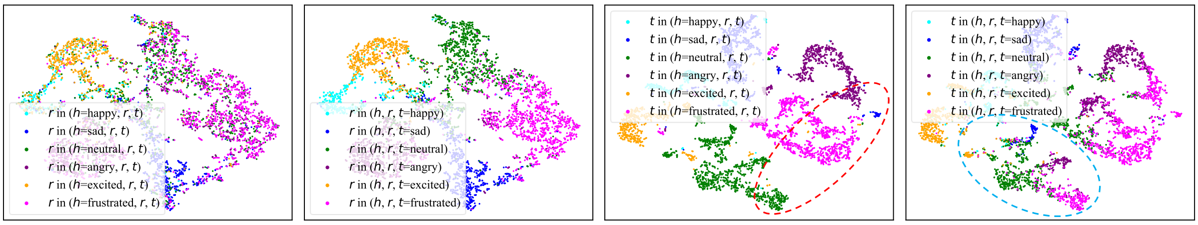
The overall results are showcased in Table 2, where "X+EmoTransKG" indicates the combination of model X with the emotion transformation knowledge obtained from our proposed EmoTransKG. EmoTransKG shows consistent improvement in performance with high confidence (p<0.05) on both the IEMOCAP and EmoryNLP datasets when evaluated within the same CER framework. These improvements on CER highlight the effectiveness of our EmoTransKG and the applicability of emotion retrofitting. In particular, the performance improvements achieved by EmoTransKG on models modeling intra-speaker dependencies demonstrate the breadth and depth of its understanding of emotion transformations. Implementing EmoTransKG on the external knowledge-introduced models also enhances their performance, emphasizing that the emotion transformations constructed by EmoTransKG differ from logical interactions and provide improved performance in emotion interactions from a new perspective.

Although our approach performs acceptably on EmoryNLP, it does not yield results as remarkable as those on IEMOCAP. Table 1 provides two reasons for this discrepancy. One reason is that the ratio of the number of relations to the number of utterances in the IEMOCAP test set is nearly twice that of the EmoryNLP test set. This implies that our EmoTransNet model runs emotion retrofitting about twice as frequently in IEMOCAP as in EmoryNLP during testing. Additionally, consuming fewer utterances per relation (2.97<3.52) leads to more robust emotion transformations.

4.3 Visualization Analysis

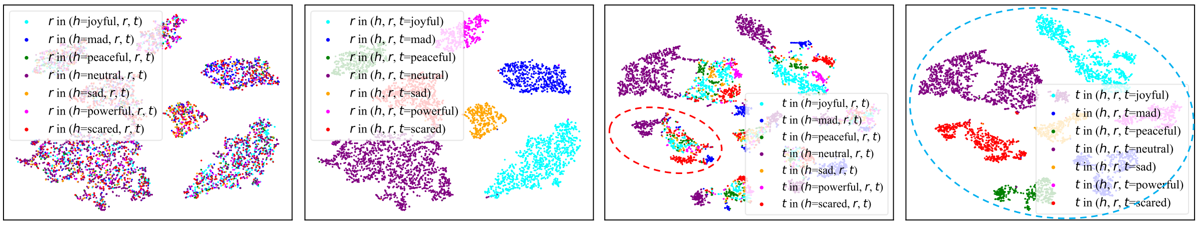
To intuitively assess the semantic relatedness of relations and the impact of the head entity and the relation on the tail entity in EmoTransKG, we project the learned relation representations ($h^{r,L}_{[\mathrm{CLS}]}$) and tail entity representations (h^t) from EmoTransNet into a 2D space using t-SNE (Maaten and Hinton, 2008). The resulting visualizations are presented in Figures 4 and 5, where each color represents a different head or tail entity category that is assigned in the EmoTransKG triples.

As you can see in Figures 4(a) and 5(a), EmoTransNet invariably generates similar relation representations for EmoTransKG triples with identical head and tail entities, demonstrating their semantic affinity. However, while Figure 4(a) reflects the categories of both head and tail entities, Figure 5(a) only shows the categories of tail entities. The difference in the EmoryNLP-derived EmoTransKG can be attributed to the strong correlation between tail entities and relations, which masks the weak correlation with head entities. Figure 4(b) demonstrates the separability of tail entity representations based on head entities (e.g., region shown in the dashed red ellipse for the same tail entity "`frustrated`") and relations (e.g., region shown in the dashed blue ellipse for the same head entity "`neutral`"). Such a visualization suggests that the head entities determine the intra-class distribution of tail entities, and the relations determine the inter-class distribution. Similarly, Figure 5(b) shows a comparable distribution; however, with more blurred boundaries in the left figure and clearer boundaries in the right figure. This discrepancy indicates that, compared to the IEMOCAP-derived EmoTransKG, the head entity has less influence, while the relation has a stronger impact on the tail entity in the EmoryNLP-derived EmoTransKG. These separable distributions explain the effectiveness of our EmoTransNet in extracting emotional knowledge and the plausibility of emotion transformations within EmoTransKG. To further support the differences in visualizations



(a) Relation (b) Tail entity

Figure 4: t-SNE visualization of relation and tail entity representations from IEMOCAP-derived EmoTransKG.



(a) Relation (b) Tail entity

Figure 5: t-SNE visualization of relation and tail entity representations form EmoryNLP-derived EmoTransKG.

between the two EmoTransKGs, we provide their quantitative analyses in Appendix C.

4.4 Ablation study

We present the following ablations of the EmoTransKG construction, without considering the model structure and training techniques for EmoTransNet. Firstly, we investigate how varying the amount of relations available for the construction of EmoTransKG affects the diversity of emotion transformations. Secondly, in our initial notation of EmoTransKG, head entities are represented as separate one-hot vectors, which requires EmoTransNet to learn emotion-specific semantic representations from scratch. As an ablation, we train a model by mapping head entities to natural language and initializing them with Word2vec embeddings (Mikolov et al., 2013) to incorporate knowledge from language, which we denote EmoTransKG-WE. We report the ablation results of COSMIC and DialogueGCN in Table 3.

We observe that increasing the number of relation facts can be increased to allow the model to learn a greater diversity of emotion transformation knowledge, thus improving its generalization ability. When using word embedding initialization, it has been noticed that there is inferior performance in CER models. One hypothesis is that this occurs because the inaccuracy of the provided Word2vec embeddings. To validate this hypothesis, we compute the cosine similarity between Word2vec embeddings of basic emotions on EmoryNLP, and find that the cosine similarity between `Joyful` and `Sad` is higher than that between `Joyful` and `Powerful`, consistent with Agrawal et al. (2018)'s observations. This result corroborates that word embeddings derived from textual corpora are inadequate for accurately capturing emotion semantics, presenting a significant challenge for word embedding learning. In contrast, the entity representations (h^h) learned from EmoTransNet demonstrate that emotions with the same polarity exhibit higher cosine similarity compared to those with different polarities, indicating more effective semantic learning of emotions. In addition, we observe a steeper performance degradation in IEMOCAP due to the stronger influence of head entities on tail entities in the IEMOCAP-derived EmoTransKG, as revealed through our visualization analysis (§4.3).

Method	% # Rel	IEMOCAP	EmoryNLP
COSMIC	-	67.35	38.59
+EmoTransKG	10%	67.54 (↑ 0.19)	38.62 (↑ 0.03)
	50%	67.82 (↑ 0.47)	38.77 (↑ 0.18)
	FULL	68.39 (↑ 1.04)	39.06 (↑ 0.47)
+EmoTransKG-WE	-	67.07 (↓ 0.28)	38.71 (↑ 0.12)
DialogueGCN	-	65.88	38.13
+EmoTransKG	10%	66.04 (↑ 0.16)	38.04 (↓ 0.09)
	50%	66.33 (↑ 0.45)	38.34 (↑ 0.21)
	FULL	67.11 (↑ 1.23)	38.80 (↑ 0.67)
+EmoTransKG-WE	-	65.66 (↓ 0.22)	38.33 (↑ 0.20)

Table 3: Effect of the amount of relations (# Rel) and emotion entity initialization using word embeddings (EmoTransKG-WE) on emotion transformations.

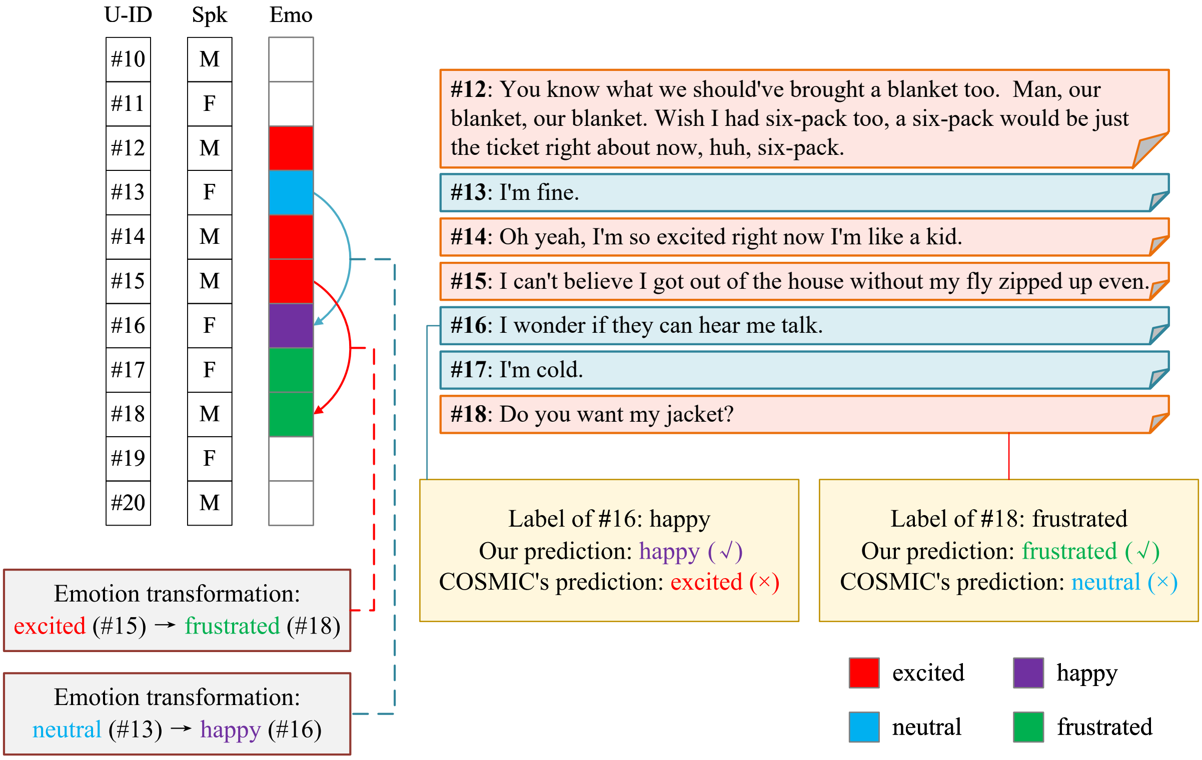


Figure 6: Two real cases that our EmoTransNet gives the correct predictions while COSMIC fails. Emotion transformations are illustrated within EmoTransNet.

4.5 Case Study

In Figure 6, we present two cases (#13 → #16 and #15 → #18) of emotion prediction, where our EmoTransNet yields accurate predictions while COSMIC falls short. In both conversation segments, the continuous emotions conveyed in the individuals' first and last utterances are transferred due to the experienced events. Therefore, the model must consider both emotional inertia resulting from prior emotions and the perception of events.

Specifically, EmoTransNet assigns the emotion label "excited" to #16, despite events #14 and #15 conveying the positive semantic polarity "happy." This suggests that emotion transformation is influenced by the initial emotion (the head entity) rather than solely by the experienced events (the relation). Conversely, COSMIC's prediction for #16 merely replicates or empathizes with the speaker M's emotion. For the prediction of #18, despite its sequential modeling, COSMIC struggles to capture or comprehend historical information necessary for accurate prediction. This highlights COSMIC's limitations in emotional interaction and contextual modeling, despite the presence of contextual utterances providing effective cues for emotion prediction. Our EmoTransNet, which adopts an emotion transformation perspective, predicts future emotions based on initial emotions and experienced events. This approach mitigates potential misjudgments resulting from emotional contagion and from information loss during long-distance transmission.

5 Conclusion

In this paper, we present EmoTransKG: an emotion knowledge graph. We define basic emotions in conversation transcripts as emotion entities, and extract event sequences between two entities with the same participant as emotion relations. EmoTransKG is the first EKG that establishes emotion transformations and connections. We train an EmoTransNet model on EmoTransKG and utilize it for emotion retrofitting in CER, as well as to visualize the roles of entities and relations in emotion transformations, thereby demonstrating the plausibility and effectiveness of EmoTransKG. Currently, we are extending our approach to other conversational resources, such as DailyDialog (Li et al., 2017) and MELD (Poria et al., 2019), and fusing these EmoTransKGs to achieve high coverage of emotion concepts. We also investigate whether EmoTransKG can offer recommendations for the engineering of emotion-related word embeddings.

Limitations

We note several limitations of our work:

(1) Our EmoTransKG is limited by the fact that the definition of basic emotions is disputed, which poses a challenge for future efforts to extend and fuse emotion knowledge graphs. Although our work has not addressed this issue, our visualization results demonstrate basic emotions from different emotion theories can be well separated from each other. These findings provide valuable insights for advancing research in emotions and cognition.

(2) The diversity of relations requires a significant amount of available emotional transfer facts for successful editing, but the linguistic richness and diversity limit its coverage.

(3) We do not currently account for data imbalance in EmoTransKG. As shown in Figures 4(b) and 5(b), the EmoTransNet model, trained on the target variable (tail entity) with such a significantly skewed distribution, exhibits biased, generally leading to better performance for emotion triples where the head and tail entities are the same.

Acknowledgements

This work was supported by National Natural Science Foundation of China (NSFC Grant No. 62076092).

References

Ameeta Agrawal, Aijun An, and Manos Papagelis. 2018. Learning emotion-enriched word representations. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 950–961.

Ivana Balazevic, Carl Allen, and Timothy M. Hospedales. 2019. Hypernetwork knowledge graph embeddings. In *Proceedings of the 28th Artificial Neural Networks and Machine Learning*, volume 11731, pages 553–565.

Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, volume 26, pages 2787–2795.

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4762–4779.

Margaret M. Bradley and Peter J. Lang. 1999. Affective norms for english words (anew): Instruction manual and affective ratings. *journal royal microscopical society*.

Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N. Chang, Sungbok Lee, and Shrikanth S. Narayanan. 2008. IEMOCAP: interactive emotional dyadic motion capture database. *Language resources and evaluation*, 42(4):335–359.

Rafael A. Calvo and Sidney D'Mello. 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1):18–37.

Yaqian Dun, Kefei Tu, Chen Chen, Chunyan Hou, and Xiaojie Yuan. 2021. KAN: knowledge-aware attention network for fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 81–89.

Manaal Faruqui, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard H. Hovy, and Noah A. Smith. 2015. Retrofitting word vectors to semantic lexicons. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1606–1615.

Christiane Fellbaum. 1998. *WordNet: An electronic lexical database*. MIT press.

Deepanway Ghosal, Navonil Majumder, Alexander F. Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. COSMIC: commonsense knowledge for emotion identification in conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2470–2481.

Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander F. Gelbukh. 2019. Dialoguegcn: A graph convolutional neural network for emotion recognition in conversation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 154–164.

Alastair J Gill, Robert M French, Darren Gergle, and Jon Oberlander. 2008. Identifying emotional characteristics from short blog texts. In *Proceedings for the 30th Annual Meeting of the Cognitive Science Society*, pages 2237–2242.

Leslie S. Greenberg. 2002. *Emotion-Focused Therapy: Coaching Clients to Work Through Their Feelings*. American Psychological Association.

David Helbig, Enrica Troiano, and Roman Klinger. 2020. Challenges in emotion style transfer: An exploration with a lexical substitution pipeline. In *Proceedings of the 8th International Workshop on Natural Language Processing for Social Media*, pages 41–50.

Taichi Ishiwatari, Yuki Yasuda, Taro Miyazaki, and Jun Goto. 2020. Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 7360–7370.

Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 687–696.

Tae Kyun Kim. 2015. T test as a parametric statistic. *Korean journal of anesthesiology*, 68(6):540–546.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Jiangnan Li, Zheng Lin, Peng Fu, and Weiping Wang. 2021. Past, present, and future: Conversational emotion recognition through structural modeling of psychological knowledge. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1204–1214.

Juncen Li, Robin Jia, He He, and Percy Liang. 2018. Delete, retrieve, generate: a simple approach to sentiment and style transfer. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1865–1874.

Wei Li, Wei Shao, Shaoxiong Ji, and Erik Cambria. 2022. Bieru: Bidirectional emotional recurrent unit for conversational sentiment analysis. *Neurocomputing*, 467:73–82.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. Dailydialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of the 8th International Joint Conference on Natural Language Processing*, pages 986–995.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, pages 2181–2187.

Hugo Liu and Push Singh. 2004. Conceptnet—a practical commonsense reasoning tool-kit. *BT technology journal*, 22(4):211–226.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.

Fuli Luo, Peng Li, Jie Zhou, Pengcheng Yang, Baobao Chang, Xu Sun, and Zhifang Sui. 2019. A dual reinforcement learning framework for unsupervised text style transfer. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 5116–5122.

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(11).

Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander F. Gelbukh, and Erik Cambria. 2019. Dialoguernn: An attentive RNN for emotion detection in conversations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6818–6825.

Cynthia Matuszek, Michael Witbrock, John Cabral, and John DeOliveira. 2006. An introduction to the syntax and content of cyc. *UMBC Computer Science and Electrical Engineering Department Collection*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119.

Saif Mohammad. 2018a. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 174–184.

Saif M. Mohammad. 2018b. Obtaining reliable human ratings of valence, arousal, and dominance for 20, 000 english words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 174–184.

Dai Quoc Nguyen, Tu Dinh Nguyen, Dat Quoc Nguyen, and Dinh Q. Phung. 2018. A novel embedding model for knowledge base completion based on convolutional neural network. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics*, pages 327–333.

Andrew Ortony, Gerald L. Clore, and Allan Collins. 1988. *The Cognitive Structure of Emotions*. Cambridge University Press.

Charles Egerton Osgood, George J Suci, and Percy H Tannenbaum. 1957. *The measurement of meaning*. University of Illinois press.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. Meld: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 527–536.

James A Russell. 2003. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145.

Klaus R. Scherer. 1993. Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition & Emotion*, 7:325–355.

Tianxiao Shen, Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2017. Style transfer from non-parallel text by cross-alignment. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, volume 30, page 6833–6844.

Weizhou Shen, Siyue Wu, Yunyi Yang, and Xiaojun Quan. 2021. Directed acyclic graph network for conversational emotion recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pages 1551–1560.

Amit Singhal. 2012. Introducing the knowledge graph: Things, not strings. Google Blog. Retrieved from https://www.blog.google/products/search/introducing-knowledge-graph-things-not/.

Rui Song, Fausto Giunchiglia, Lida Shi, Qiang Shen, and Hao Xu. 2023. SUNET: speaker-utterance interaction graph neural network for emotion recognition in conversations. *Engineering Applications of Artificial Intelligence*, 123:106315.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, pages 4444–4451.

Ryan A Stevenson, Joseph A Mikels, and Thomas W James. 2007. Characterization of the affective norms for english words by discrete emotional categories. *Behavior research methods*, 39(4):1020–1024.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008.

Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29:2724–2743.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, pages 1112–1119.

Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45:1191–1207.

Donghan Yu, Yiming Yang, Ruohong Zhang, and Yuexin Wu. 2021. Knowledge embedding based graph convolutional network. In *Proceedings of the Web Conference 2021*, pages 1619–1628.

Sayyed M. Zahiri and Jinho D. Choi. 2018. Emotion detection on TV show transcripts with sequence-based convolutional neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 18, pages 44–52.

Dong Zhang, Liangqing Wu, Changlong Sun, Shoushan Li, Qiaoming Zhu, and Guodong Zhou. 2019. Modeling both context- and speaker-sensitive dependence for emotion detection in multi-speaker conversations. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 5415–5421.

Duzhen Zhang, Feilong Chen, and Xiuyi Chen. 2023. Dualgats: Dual graph attention networks for emotion recognition in conversations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 7395–7408.

Hongming Zhang, Xin Liu, Haojie Pan, Yangqiu Song, and Cane Wing-Ki Leung. 2020. ASER: A large-scale eventuality knowledge graph. In *Proceedings of the web conference 2020*, pages 201–211.

Weixiang Zhao, Yanyan Zhao, and Xin Lu. 2022. Cauain: Causal aware interaction network for emotion recognition in conversations. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence*, pages 4524–4530.

Peixiang Zhong, Di Wang, and Chunyan Miao. 2019. Knowledge-enriched transformer for emotion detection in textual conversations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 165–176.

A Baseline Models

We use the following baseline models in our evaluation, and their corresponding codes and parameters are uploaded in the supplementary material:

Recurrence-based Models: DialogueRNN (Majumder et al., 2019), COSMIC (Ghosal et al., 2020), and BiERU (Li et al., 2022).

Graph-based Models: DialogueGCN (Ghosal et al., 2019), RGAT (Ishiwatari et al., 2020), SKAIG (Li et al., 2021), DAG-ERC (Shen et al., 2021), CauAIN (Zhao et al., 2022), SUNET (Song et al., 2023), and DualGATs (Zhang et al., 2023).

These models are chosen based on their strong performance and prominence in the CER field, as well as the availability of source code to ensure the reliability and reproducibility of results. All models, except for BiERU, explicitly build intra-speaker dependencies to facilitate the comprehension of emotion propagations and connections of individual speakers in conversations. Additionally, COSMIC, SKAIG, and CauAIN incorporate external knowledge into their frameworks to enhance the logical interactions between utterances/events.

B Implement Details

This section details the hyper-parameters and optimal combinations selected. Specifically, we set d_h to 300, d to 1024, and γ in Eq. 11 to 0.95. During the training phase of EmoTransNet, we utilize the Adam optimization algorithm (Kingma and Ba, 2014) with a batch size of 64, 100 epochs, a learning rate of 2e-5, a dropout rate of 0.5, and a weight decay of 3e-4. To balance performance and computational complexity, we set the number of Transformer blocks L to 6. Consistently, we use the RoBERTa (Liu et al., 2019) feature extractor to obtain utterance embeddings for the Feature Extraction module in all baseline models mentioned. For the re-implementations and emotion retrofitting experiments, we strictly adhere to the settings reported in the original papers. The experiments are conducted on a server equipped with a 3090 GPU card and the Ubuntu operating system version 20.04. The reported results are based on the average score obtained from 20 random runs on the test set.

C Quantitative Analysis

In this section, we perform a quantitative analyses of emotion transformations within EmoTransKG. We examine the two following aspects:

(1) **Impact of head entities on emotion transformations.** Specifically, for each triple in the validation set, we randomly select a different emotion entity to replace its head entity, and then input the replaced head entity and its corresponding relation into EmoTransNet to assess the probability of correctly predicting the tail entity. The EmoryNLP dataset shows a significantly higher probability of correctly predicting the tail entity (85%) compared to the IEMOCAP dataset (62%). This result indicates that the head entities in the IEMOCAP-derived EmoTransKG have a greater impact on predicting the tail entities than those in the EmoryNLP-derived EmoTransKG.

(2) **Impact of relations on emotion transformations.** We find that in the IEMOCAP-derived EmoTransKG, 70% of the triples have the same head and tail entities. However, this proportion decreases to only 36% in the EmoryNLP-derived EmoTransKG. These results suggest that the relations derived from EmoryNLP are more likely to transform the head entity into a tail entity of a different category.

In conclusion, our quantitative analyses indicate that both head entities and relations in EmoTransKG triples have an impact on the emotional expressions of tail entities. Furthermore, the influence of head entities is relatively weaker in the EmoryNLP-derived EmoTransKG compared to the IEMOCAP-derived EmoTransKG, while the impact of relations is stronger. These findings are consistent with the visualization results (§4.3).