

Supplementary materials of ECFCON: Emotion Consequence Forecasting in Conversations

1 DATASET

1.1 Why the *Home with Kids*

It is a situational comedy centered on children’s themes, telling the various interesting stories that occur between parents and three children after two divorced families merge. This comedy is full of humor and warmth, with vivid and interesting characters, making it very suitable for the study of emotional consequences in conversations. This drama is somewhat similar to *Friends* and can be considered as a drama of the same genre.

1.2 Meaning of the Emotions

- *anger*: It is a strong emotion characterized by feelings of hostility, frustration, and antagonism towards someone or something. It can arise from experiences of perceived wrongs, injustice, negligence, or threats.
- *disgust*: It is a strong feeling of aversion or repulsion towards something that is perceived as unpleasant, offensive, or revolting. This emotion can be triggered by a variety of stimuli such as unpleasant tastes, smells, sights, or actions.
- *fear*: It is a fundamental emotion experienced in anticipation of potential danger or a threat. It is a basic survival mechanism that signals our bodies to respond to danger with a fight or flight response.
- *happiness*: It is an emotional state characterized by feelings of joy, satisfaction, contentment, and fulfillment. While happiness has many different definitions, it generally involves positive emotions and life experiences.
- *sadness*: It is an emotional state characterized by feelings of unhappiness, grief, and sorrow. It is a natural human emotion that everyone experiences at various times in their lives, often as a reaction to situations involving loss, disappointment, or misfortune.
- *surprise*: It is an emotion experienced when something unexpected happens. It is characterized by a brief state of heightened alertness and is usually accompanied by a startle response, which might include raising the eyebrows, widening the eyes, and opening the mouth.
- *neutral*: It refers to a state where a person feels neither positive nor negative emotions particularly strongly. It’s a state of emotional balance or calmness where no specific emotion dominates the individual’s mood.

1.3 Datasets Comparison

As illustrated in Table 1, we have compiled a list of datasets for Emotion Recognition in Conversation (ERC), Emotion Cause Extraction (ECE) and Emotion Consequence Forecasting (ECF). Our dataset, ECFCON, is the first dataset specifically designed for emotion consequence forecasting in conversations. Particularly, we provide three modalities for ECFCON, including text, audio and vision, which significantly enriches the content and enhances the diversity of the dataset. Our dataset contains a total of 39,950 utterances, which

is relatively larger than other datasets. Concretely, the size of our dataset surpasses all the datasets in ECE, as well as most datasets in ERC.

Table 1: Comparison of datasets for emotion, cause and consequence analysis. ERC: Emotion Recognition in Conversation; ECE: Emotion Cause Extraction; ECF: Emotion Consequence Forecasting. Languages included are EN: English, ZH: Chinese. Modalities provided are T: text, A: audio, V: vision. ‘Conv’ stands for conversation. Units of analysis are utterance (u), sentence (s), document (d), post (p).

Dataset	Task	Lng	Modality	Src	# Ins
IEMOCAP[2]	ERC	EN	T,A,V	Conv	10,239u
SEMAINE [12]	ERC	EN	T,A,V	Conv	5,798u
DailyDialog[11]	ERC	EN	T	Conv	102,979u
EmotionLines[9]	ERC	EN	T	Conv	14,503u
EmoContext[3]	ERC	EN	T	Conv	115,272u
MELD[13]	ERC	EN	T,A,V	Conv	13,708u
MELSD[5]	ERC	EN	T,A,V	Conv	20,000u
Emotion-Stimulus[7]	ECE	EN	T	-	2,414s
ECE Corpus [8]	ECE	ZH	T	News	2,105d
NTCIR-13-ECA[6]	ECE	ZH	T	News	2,403d
Weibo-Emotion[4]	ECE	ZH	T	Blog	7,000p
REMAN[10]	ECE	EN	T	Fiction	1,720d
GoodNewsEveryone[1]	ECE	EN	T	News	5,000s
RECCON[14]	ECE	EN	T	Conv	11,769u
ECF[15]	ECE	EN	T,A,V	Conv	13,509u
ECFCON	ECF	ZH	T,A,V	Conv	39,950u

1.4 Dataset Statistics

The distribution statistics of the train, valid, and test sets are shown in Table 1. We chose an 8:1:1 split ratio based on the number of episodes. The number of emotional dialogue videos per episode is inconsistent, which means that the number of videos does not strictly follow this ratio. Nevertheless, the combined total of the test and valid sets are still approximates 20%, generally meeting this ratio. We adopt this division to ensure that the train, valid, and test sets each come from different episodes, thereby avoiding mutual interference.

Table 2: The statistics of train, valid and test sets in ECFCON.

Sets	Videos	Utterances	Emotions	Consequences
train	2286	33,107	10,124	7,104
valid	173	2,510	831	623
test	321	4,333	1,436	1,083

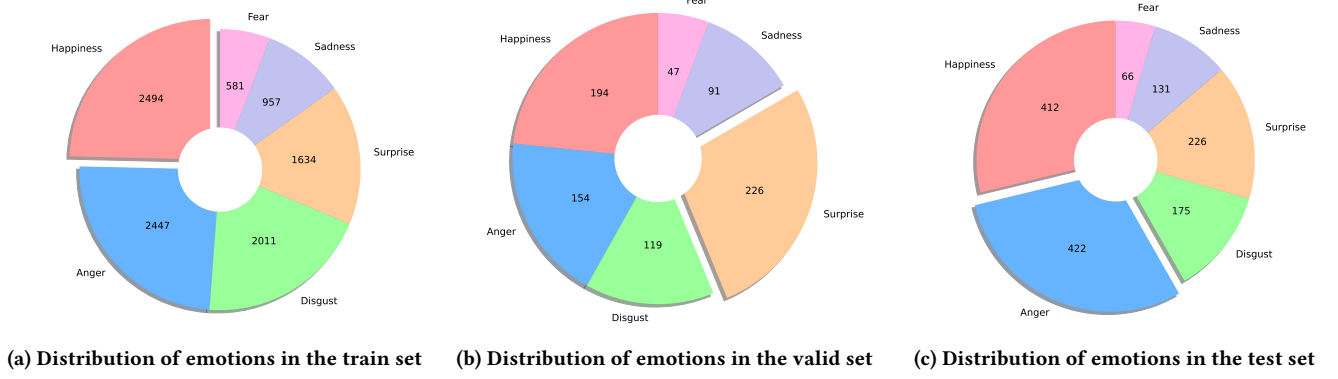


Figure 1: The distribution of emotions in ECFCO.

1.5 Emotion Distribution

The distribution of emotions in ECFCO is shown in Figure 1. It can be observed that the distribution of emotions across the three datasets differs. This variation is due to the division of the dataset according to complete episodes, where many episodes exhibit emotional biases, such as frequent occurrences of anger or surprise.

1.6 Consequence Types

Figure 2 illustrates the samples of consequence types in ECFCO. Based on the different speakers of the consequence utterances, the consequence types can be further subdivided into 4 categories as follows:

- *self-objective*: the objective consequence is generated by the same speaker as the emotion utterance.
- *inter-objective*: the objective consequence is generated by another speaker different from the one involved in the emotion utterance.
- *self-subjective*: the subjective consequence is generated by the same speaker as the emotion utterance.
- *inter-subjective*: the subjective consequence is generated by another speaker different from the one involved in the emotion utterance.

2 DETAILS OF THE EXPERIMENTS

2.1 Details of Few-shot Learning

Tables 3, 4, 5 shows the concrete result of few-shot learning for three subtasks: CF, ECPF, and ECPFC, respectively. It can be observed that although there are fluctuations in the scores, overall, the clue-driven hybrid approaches surpass traditional approaches. For instance, the clue-driven hybrid methods, i.e., ECFCO-RoBERTa in the CF task, MECPE-2steps in the ECPF task, and MECPE-2steps and ECFCO-BERT in the ECPFC task, all outperform traditional methods across different data sizes in the few-shot setting. Moreover, the performance of directly fine-tuning MLLMs is relatively poor.

Some of the clue-driven results are not as good as expected. This is because, in the few-shot training process, the limited number of samples may lead to significant fluctuations, making it easy to produce some unexpected results.

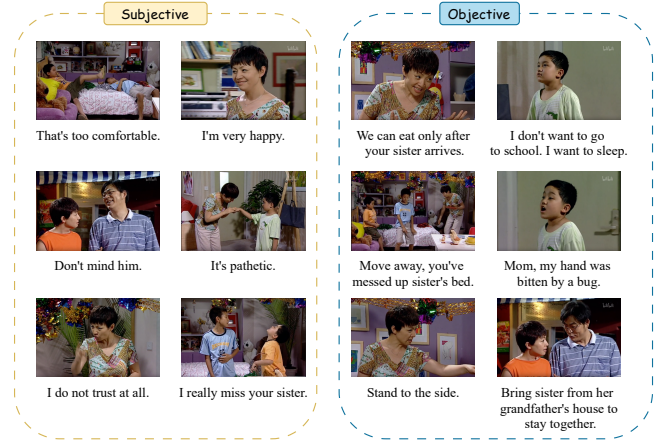


Figure 2: The samples of consequence types in ECFCO.

2.2 Error Analysis

As shown in Figure 3, we present two cases concerning emotion recognition and consequence forecasting.

In case (a), clues indicate laughter through both facial expressions and gestures, thereby inferring that the emotion is *happiness*. The traditional method fails to capture this clue, leading to incorrect predictions, while the clue-driven hybrid method successfully identifies the emotion.

In case (b), the *impact* and *why* clues suggest that utterance 7 is a consequence of utterance 4, aiding the clue-driven hybrid method in making accurate predictions. Specifically, regarding the *impact* clue, utterance 4 mentions jet lag and speculates that it might lead to the father's questioning and confusion. The *why* clue states that the reason for saying utterance 7 is the jet lag. This pairing of impact and why clues builds a logical loop, enhancing understanding of the relationship between the two utterances. Conversely, the traditional method fails to grasp this logical relationship, resulting in incorrect predictions.

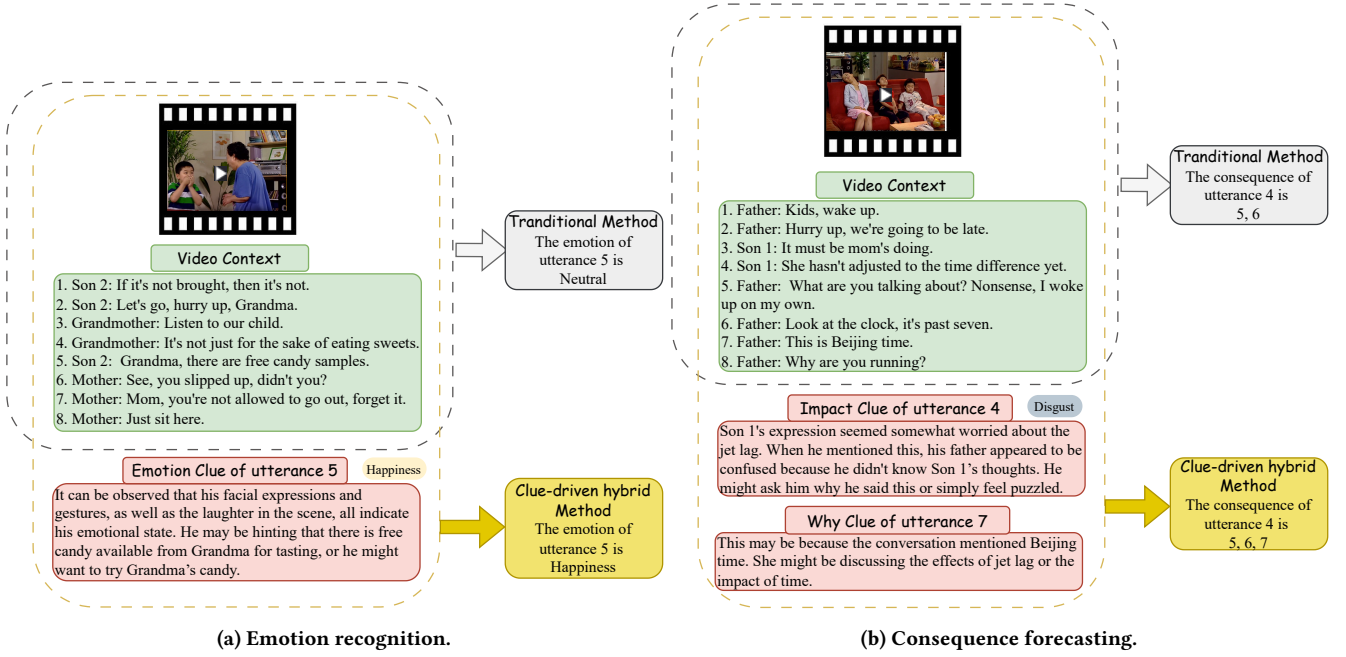


Figure 3: Error Analysis

Table 3: Few-shot for Consequence Forecasting (CF)

Method	0	1	10	100	1000	all
MECPE-2steps	0.0000	0.0000	0.4067	0.5104	0.5526	0.6137
MECPE-2steps + clues	0.0000	0.0000	0.3883	0.5530	0.5548	0.6194
ECFCON-LSTM(TAV)	0.0000	0.0000	0.4187	0.5326	0.5443	0.6187
ECFCON-LSTM(TAV) + clues	0.0000	0.0000	0.4036	0.5711	0.5332	0.6255
ECFCON-BERT(TAV)	0.0000	0.0000	0.3848	0.5742	0.5317	0.6299
ECFCON-BERT(TAV) + clues	0.0000	0.0000	0.3984	0.5732	0.5631	0.6344
ECFCON-RoBERTa(TAV)	0.0000	0.0000	0.4063	0.5328	0.5316	0.6210
ECFCON-RoBERTa(TAV) + clues	0.0000	0.0000	0.4366	0.5459	0.5767	0.6288
ECFCON-MLLMs	0.4453	0.0000	0.0037	0.0191	-	-

Table 4: Few-shot for Emotion Consequence Pair Forecasting (ECPF)

Method	0	1	10	100	1000	all
MECPE-2steps	0.0000	0.1347	0.1194	0.0314	0.2801	0.2872
MECPE-2steps + clues	0.0000	0.1480	0.1569	0.1389	0.2962	0.3312
ECFCON-LSTM(TAV)	0.0000	0.1052	0.1328	0.1134	0.2229	0.3200
ECFCON-LSTM(TAV) + clues	0.0000	0.1014	0.1387	0.1239	0.2278	0.3197
ECFCON-BERT(TAV)	0.0000	0.1282	0.1407	0.1538	0.2481	0.3203
ECFCON-BERT(TAV) + clues	0.0000	0.1418	0.1420	0.1510	0.2948	0.3495
ECFCON-RoBERTa(TAV)	0.0000	0.0439	0.1366	0.1893	0.2360	0.3443
ECFCON-RoBERTa(TAV) + clues	0.0000	0.0651	0.1239	0.1901	0.2428	0.3323
ECFCON-MLLMs	0.2021	0.0000	0.0000	0.0030	-	-

Table 5: Few-shot for Emotion Consequence Pair Forecasting (ECPF) with Categories

Method	0	1	10	100	1000	all
MECPE-2steps	0.0000	0.0386	0.0106	0.0000	0.1197	0.1506
MECPE-2steps + clues	0.0000	0.0000	0.0234	0.0015	0.1341	0.1662
ECFCON-LSTM(TAV)	0.0000	0.0043	0.0192	0.0000	0.0825	0.1382
ECFCON-LSTM(TAV) + clues	0.0000	0.0050	0.0228	0.0309	0.0801	0.1536
ECFCON-BERT(TAV)	0.0000	0.0283	0.0222	0.0000	0.0845	0.1554
ECFCON-BERT(TAV) + clues	0.0000	0.0352	0.0241	0.0057	0.1027	0.1803
ECFCON-RoBERTa(TAV)	0.0000	0.0183	0.0064	0.0000	0.0576	0.1548
ECFCON-RoBERTa(TAV) + clues	0.0000	0.0088	0.0243	0.0043	0.1145	0.1668
ECFCON-MLLMs	0.0776	0.0000	0.0000	0.0000	-	-

REFERENCES

- [1] Laura Ana Maria Bostan, Evgeny Kim, and Roman Klinger. 2020. GoodNewsEveryone: A Corpus of News Headlines Annotated with Emotions, Semantic Roles, and Reader Perception. In *Proceedings of LREC 2020*. European Language Resources Association, 1554–1566.
- [2] 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. *Lang. Resour. Evaluation* 42, 4 (2008), 335–359.
- [3] Ankush Chatterjee, Umang Gupta, Manoj Kumar Chinnakotla, Radhakrishnan Srikanth, Michel Galley, and Puneet Agrawal. 2019. Understanding Emotions in Text Using Deep Learning and Big Data. *Comput. Hum. Behav.* 93 (2019), 309–317.
- [4] Xiyao Cheng, Ying Chen, Bixiao Cheng, Shoushan Li, and Guodong Zhou. 2017. An Emotion Cause Corpus for Chinese Microblogs with Multiple-User Structures. *ACM Trans. Asian Low Resour. Lang. Inf. Process.* 17, 1 (2017), 6:1–6:19.
- [5] Mauajama Firdaus, Hardik Chauhan, Asif Ekbal, and Pushpak Bhattacharyya. 2020. MEISD: A Multimodal Multi-Label Emotion, Intensity and Sentiment Dialogue Dataset for Emotion Recognition and Sentiment Analysis in Conversations. In *Proceedings of COLING 2020*. International Committee on Computational Linguistics, 4441–4453.
- [6] Qinghong Gao, Jiannan Hu, Ruifeng Xu, Lin Gui, Yulan He, Kam-Fai Wong, and Qin Lu. 2017. Overview of NTCIR-13 ECA Task. In *Proceedings of NTCIR 2017*. National Institute of Informatics (NII).
- [7] Diman Ghazi, Diana Inkpen, and Stan Szpakowicz. 2015. Detecting Emotion Stimuli in Emotion-Bearing Sentences. In *Proceedings of CICLing 2015 (Lecture Notes in Computer Science, Vol. 9042)*. Springer, 152–165.
- [8] Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou. 2016. Event-Driven Emotion Cause Extraction with Corpus Construction. In *Proceedings of EMNLP 2016*, Jian Su, Xavier Carreras, and Kevin Duh (Eds.). The Association for Computational Linguistics, 1639–1649. <https://doi.org/10.18653/v1/d16-1170>
- [9] Chao-Chun Hsu, Sheng-Yeh Chen, Chuan-Chun Kuo, Ting-Hao K. Huang, and Lun-Wei Ku. 2018. EmotionLines: An Emotion Corpus of Multi-Party Conversations. In *Proceedings of LREC 2018*. European Language Resources Association (ELRA).
- [10] Evgeny Kim and Roman Klinger. 2018. Who Feels What and Why? Annotation of a Literature Corpus with Semantic Roles of Emotions. In *Proceedings of COLING 2018*. Association for Computational Linguistics, 1345–1359.
- [11] 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 IJCNLP 2017*. Asian Federation of Natural Language Processing, 986–995.
- [12] Gary McKeown, Michel François Valstar, Roddy Cowie, Maja Pantic, and Marc Schröder. 2012. The SEMAINE Database: Annotated Multimodal Records of Emotionally Colored Conversations between a Person and a Limited Agent. *IEEE Trans. Affect. Comput.* 3, 1 (2012), 5–17.
- [13] 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 ACL 2019*. Association for Computational Linguistics, 527–536.
- [14] Soujanya Poria, Navonil Majumder, Devamanyu Hazarika, Deepanway Ghosal, Rishabh Bhardwaj, Samson Yu Bai Jian, Pengfei Hong, Romila Ghosh, Abhinaba Roy, Niyati Chhaya, Alexander F. Gelbukh, and Rada Mihalcea. 2021. Recognizing Emotion Cause in Conversations. *Cogn. Comput.* 13, 5 (2021), 1317–1332.
- [15] Fanfan Wang, Zixiang Ding, Rui Xia, Zhao Yu Li, and Jianfei Yu. 2023. Multimodal Emotion-Cause Pair Extraction in Conversations. *IEEE Trans. Affect. Comput.* 14, 3 (2023), 1832–1844.