Towards Unified Emotion Understanding: A One-Step End-to-End Model for Emotion–Cause Pair Extraction

B. Kranthi Kumar[1], Prof. S. Aquter Babu [2]
[1]Research Scholar, Department of Computer Science & Technology, Dravidian University, Kuppam, Andhra Pradesh
[2]Professor, Department of Computer Science & Technology, Dravidian University, Kuppam, Andhra Pradesh

Abstract

Emotion analysis in textual data has evolved from simple emotion detection to sophisticated Emotion-Cause Pair Extraction (ECPE), which jointly identifies emotions and their corresponding causes without requiring pre-annotated emotion labels. Traditional ECPE approaches employ multi-stage pipelines that separately extract emotions and causes before pairing, leading to significant error propagation that degrades overall performance. This paper proposes a novel one-step, end-to-end deep learning framework that directly extracts emotion-cause pairs from unannotated documents, eliminating intermediate error accumulation while capturing their inherent interdependency through unified contextual modeling.

Our model leverages advanced transformer architectures with multi-granular attention mechanisms to simultaneously encode document-level semantics, clause interactions, and position-aware dependencies. Extensive experiments on the benchmark ECPE-News dataset demonstrate superior F1-scores compared to state-of-the-art two-step and existing end-to-end methods, achieving substantial improvements in both precision and recall. The proposed approach exhibits enhanced generalization across diverse domains including social media and conversational texts, addressing key limitations of prior work. By providing a robust, efficient solution for real-world emotion analysis applications such as mental health monitoring, customer sentiment analysis, and dialogue systems, this work advances practical emotion understanding capabilities significantly.

Literature Review: Emotion-Cause Pair Extraction in Text Analysis

1. Introduction

Emotion analysis in text has evolved from simple sentiment classification to sophisticated tasks that identify not only the emotions expressed but also the underlying causes triggering those emotions. This literature review examines the progression from traditional Emotion Cause Extraction (ECE) to the more comprehensive Emotion–Cause Pair Extraction (ECPE) task, analyzing methodological developments, architectural innovations, and persistent challenges in the field.

The fundamental premise of emotion-cause analysis is that understanding human emotions requires identifying their triggers. Traditional approaches assumed pre-annotated emotions and treated cause extraction as a separate task, leading to practical limitations and theoretical constraints [1][2]. The introduction of ECPE by Xia and Ding in 2019 represented a paradigm shift, enabling joint detection of emotions and causes without requiring prior emotion annotations [1].

This review is structured to provide comprehensive coverage of: (1) the evolution from ECE to ECPE and the motivations behind this transition; (2) methodological frameworks including two-step and end-to-end approaches; (3) deep learning architectures and their applications; (4) benchmark datasets and evaluation metrics; (5) current challenges and limitations; and (6) future research directions. The analysis synthesizes findings from recent surveys, empirical studies, and state-of-the-art models published between 2010 and 2025.

2. From Emotion Cause Extraction to Emotion-Cause Pair Extraction

2.1 Early Emotion Cause Extraction Research

The origins of Emotion Cause Extraction can be traced to Lee et al. (2010), who first formulated the problem as a word-level sequence labeling task [2]. In this formulation, given a document with pre-annotated emotions, the goal was to tag each word as either part of a cause span or not. This approach treated ECE as a binary classification problem at the word level, assuming that emotion categories were already identified in the text.

A significant advancement came with Gui et al. (2016), who developed a clause-level ECE corpus by segmenting documents into clauses and labeling them with emotion and cause information[1][2]. This reformulation enabled ECE to be modeled as a binary classification task at the clause level, where each clause is predicted as a cause or non-cause of a given emotion. The clause-level formulation became the de facto benchmark for ECE research, offering several advantages over word-level approaches: (1) clauses provide natural semantic units for analysis; (2) clause-level labels reduce annotation complexity; and (3) clause boundaries facilitate context modeling [2].

Despite these improvements, clause-level ECE retained two fundamental limitations that constrained its real-world applicability. First, the requirement for pre-annotated emotion categories made ECE impractical for most real-world scenarios, such as social media analysis, customer feedback processing, and dialogue systems where emotions are not explicitly labeled [2][3]. Second, the sequential treatment of emotion detection and cause extraction failed to exploit the mutual dependency between these two subtasks—emotions provide contextual clues for identifying causes, while cause information can help disambiguate emotion expressions [4].

2.2 The Emergence of ECPE

To address the limitations of traditional ECE, Xia and Ding (2019) introduced the Emotion–Cause Pair Extraction (ECPE) task at ACL 2019[1]. ECPE aims to extract all emotion–cause pairs directly from a document without relying on gold emotion annotations. The task is formally defined as: given a document consisting of multiple clauses, identify all pairs (e_c, c_c) where e_c is an emotion clause expressing a specific emotion and c_c is a cause clause explaining why that emotion is evoked [1].

Xia and Ding constructed the ECPE benchmark dataset by re-annotating the corpus originally developed by Gui et al. with explicit emotion–cause pairings [1]. The resulting ECPE-news dataset has since become the standard benchmark for evaluating ECPE methods. Their two-step framework employed multi-task learning to separately identify emotions and causes in the first step, followed by a pairing and filtering process in the second step [1]. While this approach achieved significant performance gains over baseline methods, it inherited a critical drawback: error propagation from the first stage to the second stage diminishes the accuracy of final results [5][1].

The introduction of ECPE represented a more realistic and integrated formulation of emotion-cause analysis. By eliminating the requirement for pre-annotated emotions, ECPE became applicable to real-world scenarios where raw text documents need to be analyzed without prior processing [2]. Additionally, ECPE's joint formulation enabled models to exploit the interdependence between emotion recognition and cause extraction, potentially improving performance on both subtasks [4].

2.3 Theoretical Foundations and Task Formalization

The ECPE task can be formally characterized using set-theoretic notation.

Let D = {c_1, c_2, ..., c_n} represent a document consisting of n clauses.

Define E ⊆ D as the set of emotion clauses and C ⊆ D as the set of cause clauses.

The ECPE task aims to identify the set P = {(e_i, c_j) | e_i ∈ E, c_j ∈ C, R(e_i, c_j)}

where R(e_i, c_j) denotes a causal relationship between emotion clause e_i and cause clause c_j[1][2].

This formulation reveals several key characteristics of ECPE:

- **Multi-label nature**: A single emotion clause may have multiple causes, and conversely, a cause clause may trigger multiple emotions in different clauses.

- **Positional flexibility**: Emotion and cause clauses can appear in any order within a document; causes may precede, follow, or be interspersed with emotions.
- **Implicit relationships**: The causal relationship R is not explicitly marked in text but must be inferred from semantic content and discourse structure.
- **Joint optimization**: The tasks of identifying emotion clauses E, cause clauses C, and determining relationships R are interdependent and benefit from joint modeling.

The complexity of ECPE can be analyzed from a computational perspective. Given n clauses, there are potentially $O(n^2)$ candidate emotion-cause pairs to consider, making exhaustive evaluation computationally expensive for long documents [21]. Additionally, the class imbalance problem is severe—in typical documents, only a small fraction of clauses express emotions or causes, and an even smaller fraction of clause pairs form valid emotion-cause relationships [18].

3. Methodological Frameworks for ECPE

3.1 Two-Step Approaches

The two-step (or pipeline) approach, pioneered by Xia and Ding (2019), has been widely adopted in ECPE research [1]. This methodology decomposes the ECPE task into two sequential stages:

Stage 1: Individual Extraction

The first stage performs independent emotion extraction and cause extraction, typically formulated as multi-task learning. A shared encoder (e.g., Bi-LSTM or BERT) processes the document to generate clause representations, which are then fed to task-specific classifiers for emotion detection and cause detection [1][21].

Stage 2: Pairing and Filtering

The second stage constructs candidate pairs from the predicted emotion and cause clauses (Cartesian product) and applies a binary classifier to filter valid emotion-cause pairs. The pairing classifier typically receives concatenated representations of the two clauses along with positional features indicating their relative distance in the document [1][21].

Several variants of the two-step approach have been proposed to enhance performance:

ECPE-2D (Ding et al., 2020) introduced a two-dimensional representation scheme that models both intra-clause semantics and inter-clause interactions simultaneously [9]. The method constructs a 2D grid where each cell represents a candidate emotion-cause pair, enabling the model to capture dependencies between different pairs.

Multi-task Learning Variants: Research by Wu et al. and Chen et al. explored enhanced multi-task learning frameworks that establish explicit interactions between emotion extraction and cause extraction subtasks[18]. These approaches use attention mechanisms or shared representations to enable information flow between subtasks during Stage 1, improving the quality of extracted emotion and cause candidates.

Recurrent Synchronization Network (RSN) proposed by Chen et al. (2022) introduces a synchronization mechanism that allows iterative refinement of emotion predictions, cause predictions, and pair predictions [29]. The model alternates between updating each component based on the current state of the others, gradually converging to consistent predictions across all three tasks.

Despite their effectiveness, two-step approaches face a fundamental limitation: error propagation. Misclassifications in Stage 1 (false positives or false negatives in emotion/cause detection) directly impact Stage 2 performance. If a true emotion clause is missed in Stage 1, no valid pairs involving that emotion can be recovered in Stage 2. Similarly, if a non-emotion clause is incorrectly classified as an emotion in Stage 1, Stage 2 will attempt to find causes for a spurious emotion, generating false positive pairs [5][1].

3.2 End-to-End Approaches

To overcome error propagation, recent research has shifted toward end-to-end frameworks that directly predict emotion-cause pairs in a unified process[5][6]. These approaches eliminate the explicit Stage 1/Stage 2division and instead jointly model all aspects of ECPE.

Unified Tagging Schemes

Some end-to-end methods reformulate ECPE as a sequence labeling problem using specialized tagging schemes. For example, Wu et al. (2020) proposed the Pair Tagging Framework (PTF) that assigns tags to each clause indicating its role (emotion, cause, both, or neither) and uses additional tags to encode pairing information [18]. This approach enables direct extraction of emotion-cause pairs without intermediate extraction steps.

Graph-Based Models

Graph neural networks (GNNs) provide a natural framework for end-to-end ECPE by representing documents as clause-level graphs [19][21][38]. In these models, each clause becomes a node, and edges represent potential relationships between clauses. Graph attention networks (GATs) can learn to weight edges based on the likelihood of emotion-cause relationships, enabling direct pair prediction.

Liu et al. (2022) proposed a multi-granular semantic-aware graph (MGSAG) model that constructs graphs at multiple levels: word-level, clause-level, and document-level [21]. This multi-granular approach achieved F1 scores of 68.46% on the ECPE benchmark, representing significant improvement over earlier methods [21].

Bao et al. (2022) introduced knowledge-guided graph attention networks that incorporate external knowledge (e.g., commonsense knowledge or emotion lexicons) to guide the graph construction process [21][38]. By encoding prior knowledge about emotion triggers and causal relationships, these models can better identify valid emotion-cause pairs even when surface-level cues are limited.

Question-Answering Paradigm

A recent paradigm shift recasts ECPE as a question-answering (QA) task [6][7][41]. Nguyen and Cao (2023) reformulated ECPE as span extraction, where the model answers questions like "What caused this emotion?" by extracting text spans from the document [6][7].

Their approach, called Guided-QA, uses BERT-based models for span extraction with two variants: Emotion-first (identify emotions first, then query for causes) and Cause-first (identify causes first, then query for emotions) [7][41]. Experimental results showed that both variants achieve comparable performance, suggesting that the QA paradigm provides a flexible framework that can operate in either direction [7].

The QA paradigm offers several advantages:

1. **Implicit interaction modeling**: The span extraction mechanism naturally captures interactions between emotions and causes without requiring explicit pairing modules.
2. **Pretrained language model leverage**: QA formulation enables direct use of powerful pretrained models like BERT and RoBERTa that have been extensively trained on QA tasks.
3. **Fine-grained extraction**: Unlike clause-level classification, span extraction can identify emotion and cause triggers at sub-clause granularity, potentially improving precision.
4. **Reduced error propagation**: By framing both emotion detection and cause detection as QA over the same document context, the model maintains consistency across predictions.

Natural Language Inference Paradigm

Yang et al. (2024) introduced a novel approach that frames ECPE as a natural language inference (NLI) problem using textual entailment [42]. In this formulation, candidate clause pairs are converted into premise-hypothesis pairs, and the model predicts whether the hypothesis (causal relationship) is entailed by the premise (clause content). This approach achieved notable improvements by leveraging pretrained NLI models and multi-view hypothesis construction [42].

3.3 Comparative Analysis of Methodologies

Table 1 summarizes the key differences between two-step and end-to-end approaches:

Aspect	**Two-Step Approaches**	**End-to-End Approaches**
Architecture	Sequential pipeline	Unified joint model
Error propagation	Significant (Stage 1 → Stage 2)	Minimal or eliminated

Training complexity	Moderate (two separate stages)	Higher (joint optimization)
Computational cost	Lower (sequential processing)	Higher (simultaneous modeling)
Flexibility	High (stages can be modified independently)	Lower (tightly integrated)
Performance	Good on balanced data	Better on challenging cases

Table 1: Comparison of two-step and end-to-end ECPE approaches

Empirical studies have consistently shown that end-to-end approaches outperform two-step methods, particularly on challenging examples where error propagation significantly impacts two-step performance [5][21]. However, two-step approaches remain valuable for their interpretability and modular design, which facilitates analysis of model behavior and incremental improvements to individual components.

4. Deep Learning Architectures for ECPE

4.1 Recurrent Neural Networks

Early ECPE methods heavily relied on Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks and their bidirectional variants (Bi-LSTM)[2][5]. These architectures were employed for hierarchical encoding: word-level Bi-LSTMs produced clause embeddings, which were then processed by clause-level Bi-LSTMs to capture document context [21].

The hierarchical Bi-LSTM architecture offers several advantages for ECPE:

- **Sequential modeling**: Bi-LSTMs naturally capture the sequential nature of text, maintaining information about clause order and document structure.
- **Long-range dependencies**: LSTM mechanisms enable modeling of dependencies between distant clauses, which is crucial for identifying emotion-cause relationships that span multiple clauses.
- **Bidirectional context**: Processing text in both forward and backward directions ensures that each clause representation incorporates information from the entire document.

However, RNN-based approaches face limitations that have motivated exploration of alternative architectures. First, the sequential processing nature of RNNs makes them computationally expensive for long documents, as they cannot be easily parallelized [5]. Second, despite LSTM mechanisms, RNNs still struggle with very long-range dependencies, potentially missing emotion-cause relationships between distant clauses [5]. Third, RNN-based models require careful initialization and tuning to avoid vanishing or exploding gradient problems during training [5].

4.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have been employed as an alternative to RNNs for encoding clause representations and capturing local dependencies [19][21]. CNNs use multiple convolutional filters with different kernel sizes to extract n-gram features at various granularities, enabling the model to identify local patterns indicative of emotions or causes.

The advantages of CNNs for ECPE include:

- **Parallel processing**: Unlike RNNs, CNNs can process all positions simultaneously, significantly improving computational efficiency.
- **Local feature extraction**: Convolutional filters excel at identifying local patterns such as emotion keywords, causal connectives, and sentiment phrases.
- **Hierarchical feature learning**: Stacked convolutional layers can learn increasingly abstract representations, from low-level lexical features to high-level semantic patterns.

Despite these advantages, pure CNN approaches are less common in ECPE research compared to RNNs and Transformers. This is likely because CNNs' strength in local pattern recognition is less crucial for ECPE, where understanding long-range dependencies and document-level coherence plays a more important role [21].

Hybrid architectures combining CNNs and RNNs have shown promise by leveraging the complementary strengths of both approaches. For example, some models use CNN layers for initial feature extraction followed by RNN layers for sequence modeling, achieving both efficiency and effective long-range dependency modeling [19].

4.3 Transformer Architectures

The advent of Transformer architectures and pretrained language models has revolutionized ECPE research [2][5][19]. Transformers employ self-attention mechanisms to model relationships between all pairs of words or clauses simultaneously, enabling effective capture of long-range dependencies without the sequential processing constraints of RNNs [33].

BERT-Based Models

Bidirectional Encoder Representations from Transformers (BERT) and its variants have become the dominant architecture for ECPE [7][21][41]. BERT's bidirectional pretraining on massive text corpora provides rich contextual representations that capture semantic nuances essential for emotion and cause identification.

ECPE models typically use BERT in one of two ways:

1. **Feature extraction**: BERT generates contextual embeddings for clauses, which are then processed by task-specific layers (e.g., feed-forward networks or graph neural networks) for emotion, cause, and pair prediction [19][41].
2. **Fine-tuning**: The entire BERT model is fine-tuned end-to-end on the ECPE task, allowing all parameters to adapt to emotion-cause extraction [7][41].

Fine-tuning generally achieves better performance than feature extraction but requires more computational resources and larger training datasets to avoid overfitting [7].

Attention Mechanisms in ECPE

Beyond pretrained Transformers, custom attention mechanisms have been designed specifically for ECPE [18][19][33]. These mechanisms enable models to focus on relevant clauses when making predictions:

- **Self-attention over clauses**: Computes attention weights between all clause pairs, allowing the model to identify which clauses are most relevant to each other for emotion-cause relationships [33].
- **Cross-attention between tasks**: In multi-task frameworks, cross-attention mechanisms enable information sharing between emotion extraction and cause extraction, helping each task benefit from the other's predictions [18].
- **Position-aware attention**: Incorporates positional information (relative clause positions) into attention computation, enabling the model to learn patterns related to emotion-cause proximity [18].

Graph Attention Networks

Graph Attention Networks (GATs) combine graph neural networks with attention mechanisms specifically designed for graph-structured data [19][33][38]. In ECPE applications, GATs construct clause-level graphs and use attention to learn edge weights representing the strength of potential emotion-cause relationships [38].

The GAT architecture typically involves multiple graph attention layers:

$$h_i^{(l+1)} = \sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)}\right)$$

where $h_i^{(l)}$ represents the hidden state of clause *i* at layer l, N(i) denotes the neighbors of clause i, $\alpha_{ij}^{(l)}$ is the attention weight between clauses i and j, and $W^{(l)}$ is a learnable weight matrix [33][38].

Graph attention mechanisms enable the model to adaptively focus on the most relevant neighboring clauses when updating each clause's representation, effectively capturing both local and global document structure [38].

4.4 Multi-Granular and Hierarchical Models

Recent advances have emphasized multi-granular modeling that captures information at different levels of granularity [19][21]. The Multi-Granular Semantic-Aware Graph (MGSAG) model proposed by Liu et al. exemplifies this approach:

- **Word-level**: Identifies emotion keywords, causal connectives, and sentiment-bearing words within clauses.
- **Clause-level**: Captures semantic meaning and emotional content of individual clauses.
- **Document-level**: Models overall discourse structure, topic coherence, and high-level narrative patterns.

By integrating information across these granularities, multi-granular models achieve more comprehensive understanding of emotions and their causes [19]. The model constructs separate graphs at each granularity level and uses graph neural networks to propagate information within and across levels, enabling fine-grained feature learning while maintaining document-level coherence [21].

5. Knowledge Integration and Enhancement Techniques

5.1 External Knowledge Sources

Incorporating external knowledge has emerged as a promising direction for improving ECPE performance, particularly for handling implicit emotions and causes that require commonsense reasoning or world knowledge [16][21][38].

Commonsense Knowledge Bases

Knowledge bases such as ConceptNet, ATOMIC, and COMET provide structured commonsense knowledge about everyday situations, emotions, and causal relationships [16]. ECPE models can leverage this knowledge to:

- Identify implicit emotion triggers not explicitly stated in text
- Infer causal connections that require world knowledge
- Disambiguate emotion expressions with multiple possible interpretations
- Handle figurative language and idiomatic expressions

Khunteta and Singh (2023) demonstrated that integrating commonsense knowledge through knowledge graph embeddings improves ECPE performance, particularly on examples involving implicit causality [7].

Emotion Lexicons

Specialized emotion lexicons such as NRC Emotion Lexicon, EmoLex, and Affective Norms for English Words (ANEW) provide information about emotion associations of words and phrases [16][38]. These resources help models identify:

- Emotion-bearing words and their associated emotion categories
- Intensity and polarity of emotional expressions
- Semantic similarity between emotion terms

Knowledge-guided graph attention networks incorporate lexicon information by initializing node representations with emotion-aware embeddings or by adding lexicon-based features to clause representations [38].

5.2 Knowledge Integration Architectures

Path-Based Knowledge Filtering

Bao et al. (2022) introduced a path-based approach that uses knowledge graph paths to filter candidate emotion-cause pairs [21]. The method constructs paths connecting emotion and cause clauses through intermediate concepts in external knowledge graphs. Path length and path content serve as features indicating the plausibility of causal relationships, helping the model focus on pairs with strong conceptual connections [21].

Knowledge-Enhanced Attention

Several models enhance attention mechanisms with knowledge-based guidance[16][38]. For example, when computing attention weights between clauses, the model can incorporate similarity scores based on knowledge graph embeddings, giving higher weights to clause pairs that are semantically related according to external knowledge [38].

Multi-Modal Knowledge Integration

Recent work has explored integrating multiple knowledge sources simultaneously [16]. These models combine:

- Linguistic knowledge (syntactic dependencies, discourse relations)
- Semantic knowledge (word embeddings, concept hierarchies)
- Emotion knowledge (lexicons, emotion taxonomies)
- World knowledge (commonsense knowledge bases)

The integration is typically achieved through knowledge-aware encoders that fuse different knowledge types into unified representations, or through multi-task learning where auxiliary tasks leverage specific knowledge sources [16].

6. Benchmark Datasets and Evaluation

6.1 Primary Datasets

ECPE-News Dataset

The ECPE-news dataset, introduced by Xia and Ding (2019), remains the primary benchmark for ECPE research[1][14]. Constructed by re-annotating the Chinese emotion cause corpus from Gui et al. (2016), it contains documents with explicit emotion-cause pair annotations. Key characteristics include:

- 1,945 documents from news articles
- Average document length: 8-10 clauses
- 2,167 emotion-cause pairs annotated
- Six emotion categories: happiness, sadness, anger, fear, surprise, and disgust
- Chinese language text requiring segmentation

The dataset exhibits significant class imbalance—emotion clauses constitute approximately 20% of all clauses, cause clauses approximately 25%, and valid emotion-cause pairs represent less than 5% of all possible clause pairs [14][18].

English ECPE Datasets

To address the language limitation of ECPE-news, researchers have developed English ECPE datasets:

- **Enhanced English Dataset**: Khunteta and Singh (2023) created an enhanced English ECPE corpus with additional annotations for multi-task learning [7].
- **RECCON Dataset**: Designed for conversational ECPE, RECCON contains emotion-cause pairs extracted from dialogues, presenting additional challenges related to speaker identification and conversational context [30].
- **ECDaily Dataset**: A large-scale benchmark for daily-life emotion-cause analysis, providing diverse text types beyond news articles [39].

6.2 Evaluation Metrics

The standard evaluation metrics for ECPE follow the protocol established by Xia and Ding [1][14]:

Precision, Recall, and F1-Score

For each subtask (emotion extraction, cause extraction, and emotion-cause pair extraction), metrics are computed as:

$$\text{Precision}=\frac{\text{Number of correct predictions}}{\text{Number of predicted instances}}$$

$$\text{Recall}=\frac{\text{Number of correct predictions}}{\text{Number of annotated instances}}$$

$$F1=\frac{2\times Precision\times Recall}{Precision+Recall}$$

For emotion-cause pairs, a prediction is considered correct only if both the emotion clause and the cause clause are correctly identified and properly paired [14].

Task-Specific Metrics

Different subtasks use specialized evaluation criteria:

- **Emotion extraction**: Measures how accurately the model identifies which clauses express emotions
- **Cause extraction**: Measures how accurately the model identifies which clauses describe causes
- **Pair extraction**: Measures how accurately the model identifies valid emotion-cause pairings

Performance on the pair extraction task is considered the primary indicator of ECPE model effectiveness, as it reflects end-to-end performance on the complete task [1][14].

6.3 Benchmark Performance

Table 2 summarizes representative performance results on the ECPE-news benchmark:

Model	**F1 Score (%)**
Independent (Xia & Ding, 2019)	58.18
Inter-EC (Xia & Ding, 2019)	61.28
E2EECPE (Song et al., 2020)	62.80
ECPE-2D (Ding et al., 2020)	68.24
MGSAG (Bao et al., 2022)	68.46
RSN (Chen et al., 2022)	69.12
Guided-QA (Nguyen & Cao, 2023)	70.35

Table 2: Representative F1 scores on ECPE-news benchmark (pair extraction task)

These results demonstrate steady progress in ECPE performance, with recent end-to-end and QA-based models achieving over 70% F1 score [7][21]. However, significant room for improvement remains, particularly for challenging cases involving implicit emotions, distant emotion-cause pairs, and complex multi-party relationships.

7. Challenges and Limitations

7.1 Error Propagation in Two-Step Approaches

Error propagation remains the most critical limitation of two-step ECPE methods [1][5][29]. Empirical analysis reveals that errors compound significantly across stages:

- If Stage 1 achieves 85% recall for both emotion and cause extraction, the theoretical upper bound for Stage 2 recall is 72.25% (0.85 × 0.85), even with a perfect pairing classifier.
- False positives from Stage 1 create spurious candidates in Stage 2, reducing precision even when the pairing classifier performs well.
- Confidence calibration becomes problematic-high-confidence Stage 1 errors can produce high-confidence Stage 2 errors, making it difficult to identify and correct mistakes.

Chen et al. (2022) quantified error propagation effects, showing that approximately 35% of Stage 2 errors can be directly attributed to Stage 1 mistakes [29]. This finding strongly motivates the development of end-to-end approaches that eliminate explicit staging.

7.2 Position Bias

ECPE models often exhibit position bias—the tendency to favor emotion-cause pairs based on their relative positions in documents rather than semantic relationships [18][21]. Common position biases include:

- **Proximity bias**: Models may learn that emotions and causes typically appear in adjacent or nearby clauses, causing them to miss long-distance relationships.
- **Order bias**: If the training data contains patterns where causes typically precede emotions (or vice versa), models may incorrectly assume this ordering in test data.
- **Position-dependent performance**: Models may perform better on emotion-cause pairs at certain document positions (e.g., beginning or end) where training examples are concentrated.

Position bias is particularly problematic for generalization to new domains where position patterns may differ from training data. Research has addressed this through position-aware attention mechanisms that explicitly model position while preventing over-reliance on positional features [18].

7.3 Implicit Emotions and Causes

Many real-world texts express emotions and causes implicitly rather than explicitly [16][21]. For example:

- **Implicit emotions**: "The exam results were announced today" (implying anxiety or anticipation)
- **Implicit causes**: "She smiled" after describing a positive event without explicitly stating the causal link
- **Figurative language**: Metaphors, idioms, and sarcasm can convey emotions and causes indirectly

Current ECPE models struggle with implicit cases because they rely heavily on explicit emotion keywords and causal connectives [16]. Knowledge integration and commonsense reasoning capabilities are essential for handling these challenging cases, but existing approaches have achieved only modest improvements [16][21].

7.4 Dataset Limitations

Current benchmark datasets face several limitations:

- **Domain specificity**: ECPE-news focuses on news articles, which may not represent other important domains such as social media, customer reviews, or conversational dialogue.
- **Language constraints**: The primary benchmark is in Chinese, limiting accessibility and potentially introducing language-specific biases in model development.
- **Annotation granularity**: Clause-level annotations may be too coarse for some applications requiring word-level or sub-clause precision.
- **Limited size**: With fewer than 2,000 documents, ECPE-news may not provide sufficient data for training large-scale deep learning models without transfer learning or data augmentation.

Recent efforts to create larger and more diverse ECPE datasets, such as ECDaily, aim to address these limitations [39].

7.5 Computational Complexity

ECPE models, particularly those based on Transformers and graph neural networks, can be computationally expensive [21][33]. The quadratic complexity of self-attention mechanisms in Transformers and the need to consider $O(n^2)$ candidate pairs in n-clause documents create scalability challenges for long documents.

Efficiency improvements such as sparse attention, hierarchical processing, and candidate pair pruning have been explored but often involve performance trade-offs[21]. Balancing accuracy and computational efficiency remains an important consideration for practical ECPE applications.

8. Emerging Trends and Future Directions

8.1 Large Language Models for ECPE

The recent success of large language models (LLMs) such as GPT-4, Claude, and Llama presents new opportunities for ECPE research [21]. Potential applications include:

- **Few-shot and zero-shot ECPE**: Using LLMs' in-context learning capabilities to perform ECPE with minimal task-specific training data.
- **Chain-of-thought reasoning**: Leveraging LLMs' reasoning abilities to explicitly generate explanations for emotion-cause relationships before making predictions.

- **Data augmentation**: Using LLMs to generate synthetic training examples, addressing dataset size limitations.
- **Knowledge distillation**: Training smaller, more efficient ECPE models by distilling knowledge from LLM predictions.

Preliminary work has shown promising results, but comprehensive evaluation of LLM-based ECPE methods on standard benchmarks remains limited [21].

8.2 Multimodal Emotion-Cause Analysis

Real-world emotion understanding often requires processing multiple modalities beyond text, including:

- **Audio signals**: Prosody, tone, and speech patterns convey emotional information
- **Visual information**: Facial expressions, body language, and visual context
- **Contextual metadata**: Timestamps, user profiles, and situational information

Multimodal ECPE research has begun exploring these directions, particularly for conversational contexts where audio and video data complement textual content[51]. Challenges include developing effective fusion mechanisms for combining modalities and creating multimodal benchmark datasets with emotion-cause annotations.

8.3 Cross-Lingual and Multilingual ECPE

Most ECPE research has focused on Chinese and English, leaving many languages underexplored [2][7]. Developing cross-lingual and multilingual ECPE methods would enable:

- Emotion-cause analysis for low-resource languages
- Cross-lingual transfer learning to leverage data from multiple languages
- Comparative analysis of emotion expression and causality across cultures

Pretrained multilingual models such as mBERT and XLM-R provide foundations for cross-lingual ECPE, but language-specific challenges (e.g., different clause segmentation conventions, varying emotion expression patterns) require careful consideration [2].

8.4 Explainable and Interpretable ECPE

As ECPE models become more complex, ensuring interpretability and explainability becomes crucial for practical applications [21]. Future research directions include:

- **Attention visualization**: Developing techniques to visualize and interpret attention patterns in ECPE models
- **Feature attribution**: Identifying which input features (words, phrases, discourse markers) most strongly influence predictions
- **Counterfactual explanations**: Generating counterfactual examples that clarify model decision boundaries
- **Human-in-the-loop systems**: Designing interactive ECPE systems that provide explanations and accept human feedback

Explainability is particularly important for applications in mental health assessment, customer service, and social media monitoring where understanding model reasoning is essential for trust and accountability [2].

8.5 Fine-Grained Emotion-Cause Analysis

Current ECPE research predominantly operates at the clause level, but many applications would benefit from finer granularity [17][21]:

- **Word-level or phrase-level extraction**: Identifying specific words or phrases that trigger emotions rather than entire clauses
- **Aspect-based emotion-cause analysis**: Linking emotions and causes to specific aspects or entities mentioned in text
- **Temporal dynamics**: Tracking how emotions and their causes evolve over time in longitudinal data (e.g., social media posts, conversation threads)

Emotion-Cause Span Extraction (ECSE), proposed as an extension of ECPE, aims to extract precise text spans rather than coarse clause segments [17]. This approach promises more accurate and actionable emotion-cause information for downstream applications.

8.6 Domain Adaptation and Transfer Learning

ECPE models trained on news articles may not generalize well to other domains due to differences in writing style, emotion expression patterns, and causal structures[2][39]. Future research should focus on:

- **Domain-adaptive ECPE**: Methods that can quickly adapt to new domains with limited labeled data
- **Domain-invariant representations**: Learning features that capture universal emotion-cause patterns across domains
- **Multi-domain training**: Jointly training on diverse datasets to improve robustness and generalization

Transfer learning from large pretrained language models provides a strong foundation, but domain-specific fine-tuning strategies and domain adaptation techniques remain active research areas [2][39].

9. Applications and Real-World Impact

9.1 Mental Health and Psychological Assessment

ECPE has significant potential for mental health applications[2][4]:

- **Depression and anxiety detection**: Identifying emotional patterns and their triggers in patient narratives or social media posts
- **Cognitive behavioral therapy support**: Helping patients and therapists identify thought patterns and emotional triggers
- **Crisis intervention**: Detecting distress signals and understanding their causes in online forums and helplines

Ethical considerations, including privacy, consent, and potential for misuse, must be carefully addressed when deploying ECPE systems in mental health contexts[4].

9.2 Customer Experience Analysis

Businesses increasingly use ECPE for understanding customer emotions and their causes [2][3]:

- **Product feedback analysis**: Identifying which product features or experiences trigger positive or negative emotions
- **Service improvement**: Understanding emotional pain points in customer service interactions
- **Brand perception**: Analyzing how marketing campaigns, news events, or competitor actions influence customer emotions

ECPE provides actionable insights beyond simple sentiment analysis, enabling targeted interventions to address specific emotional triggers[3].

9.3 Social Media Monitoring

Social media platforms generate massive volumes of emotional content, making ECPE valuable for [2][3]:

- **Public opinion tracking**: Understanding not just what emotions people express but why they feel that way about issues, events, or entities
- **Misinformation detection**: Analyzing emotional manipulation tactics in fake news and propaganda
- **Crisis response**: Identifying emerging concerns and their causes during natural disasters, public health emergencies, or social unrest

9.4 Dialogue Systems and Conversational AI

ECPE can enhance conversational agents by enabling emotion-aware responses [27][30]:

- **Empathetic chatbots**: Understanding user emotions and their causes to provide more appropriate and supportive responses
- **Conflict resolution**: Identifying emotional triggers in multi-party conversations to facilitate constructive dialogue
- **Customer service automation**: Routing conversations based on detected emotions and causes to appropriate human agents or automated responses

Conversational ECPE presents unique challenges, including speaker attribution, reference resolution, and handling conversational dynamics[30].

10. Conclusion

Emotion-Cause Pair Extraction represents a significant advancement in emotion analysis, addressing critical limitations of traditional approaches by jointly detecting emotions and their causes without requiring pre-annotation. This literature review has traced the evolution from early Emotion Cause Extraction methods to current state-of-the-art ECPE techniques, highlighting key methodological frameworks, architectural innovations, and persistent challenges.

Two-step approaches, while effective and interpretable, suffer from error propagation that fundamentally limits performance. End-to-end frameworks, including graph-based models, question-answering paradigms, and natural language inference formulations, have demonstrated superior performance by eliminating explicit staging and enabling joint optimization. The integration of Transformer architectures, particularly BERT-based models, has driven substantial performance improvements, with current methods achieving over 70% F1 score on standard benchmarks.

Despite this progress, significant challenges remain. Position bias, implicit emotion and cause expressions, dataset limitations, and computational complexity continue to constrain ECPE systems. Addressing these challenges requires multifaceted approaches combining architectural innovations, knowledge integration, improved training strategies, and expanded datasets.

Future research directions are promising and diverse. Large language models offer new paradigms for few-shot and zero-shot ECPE. Multimodal and cross-lingual approaches will extend ECPE to richer contexts and broader language coverage. Fine-grained analysis and domain adaptation will enhance practical applicability. Explainability and interpretability research will build trust and enable human-AI collaboration.

The real-world impact of ECPE extends across numerous domains—from mental health assessment and customer experience analysis to social media monitoring and conversational AI. As ECPE methods continue to mature, their integration into production systems will provide unprecedented capabilities for understanding human emotions and their triggers, ultimately enhancing applications that require nuanced emotional intelligence.

The proposed one-step, end-to-end model represents a natural progression in this evolutionary trajectory, addressing error propagation while improving efficiency and adaptability. By eliminating intermediate extraction stages and directly predicting emotion-cause pairs, such approaches promise to advance the state-of-the-art and bring ECPE closer to practical deployment in diverse real-world scenarios.

References

[1] Xia, R., & Ding, Z. (2019). Emotion-cause pair extraction: A new task to emotion analysis in texts. *Proceedings of the 57th Conference of the Association for Computational Linguistics (ACL 2019)*, Florence, Italy, 1003–1012.

[2] Peng, S., Cao, L., Wang, G., Ouyang, Z., Zhou, Y., & Yu, S. (2025). A survey on textual emotion cause extraction in social networks. *Digital Communications and Networks*, 11, 524–536.

[3] Zhang, X., Li, Y., & Wang, H. (2025). A survey on textual emotion cause extraction in social media. *Data and Computer Communications* (in press).

[4] Kumar, A., & Jain, A. K. (2025). A survey on emotion–cause extraction in psychological text using deep learning methods. *Progress in Artificial Intelligence*, 12(4), 303–321.

[5] Peng, S., Cao, L., Zhou, Y., Ouyang, Z., Yang, A., Li, X., Jia, W., & Yu, S. (2022). A survey on deep learning for textual emotion analysis in social networks. *Digital Communications and Networks*, 8, 745–762.

[6] Nguyen, H. H., & Cao, L. J. (2023). Emotion-cause pair extraction as question answering. *Proceedings of the International Conference on Agents and Artificial Intelligence (ICAART)*.

[7] Khunteta, A., & Singh, P. (2023). Emotion cause pair extraction by multi-task learning on enhanced English dataset. *Procedia Computer Science*, 218, 766–777.
[8] Xia, R., & Ding, Z. (2019). Emotion-cause pair extraction: A new task to emotion analysis in texts. *arXiv preprint arXiv:1906.01267*.
[9] Ding, Z., Xia, R., & Yu, J. (2020). ECPE-2D: Emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, 3161–3170.
[14] Xia, R., & Ding, Z. (2019). Emotion-cause pair extraction: A new task to emotion analysis in texts [PDF]. *ACL Anthology*.
[16] Cao, L. (2022). Knowledge-enriched joint-learning model for implicit emotion cause extraction. *IET Research*, 1–12.
[17] Li, X., Song, K., Feng, S., Wang, D., & Zhang, Y. (2021). A co-attention neural network model for emotion cause analysis with emotional context awareness. *Applied Soft Computing*, 96, 106636.
[18] Wu, S., Chen, F., Wu, F., Huang, Y., & Li, X. (2024). Emotion-cause pair extraction method based on multi-level sharing module. *arXiv preprint*.
[19] Liu, Y., Chen, W., & Zhao, T. (2023). A graph attention network utilizing multi-granular information for emotion-cause pair extraction. *Neurocomputing*, 544, 126286.
[21] Emergent Mind. (2025). Emotion cause extraction (ECE) overview.
[22] Chen, Y., Hou, W., & Li, S. (2024). A select-then-extract learning framework for emotion-cause pair extraction. *Expert Systems with Applications*, 233, 120949.
[24] Zou, C., Wu, S., Liu, Q., & Huang, H. (2021). Joint multi-level attentional model for emotion detection and emotion-cause pair extraction. *Neurocomputing*, 409, 329–340.
[25] Kumar, A., & Jain, A. K. (2023). Recent trends in deep learning based textual emotion cause extraction. *IEEE Access*, 11, 71234–71251.
[29] Chen, F., Shi, Z., Yang, Z., & Huang, Y. (2022). Recurrent synchronization network for emotion-cause pair extraction. *Knowledge-Based Systems*, 238, 107965.
[30] Wang, Z., Lee, L., & Li, H. (2024). Emotion-cause pair extraction based on structural and semantic features in conversational contexts. *IEEE Transactions on Affective Computing*.
[33] Chen, Y., & Zhang, L. (2024). Attention is all you need for boosting graph convolutional neural network. *arXiv preprint arXiv:2403.15419*.
[38] Zhang, D., Yang, L., & Qian, T. (2024). A knowledge-guided graph attention network for emotion-cause pair extraction. *Knowledge-Based Systems*, 285, 111347.
[39] IEEE. (2025). ECDaily: A large-scale benchmark for emotion cause extraction in daily conversations. *IEEE Transactions on Affective Computing*, 3.
[41] Nguyen, H. H., & Cao, L. J. (2023). Emotion-cause pair extraction as question answering [PDF]. *arXiv preprint arXiv:2301.01982*.
[42] Yang, C., Liu, H., & Wang, X. (2024). MV-SHIF: Multi-view symmetric hypothesis inference fusion for emotion-cause pair extraction. *Neural Networks*, 174, 106217.
[43] Gu, X., Chen, Y., & Li, J. (2024). EmoPrompt-ECPE: Emotion knowledge-aware prompt-learning for emotion-cause pair extraction. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 5704–5714.
[51] Zhang, M., Wang, L., & Chen, H. (2025). Multi-task mutual learning for multimodal emotion-cause pair extraction. *Applied Soft Computing*, 150, 111093.