Emotion Recognition in Conversation using Probabilistic Soft Logic

Eriq Augustine
UC Santa Cruz
eaugusti@ucsc.edu

Alon Albalak
UC Santa Barbara
alon_albalak@ucsb.edu

Anurag Prakash
UC Santa Cruz
aprakas6@ucsc.edu

Connor Pryor
UC Santa Cruz
cfpryor@ucsc.edu

William Wang
UC Santa Barbara
william@cs.ucsb.edu

Lise Getoor
UC Santa Cruz
getoor@ucsc.edu

Abstract

Creating agents that can both appropriately respond to conversations and understand complex human linguistic tendencies and social cues has been a long standing challenge in the NLP community. A recent pillar of research revolves around emotion recognition in conversation (ERC); a sub-field of emotion recognition that focuses on conversations or dialogues that contain two or more utterances. In this work, we explore a new method that exploits the use of complex structure in dialogues. We implement our approach in a framework called Probabilistic Soft Logic (PSL), a declarative templating language that uses first order like logical rules, that when combined with data, define a particular class of graphical model. Additionally, PSL allow for the incorporation of neural networks into PSL models. This allows our model to take advantage of advanced neural methods, such as sentence embeddings, and logical reasoning over the structure of a dialogue. We compare out method with state-of-the-art purely neural ERC systems, and, surprisingly, see almost a 20% improvement. With these results, we provide an extensive qualitative and quantitative analysis over the DailyDialog dataset.

1 Introduction

With the growing popularity of conversational agents in daily life, the need for agents that can appropriately respond to long running conversations and that can understand complex human linguistic tendencies and social cues is becoming increasingly important. This growth in popularity has sparked a large interest in conversational research. A recent pillar of this emerging field has been around emotion recognition in conversation (ERC); a subfield of emotion recognition that focuses on conversations or dialogues that contain two or more utterances. Poria et al. (2019b) provides a thorough overview of the current state of ERC. For example, Figure 1 represents a typical conversation of two friends visiting the empire state building where the speakers express surprise, neutral emotion, and then happiness. Being able to correctly identify the emotion of utterances can aid downstream tasks such as emotion-aware dialogue agents (Polzin and Waibel, 2000; André et al., 2004; Skowron, 2010; Skowron et al., 2011; Ghandeharioun et al., 2019; Ekbal, 2020) and healthcare (Tanana et al., 2021; Mowafey and Gardner, 2012; Ghandeharioun et al., 2019).

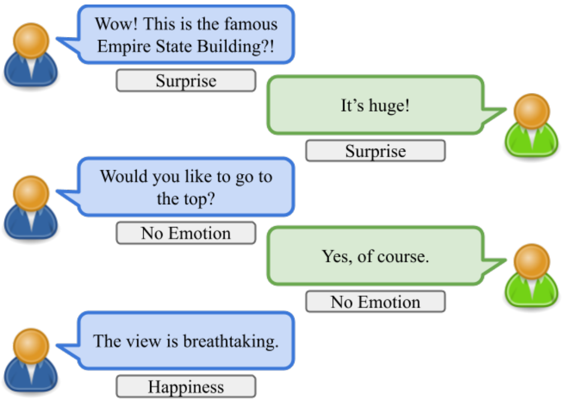

Wow! This is the famous Empire State Building?!
Surprise
It's huge!
Surprise
Would you like to go to the top?
No Emotion
Yes, of course.
No Emotion
The view is breathtaking.
Happiness

Figure 1: A sample conversation with representative emotions. Each utterance is labeled with a single emotion, or marked as having no emotion.

ERC stands out as a challenging problem because it combines the already difficult task of emotion recognition with the complexity of conversations. Conversations are distinctly intricate because they are influenced by a variety of factors such as topic, personality, argumentation logic, viewpoint, intent, location, number of speakers, and the mental and emotional states of the participants at the time of the conversation (Hovy, 1987; Schlöder and Fernández, 2015; Ghosal et al., 2020). In addition to the complexity of conversations, ERC also have to address a number of challenges stemming from emotions, such as bias in emotion annotations, emotional shift, and emotional reasoning (Poria et al., 2019b).

In this paper, we propose a general framework that uses the structure intrinsic to dialogue to aid in utterance emotion prediction. Throughout this paper we develop our method using a framework called Probabilistic Soft Logic (Bach et al., 2017), a declarative templating language that uses first order like logical rules, that when combined with data, define a particular class of graphical model. PSL provides a simple framework for incorporating structural conversational knowledge through first order logical rules, provides efficient and scalable statistical inference, and has shown to be effective in complex domains that benefit from collective inference (Tomkins et al., 2017; Kouki et al., 2019; Sridhar and Getoor, 2019; Embar et al., 2020). Furthermore, recent advances in PSL allow for the integration of neural networks into PSL models. This integration of neural methods with logical reasoning falls under the field of neural symbolic computing (NeSy) (Besold et al., 2017).

Our key contributions are as follows: 1) we create a general and extendable framework for ERC using PSL that can be applied to various ERC datasets, 2) we provide a through experimental evaluation over a popular ERC dataset, DailyDialog, 3) we show both qualitative and quantitatively that PSL outperforms the state-of-the-art models by almost 20%, and 4) we provide a qualitative exploration of the DailyDialog dataset, in which we highlight areas of potential improvement. To the best of our knowledge, our proposed framework is the first neural-symbolic approach to explicitly model structural social emotion recognition rules into conversational emotion classification.

2 Related Work

The broader task of emotion recognition has been a long standing problem across many fields of research, including machine learning, signal processing, social cognitive psychology, etc. The techniques used in emotion recognition heavily overlap with the related problems of sentiment analysis and opinion mining (Pang and Lee, 2008). All of these problems share the common goal of extracting the thoughts, feelings, and opinions of others. However, where sentiment analysis considers a person's feelings towards an entity, emotion recognition focuses more broadly on the emotion that a person feels, regardless of the target of that emotion. Additionally, sentiment analysis is typically performed on more formal text sources, such as written reviews, whereas ERC is typically performed on dialogues which are less formal and more causal in nature.

ERC has become more popular recently with the release of public conversational datasets such as social media conversations and movie/tv-show scripts (Zahiri and Choi, 2018; Poria et al., 2019a). Recent work in ERC focuses on solving the problem with deep learning architectures. One of the earliest networks to produce promising results for ERC was a bi-directional contextual LSTM model, bc-LSTM or CNN-cLSTM (Poria et al., 2017), which allowed utterances to get information from subsequent or earlier utterances. To improve upon this concept, Conversational Memory Networks (CMN) (Hazarika et al., 2018b) utilizes distinct memory for each speaker to model speaker specific information. This method was further improved by Interactive Conversational Memory Networks (ICON) (Hazarika et al., 2018a) and Interaction-aware Attention Networks (IAN) (Yeh et al., 2019), where memories were inter-connected. DialogueRNN (Majumder et al., 2019) expands on the previous methods by using Gated Recurrent Units (GRU) (Chung et al., 2014) as memory cells and is specifically modeled to exploit the speaker information. Further, DialogueGCN (Ghosal et al., 2019) and ConGCN (Zhang et al., 2019) utilize graph convolutional networks (GCN) (Defferrard et al., 2016), and model both context-sensitive and speaker-sensitive dependence for emotion detection. Additionally, KET (Zhong et al., 2019) and COSMIC (Ghosal et al., 2020) attempt to improve results by using external commonsense knowledge, while BERT DCR-Net (Qin et al., 2020) and BERT+MTL (Li et al., 2020) use BERT (Devlin et al., 2019) based features to aid in sentiment recognition. Finally, CESTa (Wang et al., 2020) models the ERC task as sequence tagging and uses conditional random fields (CRF) (Sutton et al., 2007) to model the emotional consistency in conversation.

In this paper, we utilize techniques from both the ERC and neural-symbolic computing (NeSy) communities. Neural-symbolic computing aims to incorporate logic-based reasoning with neural computation (d'Avila Garcez et al., 2019). This integration allows for interpretability and reasoning through symbolic knowledge, while providing robust learning and efficient inference with neural networks. Besold et al. (2017) and Raedt et al. (2020) provide good surveys of recent work in this field.

3 Probabilistic Soft Logic

Probabilistic Soft Logic (PSL) is a probabilistic programming language used to define a special class of Markov random fields (MRF), a hinge-loss Markov random (HL-MRF) (Bach et al., 2017). HL-MRFs are a class of conditional probabilistic models over continuous variables which allow for scalable and exact inference (Bach et al., 2013).

PSL models relational dependencies and structural constraints using weighted first-order logical clauses, referred to as *rules*. For example, consider the rule:

$$\begin{aligned} w :& \textsc{HasEmotion}(\texttt{Utterance1}, \texttt{Emotion}) \\ &\wedge \textsc{SimilarText}(\texttt{Utterance1}, \texttt{Utterance2}) \\ &\rightarrow \textsc{HasEmotion}(\texttt{Utterance2}, \texttt{Emotion}) \end{aligned}$$

where the predicates HASEMOTION and SIMILARTEXT respectively predict the emotional label for an utterance and define the similarity of two utterances, and w acts as a learnable weight for the rule that denotes the rule's relative importance in the model. This rule encodes the domain knowledge that utterances with similar texts (`Utterance1` and `Utterance2`) should probably be labeled with the same emotion, and establishes a dependency that similar utterances should share similar labels.

Given the rules for a model and data, PSL generates an HL-MRF by instantiating concrete instances of each rule where variables are replaced with actual entities from the data. This process is referred to as *grounding*, and each concrete instance of a rule is referred to as a *ground rule*. The logical atoms in the ground rules correspond to the random variables in the HL-MRF, while ground rules correspond to potential functions in the HL-MRF.

Given the observed variables X, unobserved variables Y, and potential functions, PSL defines a probability distribution over the unobserved variables as:

$$P(Y|X) = \frac{1}{Z(Y)} exp(-\sum_{i=1}^{m} w_i \phi_i(Y, X))$$

$$Z(Y) = \int_Y exp(-\sum_{i=1}^{m} w_i \phi_i(Y, X))$$

where m is the number of potential functions, ϕ_i is the i^{th} *hinge-loss potential* function, and w_i is weight of the template rule from which ϕ_i was derived. The hinge-loss potentials are defined as:

$$\phi(Y, X) = [max(0, l(Y, X))]^p$$

where l is a linear function, X and Y are in the range $[0, 1]$, and $p \in 1, 2$ optionally squares the potential.

Exact *maximum a posteriori* (MAP) inference on this distribution can be framed as the convex optimization problem:

$$\begin{aligned} Y^* &= \arg\min_Y \sum_{i=1}^{m} w_i \phi_i(Y, X) \\ &= \arg\min_Y L_{map}(w, X, Y) \end{aligned}$$

PSL uses ADMM (Boyd et al., 2010) to efficiently solve MAP inference.

3.1 Neural PSL

Neural PSL is an extension of PSL that integrates neural networks into PSL. This allows PSL to incorporate powerful neural techniques with PSL's logical inference. In Neural PSL, neural networks can be incorporated into the model as predicates. PSL does not restrict the number of networks that can be used in a single model, and places few restrictions on the types of network architectures that can be supported[1]. This flexibility allows for multiple neural networks of different architectures to be connected in a single joint model.

Neural PSL uses a method called *alternating inference* to both train the neural network and solve the HL-MRF's MAP problem. Intuitively, Neural PSL first trains the neural networks on observed data and makes initial predictions. These predictions are then incorporated as observations into the the MAP problem where inference is performed in the presence of additional relational information and constraints. The solution to the MAP problem is then used as labels to retrain the neural networks. This process of alternating between training the neural network and solving the HL-MRF's MAP problem is repeated until convergence.

[1] Neural PSL accepts any neural network that can be represented in a Keras H5 file without custom code. This allows for the easy incorporation of networks created in the most popular deep learning frameworks, PyTorch and Tensorflow.

More specifically Neural PSL defines a neural objective L_{nn} as:

$$\begin{aligned} L_{nn}(\theta, X_{nn}, X_{map}) &= \frac{1}{n}\sum_{i=1}^{n} l(g(\theta, X_{nn}^{i}), Y_{map}^{i}) \\ &= \frac{1}{m}\sum_{i=1}^{m} l(Y_{nn}^{i}, Y_{map}^{i}) \end{aligned}$$

where θ are the weights of the network, $g(\theta, X_{nn}^{i})$ is the network's prediction for X_{nn}^{i} given θ, and $l(Y_{nn}^{i}, Y_{map}^{i})$ is the network's loss function, e.g., a cross-entropy loss. In this objective, predictions from the MAP problem, Y_{map}, are used as labels.

Neural PSL also redefines the PSL MAP objective, L_{map}, to use predictions from the neural problem, Y_{nn}, as additional observations:

$$\begin{aligned} L_{map}(W, X, Y) &= L_{map}(W, X_{map} \oplus Y_{nn}, Y_{map}) \\ &= \sum_{i=1}^{m} w_i \phi_i(Y_{map}, X_{map} \oplus Y_{nn}) \end{aligned}$$

where $X_{map} \oplus Y_{nn}$ concatenates all the observations from the MAP problem, X_{map}, to predictions from the neural problem, Y_{nn}.

Now, Neural PSL defines a joint loss that minimizes each objective along with the distance between the predictions for each sub problem:

$$\begin{aligned} L_{joint}(W, \theta, X_{map}, X_{nn}, Y_{map}, Y_{nn}) = \\ L_{nn}(\theta, X_{nn}, Y_{map}) \\ + L_{map}(W, X_{map} \oplus Y_{nn}, Y_{map}) \\ + ||Y_{map} - Y_{nn}||_2^p \end{aligned}$$

where $p \in 1, 2$ optionally squares the distance penalty. Figure 2 shows the high-level interactions between the variables in Neural PSL.

4 ERC in PSL

We now describe the rules that compose our PSL model that predicts the emotion associated with each utterance. Each rule encodes structural information about conversational emotion and can be broken into the following categories: label propagation, utterance similarity, neural classification, sum constraint, and priors.

4.1 Label Propagation

In this set of rules, we take advantage of the inherent structure in the dialogue to propagate labels. First, we capture the intuition that conversations tend to have overlying dominant emotion:

$$\begin{aligned} &\textsc{NextUtterance}(\texttt{Utterance1}, \texttt{Utterance2}) \\ &\wedge \textsc{UtteranceEmotion}(\texttt{Utterance1}, \texttt{Emotion}) \\ &\rightarrow \textsc{UtteranceEmotion}(\texttt{Utterance2}, \texttt{Emotion}) \end{aligned}$$

where NextUtterance ties together an utterance, `Utterance1` with the next utterance in the conversation `Utterance2`. This rule propagates emotion from one utterance to the next utterance in a conversation. In this fashion, all utterances in a conversation are chained together and an emotional shift in one influences all others.

The next rule models a speaker maintaining a consistent emotional state between utterances:

$$\begin{aligned} &\textsc{NextSelfUtterance}(\texttt{Utterance1}, \texttt{Utterance2}) \\ &\wedge \textsc{UtteranceEmotion}(\texttt{Utterance1}, \texttt{Emotion}) \\ &\rightarrow \textsc{UtteranceEmotion}(\texttt{Utterance2}, \texttt{Emotion}) \end{aligned}$$

where NextSelfUtterance ties together an utterance, `Utterance1` with the next utterance spoken by the same speaker `Utterance2`. Figure 3 visually demonstrates the structure captured by these two rules.

4.2 Similarity

This rule ensures that similar utterances have similar emotional labels:

$$\begin{aligned} &\textsc{SimilarUtterance}(\texttt{Utterance1}, \texttt{Utterance2}) \\ &\wedge \textsc{UtteranceEmotion}(\texttt{Utterance1}, \texttt{Emotion}) \\ &\rightarrow \textsc{UtteranceEmotion}(\texttt{Utterance2}, \texttt{Emotion}) \end{aligned}$$

where SimilarUtterance is a computed similarity between two utterances. Any similarity between two utterances can be used here. In this model, we use the cosine similarity between the embeddings for each utterance. To create embeddings, we use Google's Universal Sentence Encoder version 4 (Cer et al., 2018). To reduce the size of the graphical model, we only include the highest 10 similarities for each utterance.

4.3 Neural Classification

This rule takes advantage of Neural PSL:

$$\begin{aligned} &\textsc{NeuralClassifier}(\texttt{Utterance}, \texttt{Emotion}) \\ &\rightarrow \textsc{UtteranceEmotion}(\texttt{Utterance}, \texttt{Emotion}) \end{aligned}$$

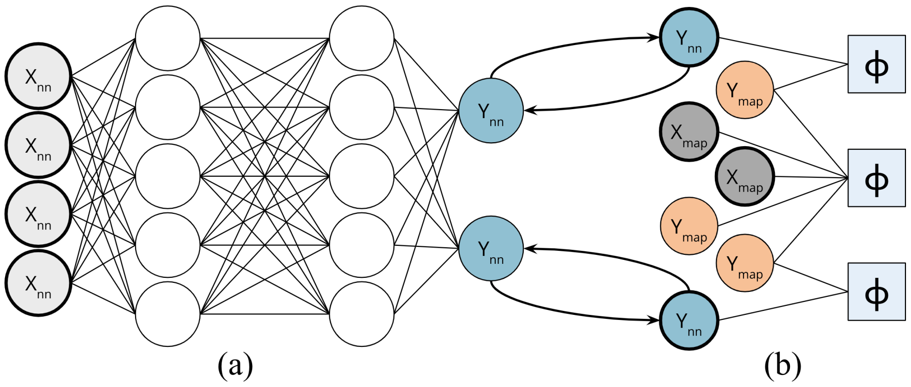
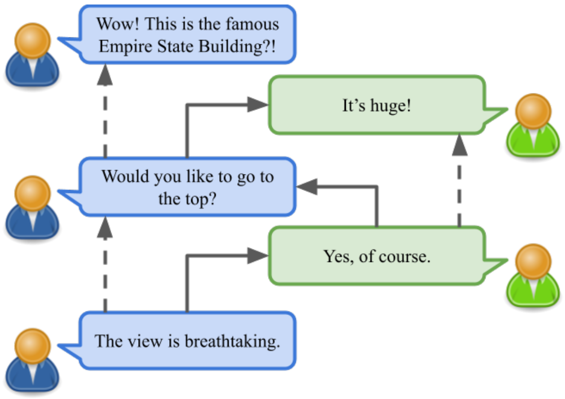


Figure 2: The high-level interactions of variables within Neural PSL. Neural PSL incorporates neural networks (a) with MAP inference over a graphical model (b). Predictions from the neural network Y_{nn} are used a observations in the graphical model. Predictions from the graphical model are then used as labels to further train the neural network. This process repeats until convergence.



Wow! This is the famous Empire State Building?!

It's huge!

Would you like to go to the top?

Yes, of course.

The view is breathtaking.

Figure 3: A sample conversation with the structure of the conversation displayed. Structure chaining together utterances with the next utterance is shown with solid arrows, while structure associated with a single speaker is shown with dashed arrows.

where NEURALCLASSIFIER is a neural network that takes in the embedding for an utterance, and predicts the emotional label for that utterance. As described in Section 3.1, PSL incorporates the network represented by the NEURALCLASSIFIER predicate into the MAP prediction. The network used here is a simple feedforward network with a single hidden layer. The input is the utterance embedding, the hidden layer has a size of 256 with a ReLu activation function, and the output layer has one neuron per emotion and uses a softmax activation function.

4.4 Sum Constraint

Next, we use a PSL hard constraint to ensure that predictions for an utterance sum to 1:

$$\textsc{UtteranceEmotion}(\texttt{Utterance}, +\texttt{Emotion}) = 1.0.$$

This constraint prevents degenerate solutions where all emotions are given full or no confidence (1 and 0 respectively). Instead all emotion predictions for an utterance must compete with one another and sum to exactly 1.

4.5 Priors

Finally, we include two negative priors into our model:

$$\textsc{UtteranceEmotion}(\texttt{Utterance}, \texttt{Emotion}) = 0.0$$
$$\textsc{UtteranceEmotion}(\texttt{Utterance}, '\textit{No Emotion}') = 0.0$$

The first prior acts as a regularizer and defaults predictions without other supporting evidence to a low value. The second prior explicitly encodes an assumption we make that every utterance is associated with an emotion. Specifically, this rules provides an additional penalty for predicting a label of *No Emotion*. This is a strong assumption that does not apply to all of ERC, but Section 5.3 goes into detail on why this assumption works well with the specific dataset we used. Therefore, in this model we assume that every utterance is associated with an emotion, and we treat every instance of an utterance labeled without emotion as a latent variable.

5 Evaluation

In this section, we evaluate the quantitative performance of our model against other recent ERC methods. We also perform a qualitative analysis on our results. Data and code will be made available upon publishing.

5.1 Dataset

We evaluate our method over a popular ERC dataset, DailyDialog (Li et al., 2017). DailyDialog is a multi-turn, dyadic text dataset that was pulled from conversations prepared by humans for the purpose of teaching English as a second language (ESL). Each conversation is designed to be a two-person conversation one may have in their typical daily communication. The conversations average around eight utterances and cover various topics such as the weather, work life, family life, and traveling. Table 1 shows conversation-level statistics on this dataset. Each utterance is labeled with one of seven emotional labels. The labeling for this dataset is heavily biased towards the *No Emotion* label, and to a lesser extent the *Happiness* label. Table 2 shows per-label statistics on this dataset. DailyDialog contains a single train-test split.

Total Conversations	13,118
Mean Utterances Per Conversation	7.9
Mean Tokens Per Conversation	114.7
Mean Tokens Per Utterance	14.6

Table 1: Conversation-level statistics about DailyDialog.

Emotion Label	Count	Percentage
Anger	1022	0.99
Disgust	353	0.34
Fear	74	0.17
Happiness	12885	12.51
Sadness	1150	1.12
Surprise	1823	1.77
No Emotion	85572	83.10

Table 2: Label-level statistics about DailyDialog. *Count* represents the total number of utterances with that emotional label (one label per utterance), while *Percentage* represents the percentage of utterances in the dataset with the associated label.

5.2 Quantitative Model Comparison

To evaluate the performance of our model, we compare against three recent ERC models: CNN+cLSTM (Poria et al., 2017), COSMIC (Ghosal et al., 2020), and CESTa (Wang et al., 2020).

CNN+cLSTM (Poria et al., 2017): Uses a CNN to obtain textual features for an utterance, then applies a context LSTM (cLSTM) over those features to learn contextual information.

COSMIC (Ghosal et al., 2020): Uses different elements of commonsense such as mental states, events, and causal relations to learn interactions between interlocutors participating in a conversation.

CESTa (Wang et al., 2020): Models ERC as a sequence tagging task where a conditional random field is leveraged to learn the emotional consistency in the conversation. Uses LSTM-based encoders that capture self and inter-speaker dependency to generate contextualized utterance representations. Uses a multi-layer transformer encoder to capture long-range global context.

Following the pattern established by the previous methods, our evaluation is performed over the single, canonical split provided with the DailyDialog dataset, and the *No Emotion* label is ignored when computing the Micro F1 score. Table 3 shows the results comparing our method with the previously discussed methods. Here we can clearly see the power of incorporating structure with neural components. Our PSL model performs nearly 20 percentage points better than the next leading method (CESTa).

To further verify our results, we evaluated our method over ten randomly generated splits of DailyDialog. To create these splits, the dataset was shuffled and 10% of conversations were assigned to the test set while the remaining 90% of conversations were assigned to the train set. For these splits, we also evaluated CNN+cLSTM to compare against our method[2]. Table 4 shows that when averaged over ten splits, PSL and CNN+cLSTM both achieve similar performance to the single canonical split. Our PSL method diverges by only 0.33 standard deviations while CNN+cLSTM diverges by only 1.24 standard deviations.

[2] CNN+cLSTM was chosen for this comparison because of its relatively quick runtime and its ease-of-use when running on a new dataset.

Model	Micro F1
CNN+cLSTM	0.518
COSMIC	0.585
CESTa	0.631
PSL	0.813

Table 3: Comparison of the Micro F1 of multiple methods across the canonical DailyDialog split. When Micro F1 is computed, the *No Emotion* label is removed.

Model	Micro F1
CNN+cLSTM	0.549 ± 0.025
PSL	0.809 ± 0.012

Table 4: Comparison of the Micro F1 of multiple methods across ten random DailyDialog splits. When Micro F1 is computed, the *No Emotion* label is removed. Standard deviation is reported along with the mean Micro F1.

5.3 Qualitative Analysis

In order to better understand the results in a complex setting like ERC, we also look at specific examples in further detail.

5.3.1 The *No Emotion* Class

As seen in Table 2, DailyDialog is heavily biased towards the *No Emotion* class. At first, it may seem that this class represents utterances that have no clear emotional context. Table 5 shows examples of uses of the *No Emotion* class where there is no clear emotional context.

Label	Utterance
No Emotion	I don't care.
No Emotion	Do you want black or white coffee?
No Emotion	She's my grandma.
No Emotion	When's your birthday?
No Emotion	I'm a doctor.
No Emotion	I certainly have.
No Emotion	About two hours ago.

Table 5: Utterances labeled as *No Emotion* and showing no clear emotional context.

However, the *No Emotion* label is also clearly used in cases where the emotional context is apparent. Table 6 shows examples where an utterance is labeled as *No Emotion*, but it is clear that a label associated with an emotion is more appropriate.

Finally, there are cases where the *No Emotion* label is used for a specific utterance, but the context of the conversation provides information on what the labeling should be. Table 7 shows additional dialog context for the last two utterances in Table 5. Both of these utterances (in bold) were labeled as *No Emotion*, and without any other context that label would make sense. However with the full context of the dialog (the speaker being bound, gagged, and robbed), a more appropriate label should be applied (e.g. *Anger*, *Fear*, or *Sadness*).

6 Conclusions and Future Work

In this paper, we proposed a neural-symbolic method for the task of ERC that combines a simple neural model with the relational inference of PSL. Our initial experiments show that even a simple neural model combined with general-purpose logical rules can outperform complex and specific state-of-the-art neural models. Furthermore, our qualitative analysis shows our model performing well even in situations where the dataset's labels are open to question.

In our future work, we plan to extend both the neural and logical components of our model. On the neural side, we can utilize more complex neural models. On the logical side, we can incorporate additional structure into our models by computing more sophisticated utterance similarity and integrating both conversation-level and user-level similarities. We also want to prove the generality of our approach by testing it on additional ERC datasets. Finally, we plan on addressing the issues discussed in Section 5.3 by relabeling the DailyDialog dataset with fine-grained emotion.

References

Elisabeth André, Matthias Rehm, Wolfgang Minker, and Dirk Bühler. 2004. Endowing spoken language dialogue systems with emotional intelligence. In *Affective Dialogue Systems (ADS)*.

Stephen Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. 2017. Hinge-loss markov random fields and probabilistic soft logic. *Journal of Machine Learning Research (JMLR)*, 18(1):1–67.

Stephen H. Bach, Bert Huang, Ben London, and Lise Getoor. 2013. Hinge-loss markov random fields: Convex inference for structured prediction. In *Uncertainty in Artificial Intelligence (UAI)*.

Tarek R. Besold, Artur S. d'Avila Garcez, Sebastian Bader, Howard Bowman, Pedro M. Domingos, Pas-

Label	Prediction	Utterance
No Emotion	*Anger*	Damp it! How are you killing me with a single shot? It's not fair! I don't want to play anymore!
No Emotion	*Disgust*	What a creep! Phony good luck e-mails are one thing, but sexual harassment is crossing the line.
No Emotion	*Fear*	Oh , doctor. Do I have to? I am afraid of needles!
No Emotion	*Sadness*	I don't know, but I feel terrible.
No Emotion	*Happiness*	And now we have a two-year-old boy. We're very happy that he's healthy and smart.
No Emotion	*Surprise*	Ah! You're bleeding all over! What happened?

Table 6: Utterances labeled as *No Emotion*, but showing clear emotional context. Predictions made the the PSL model are included.

Speaker	Utterance
Speaker 1	Good evening, sir. I understand that you have been robbed.
Speaker 2	**I certainly have.**
Speaker 1	When did this happen?
Speaker 2	**About two hours ago.**
Speaker 1	Why didn't you report it before?
Speaker 2	I couldn't. I was bound and gagged.

Table 7: A conversation that demonstrates the overuse of the *No Emotion* label. The bold utterances were labeled as *No Emotion*, but with the context of the full conversation could have been more accurately labeled.

cal Hitzler, Kai-Uwe Kühnberger, Luís C. Lamb, Daniel Lowd, Priscila Machado Vieira Lima, Leo de Penning, Gadi Pinkas, Hoifung Poon, and Gerson Zaverucha. 2017. Neural-symbolic learning and reasoning: A survey and interpretation. *arXiv*.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. 2010. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122.

Daniel Cer, Yinfei Yang, Sheng yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. 2018. Universal sentence encoder for english. In *Empirical Methods in Natural Language Processing (EMNLP)*.

Junyoung Chung, Çaglar Gülçehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *Neural Information Processing Systems (NeurIPS)*.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *Neural Information Processing Systems (NeurIPS)*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Association for Computational Linguistics (ACL)*.

Asif Ekbal. 2020. Towards building an affect-aware dialogue agent with deep neural networks. *CSI Transactions on ICT*, 8(2):249–255.

Varun Embar, Bunyamin Sisman, Hao Wei, Xin Luna Dong, Christos Faloutsos, and Lise Getoor. 2020. Contrastive entity linkage: Mining variational attributes from large catalogs for entity linkage. In *Automated Knowledge Base Construction (AKBC)*, Virtual.

Artur S. d'Avila Garcez, Marco Gori, Luís C. Lamb, Luciano Serafini, Michael Spranger, and Son N. Tran. 2019. Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. *Journal of Applied Logics*, 6(4):611–632.

Asma Ghandeharioun, Daniel McDuff, Mary Czerwinski, and Kael Rowan. 2019. EMMA: an emotion-aware wellbeing chatbot. In *Affective Computing and Intelligent Interaction (ACII)*.

Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. COSMIC: COmmonSense knowledge for eMotion identification in conversations. In *Empirical Methods in Natural Language Processing (EMNLP)*.

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 *Empirical Methods in Natural Language Processing (EMNLP)*.

Devamanyu Hazarika, Soujanya Poria, Rada Mihalcea, Erik Cambria, and Roger Zimmermann. 2018a. ICON: interactive conversational memory network for multimodal emotion detection. In *Empirical Methods in Natural Language Processing (EMNLP)*.

Devamanyu Hazarika, Soujanya Poria, Amir Zadeh, Erik Cambria, Louis-Philippe Morency, and Roger Zimmermann. 2018b. Conversational memory network for emotion recognition in dyadic dialogue videos. In *Association for Computational Linguistics (ACL)*.

Eduard Hovy. 1987. Generating natural language under pragmatic constraints. *Journal of Pragmatics*, 11(6):689–719.

Pigi Kouki, Jay Pujara, Christopher Marcum, Laura Koehly, and Lise Getoor. 2019. Collective entity resolution in multi-relational familial networks. *International Journal on Knowledge and Information Systems (KAIS)*, 61(3):1547–1581.

Jingye Li, Meishan Zhang, Donghong Ji, and Yijiang Liu. 2020. Multi-task learning with auxiliary speaker identification for conversational emotion recognition. *arXiv*.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. Dailydialog: A manually labelled multi-turn dialogue dataset. In *International Joint Conference on Natural Language Processing (IJCNLP)*.

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 *Association for the Advancement of Artificial Intelligence (AAAI)*.

Sherief Mowafey and Steve Gardner. 2012. A novel adaptive approach for home care ambient intelligent environments with an emotion-aware system. In *UKACC International Conference on Control (CONTROL)*. IEEE.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval (FTIR)*, 2(1–-2):1–135.

Thomas S Polzin and Alexander Waibel. 2000. Emotion-sensitive human-computer interfaces. In *ISCA Tutorial and Research Workshop on Speech and Emotion*.

Soujanya Poria, Erik Cambria, Devamanyu Hazarika, Navonil Majumder, Amir Zadeh, and Louis-Philippe Morency. 2017. Context-dependent sentiment analysis in user-generated videos. In *Association for Computational Linguistics (ACL)*.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019a. MELD: A multimodal multi-party dataset for emotion recognition in conversations. In *Association for Computational Linguistics (ACL)*.

Soujanya Poria, Navonil Majumder, Rada Mihalcea, and Eduard H. Hovy. 2019b. Emotion recognition in conversation: Research challenges, datasets, and recent advances. *IEEE Access*, 7:100943–100953.

Libo Qin, Wanxiang Che, Yangming Li, Minheng Ni, and Ting Liu. 2020. Dcr-net: A deep co-interactive relation network for joint dialog act recognition and sentiment classification. In *Association for the Advancement of Artificial Intelligence (AAAI)*.

Luc De Raedt, Sebastijan Dumančić, Robin Manhaeve, and Giuseppe Marra. 2020. From statistical relational to neuro-symbolic artificial intelligence. In *International Joint Conference on Artificial Intelligence (IJCAI)*.

Julian J. Schlöder and Raquel Fernández. 2015. Clarifying intentions in dialogue: A corpus study. In *Computational Semantics (IWCS)*.

Marcin Skowron. 2010. Affect listeners: Acquisition of affective states by means of conversational systems. In *Development of Multimodal Interfaces: Active Listening and Synchrony*.

Marcin Skowron, Stefan Rank, Mathias Theunis, and Julian Sienkiewicz. 2011. The good, the bad and the neutral: Affective profile in dialog system-user communication. In *Affective Computing and Intelligent Interaction (ACII)*. Springer.

Dhanya Sridhar and Lise Getoor. 2019. Estimating causal effects of tone in online debates. In *International Joint Conference on Artificial Intelligence (IJCAI)*.

Charles Sutton, Andrew McCallum, and Khashayar Rohanimanesh. 2007. Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data. *Journal of Machine Learning Research (JMLR)*, 8:693–723.

Michael J Tanana, Christina S Soma, Patty B Kuo, Nicolas M Bertagnolli, Aaron Dembe, Brian T Pace, Vivek Srikumar, David C Atkins, and Zac E Imel. 2021. How do you feel? using natural language processing to automatically rate emotion in psychotherapy. *Behavior Research Methods*, pages 1–14.

Sabina Tomkins, Lise Getoor, Yunfei Chen, and Yi Zhang. 2017. Detecting cyber-bullying from sparse data and inconsistent labels. In *Learning from Limited Labeled Data Workshop (LLD)*.

Yan Wang, Jiayu Zhang, Jun Ma, Shaojun Wang, and Jing Xiao. 2020. Contextualized emotion recognition in conversation as sequence tagging. In *Special Interest Group on Discourse and Dialogue (SIGdial)*.

Sung-Lin Yeh, Yun-Shao Lin, and Chi-Chun Lee. 2019. An interaction-aware attention network for speech emotion recognition in spoken dialogs. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Sayyed M. Zahiri and Jinho D. Choi. 2018. Emotion detection on TV show transcripts with sequence-based convolutional neural networks. In *Association for the Advancement of Artificial Intelligence (AAAI)*.

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 *International Joint Conference on Artificial Intelligence (IJCAI)*.

Peixiang Zhong, Di Wang, and Chunyan Miao. 2019. Knowledge-enriched transformer for emotion detection in textual conversations. In *Empirical Methods in Natural Language Processing (EMNLP)*.