

SEC : contexte émotionnel phrasique intégré pour la reconnaissance émotionnelle efficiente dans la conversation

Barbara Gendron-Audebert^{1,2} et Gaël Guibon¹
{prénom.nom}@loria.fr

(1) LORIA, Université de Lorraine, CNRS (2) Université du Luxembourg

JEP-TALN 2024, 09 juillet 2024, Toulouse

Contexte

- *Comme quoi on a de plus en plus de discours à caractère émotionnel dans les contenus générés par les utilisateurs.*
- *Il est nécessaire de bien les comprendre et c'est souvent subtile (co-références, ironie, amplification, biais, ...)*
- *Il existe déjà des modèles pour ça mais ça reste une tâche difficile à accomplir, qu'on pourrait essayer d'améliorer avec les progrès du deep learning*

Ajouter des choses à partir de l'intro du papier. Quelle illustration sympa on pourrait mettre ici ? Il faut quelque chose d'accrocheur à ce stade...

Objectifs

Cadre de l'étude :

- Détection et identification des émotions dans le contenu généré par les utilisateurs
- Dialogues sous forme de conversations dyadiques
- Reconnaissance d'Émotions en Conversation (ERC)

Objectifs

Cadre de l'étude :

- Détection et identification des émotions dans le contenu généré par les utilisateurs
- Dialogues sous forme de conversations dyadiques
- Reconnaissance d'Émotions en Conversation (ERC)

Questions de recherche :

- Comment utiliser l'information provenant du contexte conversationnel pour guider la détection d'émotions en conversation ?
- Est-ce que la prise en compte du contexte conversationnel permet d'améliorer la détection d'émotions en conversation dans le cas dyadique ?

Apport du *deep learning*

- La profondeur du réseau neuronal permet de traiter certains aspects plus subtils du discours (parallèle avec la vision si le temps ?)
- De nombreuses structures séquentielles à disposition
- Les modèles neuronaux obtiennent des résultats état-de-l'art en ERC^{1, 2}

Ici un lien avec notre approche montrant la richesse de notre exploration des réseaux neuronaux. En effet, nous ne nous sommes pas limités aux derniers Transformers mais avons aussi testé des MLP, LSTM, Encoder-only et Decoder-only. Je pense que ça mérite d'être souligné.

1. Soujanya PORIA et al. *Emotion Recognition in Conversation : Research Challenges, Datasets, and Recent Advances*. 2019.

2. Patrícia PEREIRA, Helena MONIZ et Joao Paulo CARVALHO. *Deep Emotion Recognition in Textual Conversations : A Survey*. 2022.

Travaux connexes

Présenter ici rapidement quelques approches utilisant du deep learning (en gros les usines à gaz de notre tableau SOTA, en précisant ce qu'ils rajoutent), avec peut-être une transition vers le metric learning grâce à un exemple qui présente du meta-learning / apprentissage contrastif.

Apport du *metric learning*

Motiver l'utilisation de cette forme de méta-apprentissage, qu'on peut définir par le formalisme "apprendre à apprendre" mais aussi par cette idée d'extraction de meta-representation. C'est ça qu'on va utiliser par la suite, et ça permet plein de choses pertinentes dans le cadre des émotions, notamment :

- *Un cadre de classification plus souple : introduction de nouveaux labels inconnus sans avoir besoin de changer le modèle ni l'entraînement*
- *L'extraction de relations entre les labels de manière assez naturelle puisque c'est comme ça que le modèle a été entraîné*
- *Une adaptation intrinsèque à l'apprentissage avec peu d'essais (few-shot learning), ce qui permet d'apprendre des labels émotionnels peu représentés/dotés*

Travaux connexes

Donner des exemples en metric learning, notamment protosec !

Réseaux siamois

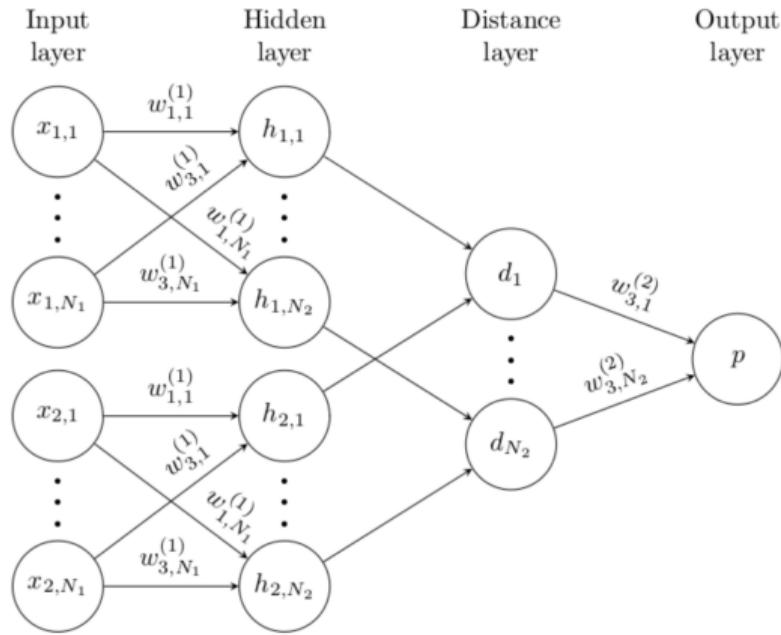


FIGURE – Architecture des réseaux siamois³

3. Gregory KOCH, Richard ZEMEL et Ruslan SALAKHUTDINOV. *Siamese Neural Networks for One-shot Image Recognition*. en. 2015.

Réseaux siamois

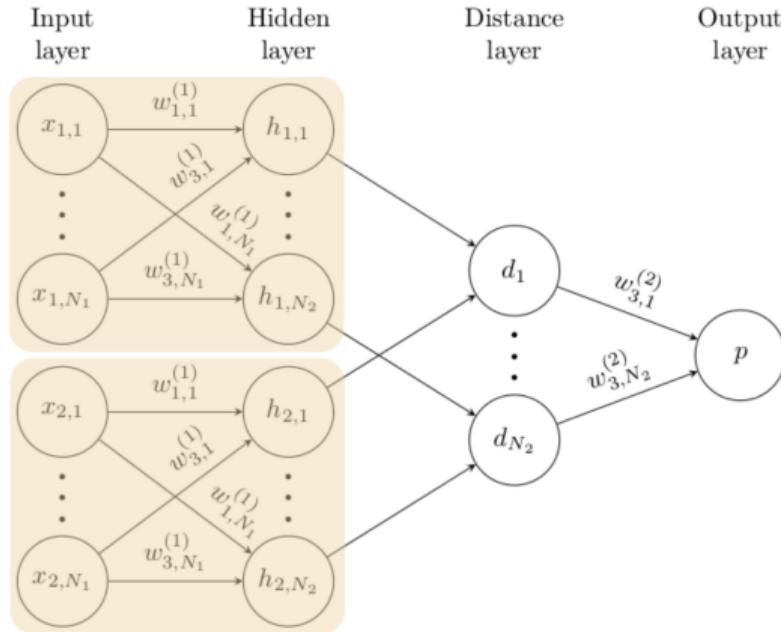


FIGURE – Architecture des réseaux siamois³

3. Gregory KOCH, Richard ZEMEL et Ruslan SALAKHUTDINOV. *Siamese Neural Networks for One-shot Image Recognition.* en. 2015.

Réseaux siamois

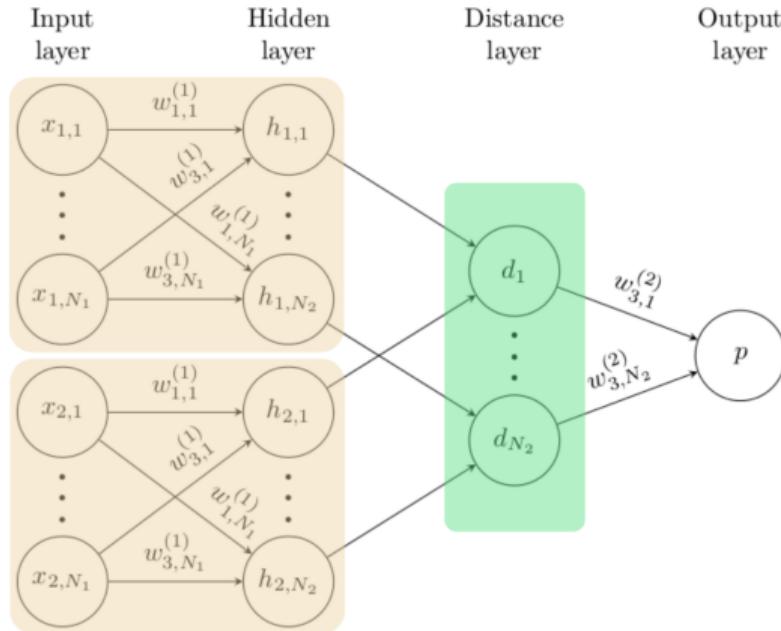


FIGURE – Architecture des réseaux siamois³

3. Gregory KOCH, Richard ZEMEL et Ruslan SALAKHUTDINOV. *Siamese Neural Networks for One-shot Image Recognition*. en. 2015.

Triplet loss : fonction de coût par triplets

Un triplet de propos :

- Ancre (A)
- Positif (P)
- Négatif (N)

A et P sont de la même classe, N est d'une autre.

Objectif de la *triplet loss* :

- Minimiser $d(A, P)$
- Maximiser $d(A, N)$

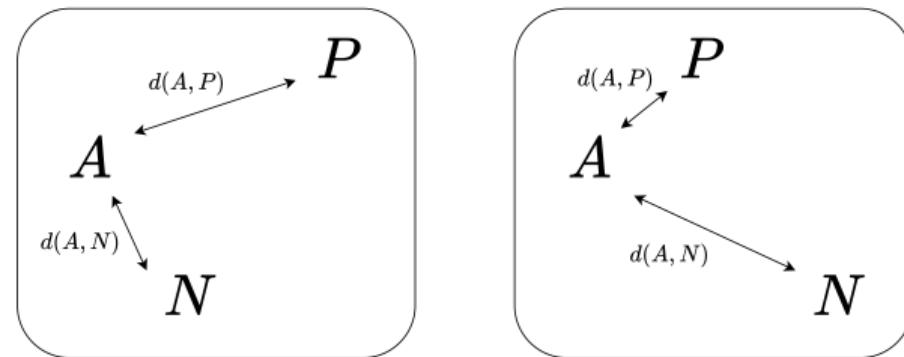
Triplet loss : fonction de coût par triplets

Un triplet de propos :

- Ancre (A)
- Positif (P)
- Négatif (N)

A et P sont de la même classe, N est d'une autre.

$$\mathcal{L}(a, p, n) = \max \{d(a, p) - d(a, n) + \text{marge}, 0\}$$



Objectif de la *triplet loss* :

- Minimiser $d(A, P)$
- Maximiser $d(A, N)$

$$\begin{aligned} d(A, P) - d(A, N) &> 0 \\ \mathcal{L}(A, P, N) &> 0 \end{aligned}$$

$$\begin{aligned} d(A, P) - d(A, N) &< 0 \\ \mathcal{L}(A, P, N) &= 0 \end{aligned}$$

FIGURE – Illustration du principe de la *triplet loss*

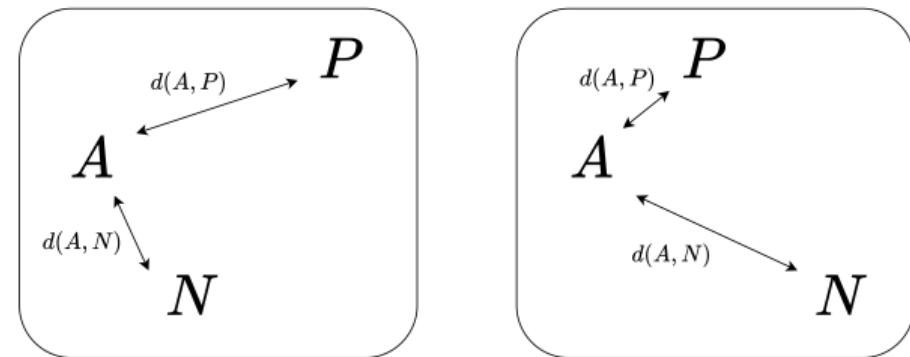
Triplet loss : fonction de coût par triplets

Un triplet de propos :

- Ancre (A)
- Positif (P)
- Négatif (N)

A et P sont de la même classe, N est d'une autre.

$$\mathcal{L}(a, p, n) = \max \{d(a, p) - d(a, n) + \text{marge}, 0\}$$



Objectif de la *triplet loss* :

- Minimiser $d(A, P)$
- Maximiser $d(A, N)$

$$\begin{aligned} d(A, P) - d(A, N) &> 0 \\ \mathcal{L}(A, P, N) &> 0 \end{aligned}$$

$$\begin{aligned} d(A, P) - d(A, N) &< 0 \\ \mathcal{L}(A, P, N) &= 0 \end{aligned}$$

FIGURE – Illustration du principe de la *triplet loss*

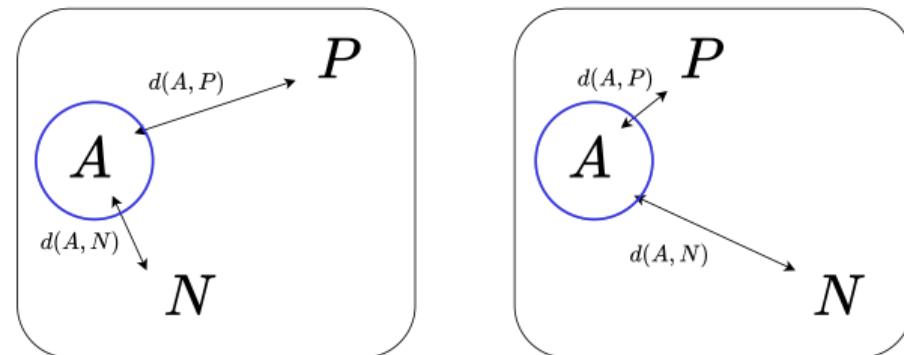
Triplet loss : fonction de coût par triplets

Un triplet de propos :

- Ancre (A)
- Positif (P)
- Négatif (N)

A et P sont de la même classe, N est d'une autre.

$$\mathcal{L}(a, p, n) = \max \{d(a, p) - d(a, n) + \text{marge}, 0\}$$



Objectif de la *triplet loss* :

- Minimiser $d(A, P)$
- Maximiser $d(A, N)$

$$\begin{aligned} d(A, P) - d(A, N) &> 0 \\ \mathcal{L}(A, P, N) &> 0 \end{aligned}$$

$$\begin{aligned} d(A, P) - d(A, N) &< 0 \\ \mathcal{L}(A, P, N) &= 0 \end{aligned}$$

FIGURE – Illustration du principe de la *triplet loss*

Protocole expérimental

Jeu de données DailyDialog⁴

- 13 118 dialogues dyadiques en anglais sur des sujets de la vie quotidienne
- Annotation au niveau du tour de parole : happiness, anger, disgust, fear, surprise, sadness et no emotion

4. Yanran LI et al. *DailyDialog : A Manually Labelled Multi-turn Dialogue Dataset*. Taipei, Taiwan, nov. 2017.

Protocole expérimental

Jeu de données DailyDialog⁴

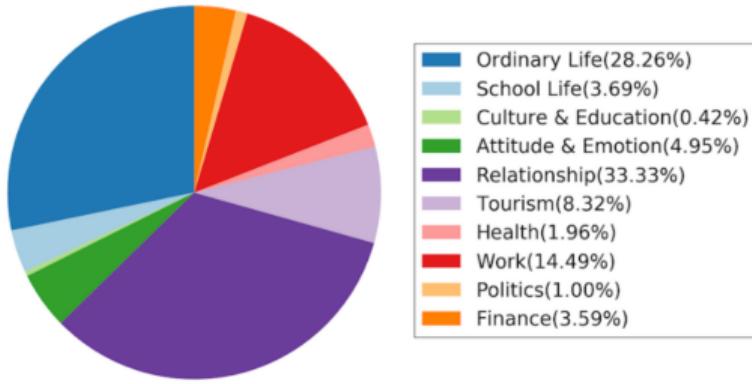
- 13 118 dialogues dyadiques en anglais sur des sujets de la vie quotidienne
- Annotation au niveau du tour de parole : happiness, anger, disgust, fear, surprise, sadness et no emotion

4. Yanran LI et al. *DailyDialog : A Manually Labelled Multi-turn Dialogue Dataset*. Taipei, Taiwan, nov. 2017.

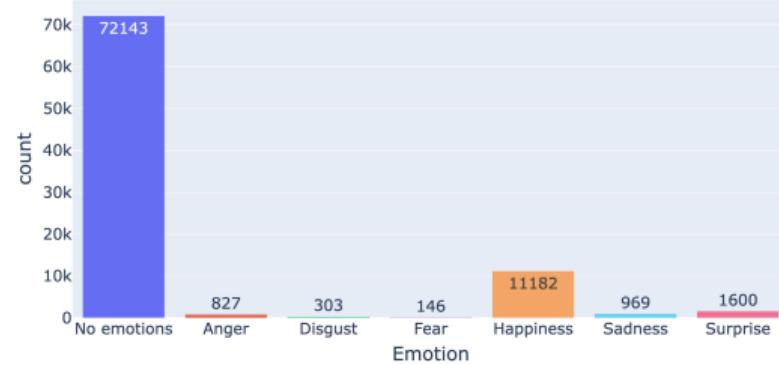
Protocole expérimental

Jeu de données DailyDialog⁴

- 13 118 dialogues dyadiques en anglais sur des sujets de la vie quotidienne
- Annotation au niveau du tour de parole : happiness, anger, disgust, fear, surprise, sadness et no emotion



(a) Répartition des sujets des dialogues



(b) Distribution des émotions dans les données d'entraînement

4. Yanran LI et al. *DailyDialog : A Manually Labelled Multi-turn Dialogue Dataset*. Taipei, Taiwan, nov. 2017.

Procédure d'entraînement

Prédictions d'émotions avec contexte (propos contextuels)

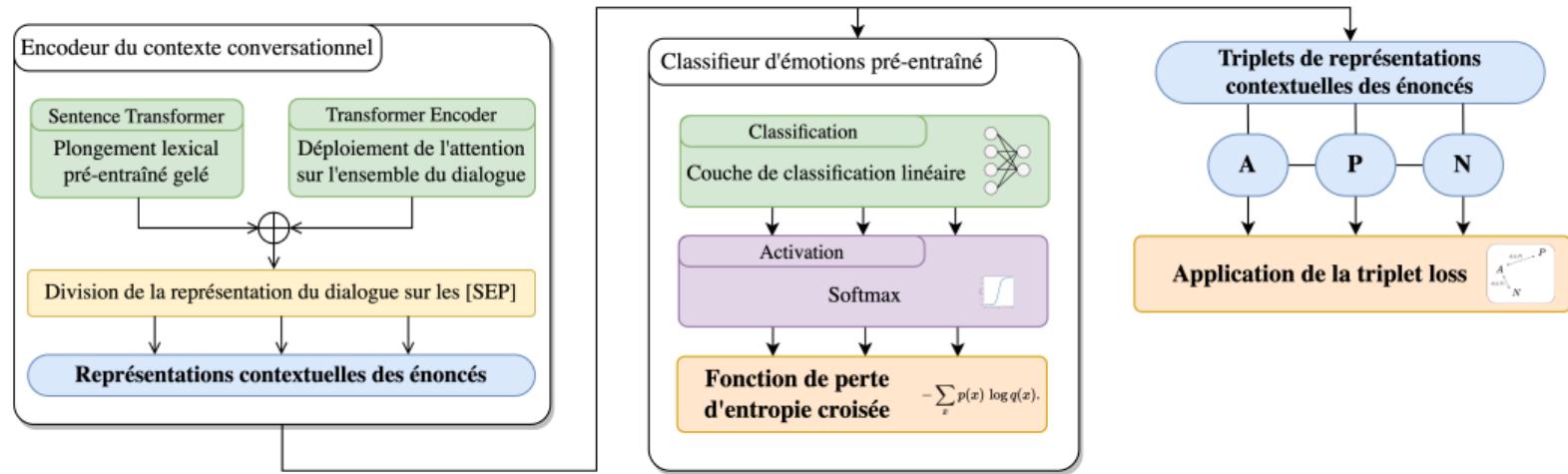


FIGURE – Prédiction d'émotions sur des représentations de propos contextuelles.

Évaluations quantitative et qualitative

Métriques d'évaluation

- *MicroF1 : choix historiquement privilégié par la littérature*
- *MacroF1 paraît plus pertinent car favorise des modèles plus polyvalents en reconnaissance des émotions, donc moins biaisé et plus aptes à distinguer des émotions nouvelles*
- *MCC (Matthews Correlation Coefficient) : métrique plus contraignante mais essentielle pour assurer la pertinence de l'entraînement.*

Le MCC est défini comme suit⁵ :

$$\text{MCC} = \frac{TP/N - S \times P}{\sqrt{PS(1-S)(1-P)}} \quad (1)$$

Où TP est le nombre de vrais positifs, N la taille des données, P la précision et S le rappel (pour *sensitivity*).

5. Brian W. MATTHEWS. « Comparison of the predicted and observed secondary structure of T4 phage lysozyme. ». In : *Biochimica et biophysica acta* (1975).

Résultats quantitatifs

Nom du modèle	macroF1*	microF1*	MCC
Modèles état-de-l'art en ERC			
CNN+cLSTM (Poria <i>et al.</i> , 2017)	–	50.24	–
KET (Zhong <i>et al.</i> , 2019)	–	53.37	–
COSMIC (Ghosal <i>et al.</i> , 2020)	51.05	58.48	–
RoBERTa (Ghosal <i>et al.</i> , 2020)	48.20	55.16	–
Rpe-RGAT (Ishiwatari <i>et al.</i> , 2020)	–	54.31	–
Glove-DRNN (Ghosal <i>et al.</i> , 2021)	41.80	55.95	–
roBERTa-DRNN (Ghosal <i>et al.</i> , 2021)	49.65	57.32	–
CNN (Ghosal <i>et al.</i> , 2021)	36.87	50.32	–
DAG-ERC (Shen <i>et al.</i> , 2021)	–	59.33	–
TODKAT (Zhu <i>et al.</i> , 2021)	<u>52.56</u>	58.47	–
SKAIG (Li <i>et al.</i> , 2021)	51.95	59.75	–
Sentic GAT (Tu <i>et al.</i> , 2022)	–	54.45	–
CauAIN (Zhao <i>et al.</i> , 2022)	–	58.21	–
DialogueRole (Ong <i>et al.</i> , 2022)	–	60.95	–
S+PAGE (Liang <i>et al.</i> , 2022)	–	64.07	–
DualGAT (Zhang <i>et al.</i> , 2023)	–	<u>61.84</u>	–
CD-ERC (Pereira <i>et al.</i> , 2023)	51.23	–	–
Notre approche			
SentEmoContext	57.71	57.75	0.49

FIGURE – Résultats en ERC sur DailyDialog

Évaluation qualitative

Discuter la question de la subjectivité de l'annotation

Comparaison avec les LLMs

Nom du modèle	macroF1*	microF1*	MCC
LLMs			
Llama2-7b (Touvron et al., 2023)	09.70	24.92	0.08
Llama2-13b (Touvron et al., 2023)	22.26	43.37	0.15
Falcon-7b (Penedo et al., 2023)	07.54	42.75	0.01
Notre approche			
SentEmoContext	57.71	57.75	0.49

[FIGURE – Résultats avec les LLMs et comparaison avec SentEmoContext](#)

	Transformers		LLMs			Notre approche
Modèle	MiniLM	MPNet	Llama2-7b	Llama2-13b	Falcon-7b	SentEmoContext
Tokens	1bn+	1bn+	2T	2T	1.5T	4M
Taille	80 MB	420 MB	13 GB	25 GB	15 GB	604,8 MB
Paramètres	22M	110M	7B	13B	7B	157M

[FIGURE – Aperçu de la taille des modèles et comparaison avec SentEmoContext](#)

Conclusion

Perspectives

Aller vers la détection plus subtile, d'émotions plus subtiles et pourquoi pas lien avec ironie ?

Merci pour votre attention !



FIGURE – Lien vers l'article



FIGURE – Code (dépôt GitHub)

Références I

- KOCH, Gregory, Richard ZEMEL et Ruslan SALAKHUTDINOV. *Siamese Neural Networks for One-shot Image Recognition.* en. 2015.
- LI, Yanran et al. *DailyDialog : A Manually Labelled Multi-turn Dialogue Dataset.* Taipei, Taiwan, nov. 2017.
- MATTHEWS, Brian W. « Comparison of the predicted and observed secondary structure of T4 phage lysozyme. ». In : *Biochimica et biophysica acta* (1975).
- PEREIRA, Patrícia, Helena MONIZ et Joao Paulo CARVALHO. *Deep Emotion Recognition in Textual Conversations : A Survey.* 2022.
- PORIA, Soujanya et al. *Emotion Recognition in Conversation : Research Challenges, Datasets, and Recent Advances.* 2019.

About Classification Metrics

Given :

- P the quantity of positive predictions
- N the quantity of negative predictions
- TP , TN , FP and FN the True Positives, True Negatives, False Positives and False Negatives

$$\text{Accuracy} = \frac{TP + TN}{P + N} \quad \text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad F_1 = \frac{2TP}{2TP + FP + FN} \quad (3)$$

About Matthews Correlation Coefficient (MCC) (Cramér, 1946)

Given :

$$N = TN + TP + FN + FP , \quad S = \frac{TP + FN}{N} \quad \text{and} \quad P = \frac{TP + FP}{N} \quad (4)$$

MCC has been defined in (Matthews, 1975) as :

$$\text{MCC} = \frac{TP/N - S \times P}{\sqrt{PS(1-S)(1-P)}} \quad (5)$$

Distribution of emotions within the dialog



FIGURE – Cumulative number of expressed emotions *w.r.t.* the utterance index

Some unexplainable predictions

===== DIALOG #82 =====

- Excuse me . Check please .
- OK , how was everything ?
- Very nice . Thank you .
- Would you like this to-go ?
- Yes , can you put it in a plastic bag ?
- Sure , no problem . Here you are . That'll be 25 dollars .
- Do you take credit cards ?
- Yes , we accept Visa and MasterCard .
- OK , here you are .
- Thanks . I'll be right back .
- OK .
- Here's your receipt .
- Thank you .
- You're welcome . Please come again .

Utterance: OK .

Emotion: no emo. Prediction: happiness X

(a)

===== DIALOG #286 =====

- I need to get my high speed internet installed .
- You'll need to make an appointment .
- Could I do that right now , please ?
- What day would you like us to do the installation ?
- Is Friday good ?
- We're only available at 3
- You can't come any earlier than that ?
- I'm sorry . That's the only available time .
- Are you available this Saturday ?
- Yes . Anytime on Saturday will be fine .
- How does 11
- We can do it . See you then .

Utterance: How does 11

Emotion: no emo. Prediction: sadness X

(b)

Figure 4.11: Two examples of unexplainable predictions.

Preprocessing Pipeline

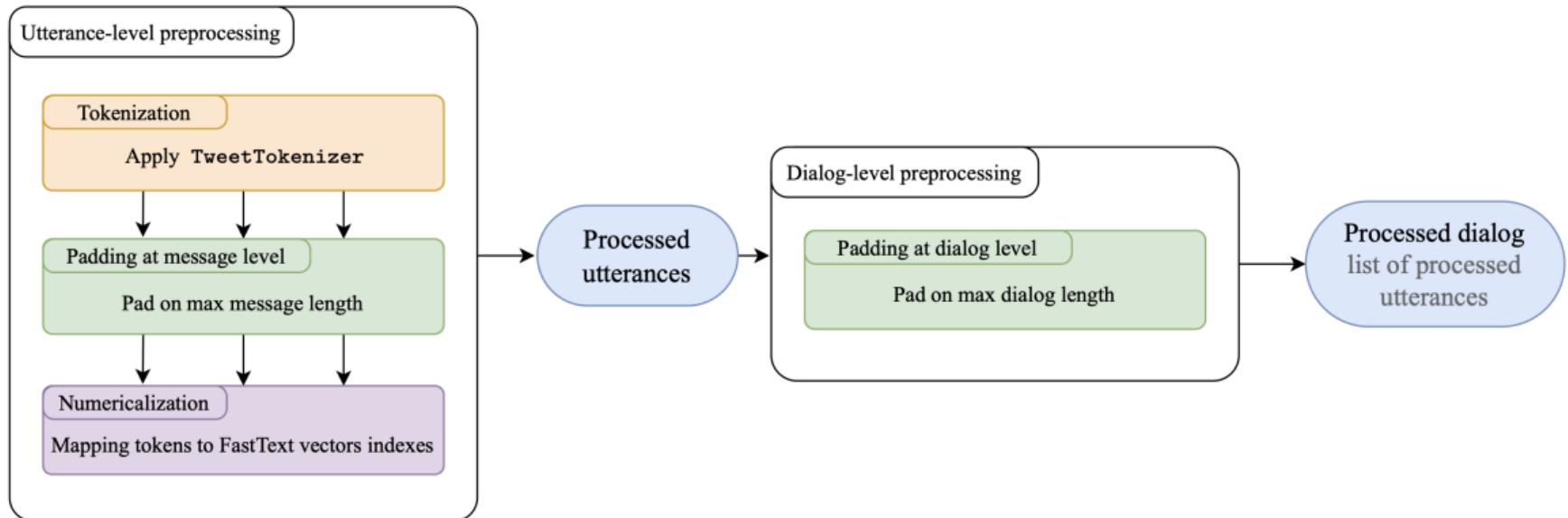


FIGURE – Preprocessing steps for isolated utterance representations

Preprocessing Pipeline

Original dialog

— Say, Jim, how about going for a few beers after dinner ?
— You know that is tempting but is really not good for our fitness.

Formatted dialog with [SEP] tokens

Say , Jim , how about going for a few beers after dinner ? [SEP] You know that is tempting
but is really not good for our fitness .

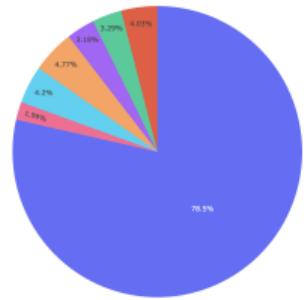
Word-piece tokenization

```
['[CLS]', 'say', ',', 'jim', ',', 'how', 'about', 'going', 'for', 'a', 'few',
'beers', 'after', 'dinner', '?', '[SEP]', 'you', 'know', 'that', 'is', 'tempting',
'but', 'is', 'really', 'not', 'good', 'for', 'our', 'fitness', '.', '[PAD]',
..., '[PAD]']
```

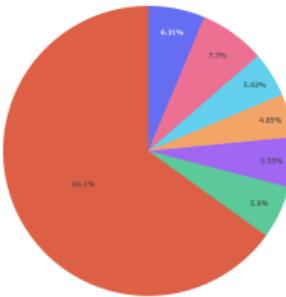
Numericalization

```
[101, 2360, 1010, 3958, 1010, 2129, 2055, 2183, 2005, 1037, 2261, 18007,
2044, 4596, 1029, 102, 2017, 2113, 2008, 2003, 23421, 2021, 2003, 2428, 2025,
2204, 2005, 2256, 10516, 1012, 0, ..., 0]
```

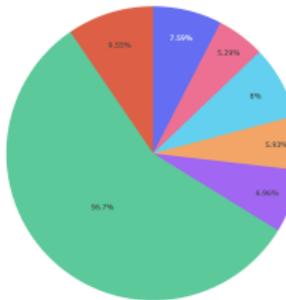
Prediction Distributions - Isolated Utterances



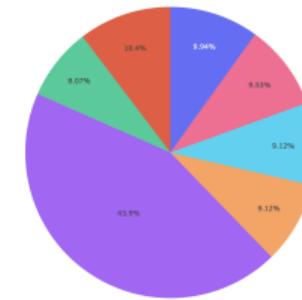
No emotion



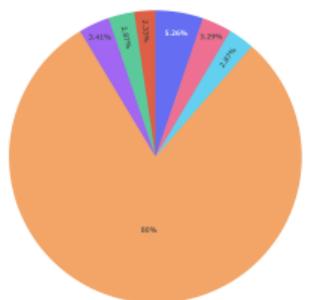
Anger



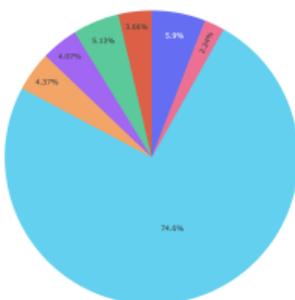
Disgust



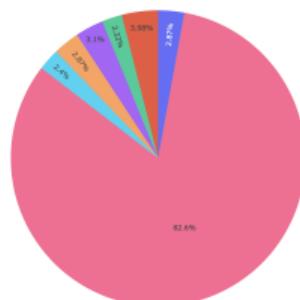
Fear



Happiness



Sadness



Surprise

Emotion

- No emotions
- Anger
- Disgust
- Fear
- Happiness
- Sadness
- Surprise

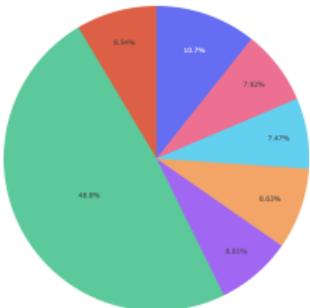
Prediction Distributions - Contextual Utterances



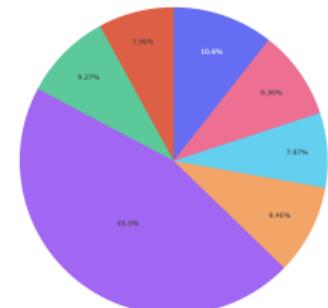
No emotion



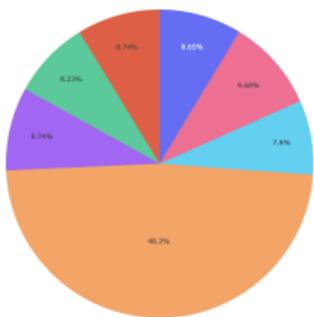
Anger



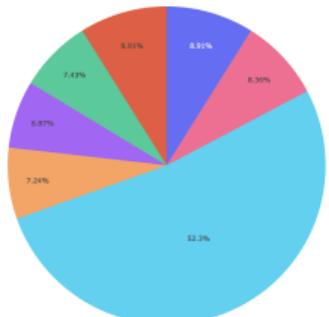
Disgust



Fear



Happiness



Sadness



Surprise

Emotion

- No emotions
- Anger
- Disgust
- Fear
- Happiness
- Sadness
- Surprise

Performances - Isolated Utterances

Layers	Accuracy↑	Loss↓	Precision↑	Recall↑	wF1 score↑
LSTM-based models					
3	68.0	0.706	0.682	0.680	0.680
4	67.9	0.734	0.678	0.679	0.677
5	70.0	0.711	0.702	0.700	0.700
6	65.4	0.756	0.658	0.654	0.655
7	66.2	0.793	0.664	0.662	0.661
MLP-based models					
3	68.4	0.747	0.685	0.684	0.684
4	62.9	0.834	0.629	0.629	0.628
5	64.7	0.801	0.650	0.647	0.647

Table 4.1: Main results using Siamese Networks on static utterances representations.
Best values are in **bold** and arrows indicates if greater or lower is better.