Software Impacts 12 (2022) 100237

Contents lists available at ScienceDirect

Software Impacts

journal homepage: www.journals.elsevier.com/software-impacts

Original software publication

Few-shot emotion recognition in conversation with sequential prototypical networks

Gaël Guibon [a,b,*], Matthieu Labeau [a], Luce Lefeuvre [b], Chloé Clavel [a]

[a] *LTCI, Télécom-Paris, Institut Polytechnique de Paris, France*
[b] *Direction Innovation & Recherche SNCF, France*

ARTICLE INFO

Keywords:
Few-shot learning
Sequence labeling
Emotion recognition in conversation
Pytorch
Prototypical networks

ABSTRACT

Detecting emotions in a conversational context benefits several industrial cases such as customer service, user appraisal from speech recognition, and so on. However, in most cases, research data differ from real data due to them being private, confidential, or difficult to label. In this work we present ProtoSeq, an adaptation of the Prototypical Networks to enable dealing with sequences in a few-shot learning way, reducing the need for labeling confidential data.

Code metadata

Current code version	1.0.0
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2021-178
Permanent link to Reproducible Capsule	https://codeocean.com/capsule/4866261/tree/v1
Legal Code License	*MIT License*
Code versioning system used	Git
Software code languages, tools, and services used	Python
Compilation requirements, operating environments & dependencies	Python >= 3.8.2 with the following dependencies PyTorch 1.7.1; torchtext 0.8.1; torchCRF 1.1.0; termcolor 1.1.0; scikit-learn 0.23.1; tweet-preprocessor 0.6.0. Works on any operating systems running Python. We recommend using a Nvidia GPU with at least 4Go VRAM.
If available Link to developer documentation/manual	
Support email for questions	gael.guibon@gmail.com, gael.guibon@telecom-paris.fr

1. Introduction

One limitation of current deep learning methods is the availability of datasets to train predictive models. In the industrial area, companies often face this problem because they deal with confidential data such as medical data, security-related data, or private communication data, to name but a few. To get around this difficulty, new approaches emerged trying to alleviate the data size dependency by considering transfer learning [1], semi-supervised learning [2], meta-learning [3], or few-shot learning (FSL) [4] for instance. In this work, we focus on recognizing emotions in conversation in the context of private communication data by adapting a well-known metric-learning based meta-learning framework commonly used for few-shot learning: the Prototypical Networks [5]. We present a variation of Prototypical Networks dedicated to sequences, that we name ProtoSeq. In its core, ProtoSeq consists of Prototypical Networks [5] along with an episodic training framework [6] both adapted to enable sequences of data.

By sharing the ProtoSeq, we seek to encourage the field of Emotion Recognition in Conversation to consider the use of FSL, as opposed to all the studies using supervised learning [7–10] which make the implicit assumption that a right amount of data will be available. Moreover, with ProtoSeq we first present this approach on a text

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

[*] Corresponding author at: LTCI, Télécom-Paris, Institut Polytechnique de Paris, France.
E-mail addresses: gael.guibon@telecom-paris.fr, ext.gael.guibon@sncf.fr (G. Guibon), matthieu.labeau@telecom-paris.fr (M. Labeau), luce.lefeuvre@sncf.fr (L. Lefeuvre), chloe.clavel@telecom-paris.fr (C. Clavel).

https://doi.org/10.1016/j.simpa.2022.100237
Received 4 December 2021; Received in revised form 27 December 2021; Accepted 11 January 2022

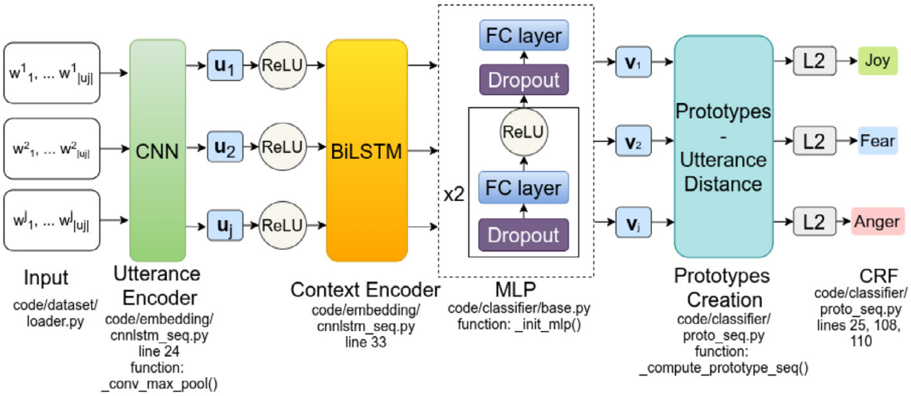
w¹₁, ... w¹|uj|
w²₁, ... w²|uj|
wʲ₁, ... wʲ|uj|
CNN
u₁
u₂
uⱼ
ReLU
BiLSTM
FC layer
Dropout
ReLU
FC layer
Dropout
x2
v₁
v₂
vⱼ
Prototypes - Utterance Distance
L2
Joy
Fear
Anger
Input
code/dataset/ loader.py
Utterance Encoder
code/embedding/ cnnlstm_seq.py line 24 function: _conv_max_pool()
Context Encoder
code/embedding/ cnnlstm_seq.py line 33
MLP
code/classifier/base.py function: _init_mlp()
Prototypes Creation
code/classifier/ proto_seq.py function: _compute_prototype_seq()
CRF
code/classifier/ proto_seq.py lines 25, 108, 110

Fig. 1. Architecture of the ProtoSeq model: from the input FastText [15] static embeddings to the CRF layer. Each part is linked to the related code in the reproducible capsule. The global training function is located in `code/train/regular.py`.

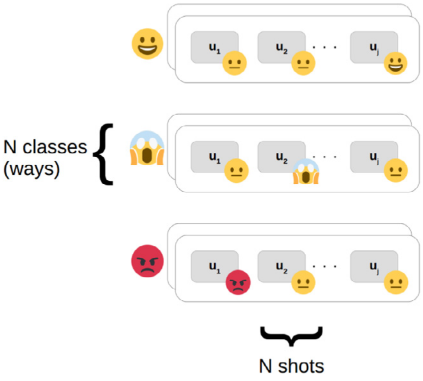

Fig. 2. Episodic training framework for sequences: an example episode (2-shots, 3-ways). Conversation labels are inferred from utterance (u) labels. Label = emoji.

modality, but share the framework in order to make it easily adaptable to different modalities such as speech or vision, the latter from which the original Prototypical Networks [5] stems from.

2. Description

ProtoSeq can be divided into two main parts: the model and the training framework. The model uses a hierarchical encoder based on convolutional networks (CNN) [11] for utterance encoding and Bi-directional Long–Short Term Memory networks (BiLSTM) [12] to adjust utterance representations with their surrounding context. This stems from recent works on sequence labeling in dialog and emotion recognition in conversation [7,13]. With the combination of an additional Multi-Layer Perceptron (MLP) and a Conditional Random Fields (CRF) [14] layer, ProtoSeq allows sequence labeling using a two-step process (overview in Fig. 1): (1) an order-aware labeling using the encoded utterance distances from the class prototypes (from input to Prototypes-Utterance Distances in Fig. 1); (2) a global context-aware sequence labeling using a CRF [14] layer to fine-tune the previous step (the last layer – CRF – in Fig. 1). We also apply and share an adapted episodic training framework [6] dedicated to sequences, where the number of instances per class for training is conditioned by the presence of at least one utterance with the relevant class in the sequence. To train the model, we generate a set of random episodes (replacing batches and reproducing the context of a few training examples) with a fixed number of random examples (*shots*) for each class (*way*) to train, and predict on a fixed number of elements (*queries*) to compute the cross entropy loss. Fig. 2 shows a training episode's *shots* and *ways* for sequences.

Considering those two main features, we introduce the ProtoSeq through a public PyTorch implementation. The software base structure is inspired from [16] but extends it in order to incorporate multiple sub-structures such as few-shot and supervised learning tasks in a more separated way. Hence, our adapted episodic strategy (see Fig. 2) can be found in `code/dataset/parallel_sampler_seq.py` while Bao's [16] implementation of Larochelle's [6] episodic strategy, the original version for non-sequences, can be found in `code/dataset/parallel_sampler.py`.

Data Utilities. Our ProtoSeq's PyTorch implementation expects data in JSON lines format. Each line should be a JSON object representing a conversation. We have opted for this format over Pandas[1] due to the hierarchical nature of the data. By default, we present an example usage on a textual conversation dataset [17] for which we share our custom parser `data/parser_gg.py`. By sharing this parser, we share another way to handle this data, which can also be used to format data from the Datasets library.[2] We also share the function dedicated to data preprocessing `creaDailyDialogSeq()` in `emotionClf.py`, along with the unique fully parsed and preprocessed data. An option is dedicated to reproduce it: `python3 emotionClf.py --task prepa_dataset`.

Additional Implementations. When comparing existing supervised approaches used in emotion recognition in conversation, the public KET [9] implementation[3] dropped from 53.37% to 41.43% in micro F1-score when applied on our private confidential dataset. We tried to measure it for CESTa, but found no available public implementation. Thus, we share our personal CESTa implementation[4] that we made from scratch by following the original paper's instructions. However, it did not yield good results neither on our private confidential dataset [18] nor on DailyDialog [17], which seems to contradict the original paper's results. As far as we know, this is the only available implementation of CESTa, this is why we share it to prove this did not work on our specific data, but also to allow possible improvements from the research community, which can re-apply it or modify it.

Our reproducible capsule also comes with several ProtoSeq variants that the user can trigger to try different encoders from a simple average of input representations to multiple Transformer [19] encoder layers.

3. Impact

With ProtoSeq, we wish to offer an example solution to deal with data privacy shortage and have an indication of the differences in performance. Most conversational data are private, unlabeled, and differ

[1] https://pandas.pydata.org/.
[2] https://huggingface.co/datasets/daily_dialog.
[3] https://github.com/zhongpeixiang/KET.
[4] Available in `code/classifier/cesta.py` with inline comments.

from the available research datasets. Hence, we seek to encourage the research field to consider the use of few-shot learning for this task by sharing ProtoSeq, which achieves 31.81% in micro f1-score (excluding the majority class) compared to 26.07% for WarmProto-CRF, another FSL method for sequence [18,20]. In our implementation, default data is expected to be hierarchical textual data (nested lists of tokens). However, it can be modified to handle any kind of data as the model is not restricted to textual data. It can be used for other modalities as long as the properties (hierarchical sequences) are met. Even if we restrict ourselves to the scope of Natural Language Processing and, more precisely, to Sentiment Analysis and Emotion Recognition, multiple other tasks are still available: named entity recognition, speaker's speech pattern recognition, part-of-speech tagging, for instance. ProtoSeq is designed to be used for private data where only a few labeled examples are available. Moreover, it is the first software on few-shot emotion recognition in conversation, which makes it a pioneer in this research sub field. This is why this work is motivated by both academic and industrial objectives and we expect it to make this research field lean towards real data usage with performance comparison to artificial (hence shareable) data, instead of only comparing performance on the latter. This software aims at sharing a baseline for other studies to compare from directly or indirectly [21]. To do so, we make the code easy to reuse.

4. Current limitations

The two parts of the ProtoSeq each have one limitation. Firstly, the adapted Prototypical Networks possess a final CRF layer which overwrites the order information from the previous hierarchical encoder. This means, even if the order is used to determine the representations, the sequence labeling phase ignores it almost completely at the end. Secondly, the adapted episodic setting yields a variable number of utterances per class. This means the strict balance between classes from the standard episodic strategy [6] is lost due to the variable number of instances per external classes in a sequence.

5. Conclusion and future improvements

In this work we present ProtoSeq, the first application of few-shot learning for emotion recognition in conversation with the hope of promoting the need of such approach in this task. This software stems from the industrial context where we do not have access to enough data due to privacy limitations. ProtoSeq is made of both a hierarchical model and a training framework for which we will try to lessen the inherent limitations in regards to element order overwriting, and variable number of utterances in the episodes.

CRediT authorship contribution statement

Gaël Guibon: Software, Methodology, Conceptualization, Data curation, Resources, Writing – original draft, Writing – review & editing. **Matthieu Labeau:** Methodology, Conceptualization, Writing – review & editing, Supervision, Project administration, Formal analysis, Validation. **Luce Lefeuvre:** Methodology, Conceptualization, Writing – review & editing, Supervision, Project administration, Validation. **Chloé Clavel:** Methodology, Conceptualization, Writing – review & editing, Supervision, Project administration, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This project has received funding from SNCF (the French National Railway Company), the French National Research Agency's grant ANR-17-MAOI, and the DSAIDIS, France chair at Télécom-Paris.

References

[1] S. Ruder, M.E. Peters, S. Swayamdipta, T. Wolf, Transfer learning in natural language processing, in: Proceedings Of The 2019 Conference Of The North American Chapter Of The Association For Computational Linguistics: Tutorials, Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 15–18, http://dx.doi.org/10.18653/v1/N19-5004, URL https://aclanthology.org/N19-5004.

[2] J.E. Van Engelen, H.H. Hoos, A survey on semi-supervised learning, Mach. Learn. 109 (2) (2020) 373–440.

[3] T. Hospedales, A. Antoniou, P. Micaelli, A. Storkey, Meta-learning in neural networks: A survey, 2020, arXiv:2004.05439 [Cs, Stat]. ArXiv: 2004.05439. URL http://arxiv.org/abs/2004.05439.

[4] Y. Wang, Q. Yao, J.T. Kwok, L.M. Ni, Generalizing from a few examples: A survey on few-shot learning, ACM Comput. Surv. 53 (3) (2020) 1–34.

[5] J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances In Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 4077–4087, URL http://papers.nips.cc/paper/6996-prototypical-networks-for-few-shot-learning.pdf.

[6] S. Ravi, H. Larochelle, Optimization as a model for few-shot learning, OpenReview (2016).

[7] S. Poria, E. Cambria, D. Hazarika, N. Majumder, A. Zadeh, L.-P. Morency, Context-dependent sentiment analysis in user-generated videos, in: Proceedings Of The 55th Annual Meeting Of The Association For Computational Linguistics (Volume 1: Long Papers), 2017, pp. 873–883.

[8] N. Majumder, S. Poria, D. Hazarika, R. Mihalcea, A. Gelbukh, E. Cambria, DialogueRNN: An attentive RNN for emotion detection in conversations, in: Proceedings Of The AAAI Conference On Artificial Intelligence, vol. 33, 2019, pp. 6818–6825, http://dx.doi.org/10.1609/aaai.v33i01.33016818, URL https://aaai.org/ojs/index.php/AAAI/article/view/4657.

[9] P. Zhong, D. Wang, C. Miao, Knowledge-enriched transformer for emotion detection in textual conversations, 2019, Cs. ArXiv: 1909.10681. URL http://arxiv.org/abs/1909.10681.

[10] Y. Wang, J. Zhang, J. Ma, S. Wang, J. Xiao, Contextualized emotion recognition in conversation as sequence tagging, in: Proceedings Of The 21th Annual Meeting Of The Special Interest Group On Discourse And Dialogue, 2020, pp. 186–195.

[11] Y. Kim, Convolutional neural networks for sentence classification, 2014, arXiv preprint arXiv:1408.5882.

[12] Z. Huang, W. Xu, K. Yu, Bidirectional LSTM-CRF models for sequence tagging, 2015, arXiv:1508.01991.

[13] E. Chapuis, P. Colombo, M. Manica, M. Labeau, C. Clavel, Hierarchical pre-training for sequence labelling in spoken dialog, in: Proceedings Of The 2020 Conference On Empirical Methods In Natural Language Processing: Findings, 2020, pp. 2636–2648.

[14] J. Lafferty, A. McCallum, F.C. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, 2001.

[15] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching word vectors with subword information, 2017, arXiv:1607.04606.

[16] Y. Bao, M. Wu, S. Chang, R. Barzilay, Few-shot text classification with distributional signatures, in: International Conference On Learning Representations, 2020.

[17] Y. Li, H. Su, X. Shen, W. Li, Z. Cao, S. Niu, DailyDialog: A manually labelled multi-turn dialogue dataset, 2017, Cs. ArXiv: 1710.03957. URL http://arxiv.org/abs/1710.03957.

[18] G. Guibon, M. Labeau, H. Flamein, L. Lefeuvre, C. Clavel, Few-shot emotion recognition in conversation with sequential prototypical networks, in: Proceedings Of The 2021 Conference On Empirical Methods In Natural Language Processing, 2021, pp. 6858–6870.

[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances In Neural Information Processing Systems, 2017, pp. 5998–6008.

[20] A. Fritzler, V. Logacheva, M. Kretov, Few-shot classification in named entity recognition task, in: Proceedings Of The 34th ACM/SIGAPP Symposium On Applied Computing, 2019, pp. 993–1000.

[21] J. Olah, S. Baruah, D. Bose, S. Narayanan, Cross domain emotion recognition using few shot knowledge transfer, 2021, arXiv:2110.05021.