**Few-shot emotion recognition in conversation with sequential prototypical networks**

Gael 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

**A R T I C L E I N F O**

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

**A B S T R A C T**

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.

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...
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 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 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 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 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
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 DSADIS, France chair at Télécom-Paris.

References