BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2024
http://excel.fit.vutbr.cz

Extraction of key information from emergency calls

Bc. Marek Sarvaš*

Abstract

Emergency line agents are often unable to process all information given in the first seconds of a call by a distressed caller. The aim of this work is to automatically extract important data to speed up the reaction of emergency rescue teams. This problem is formulated as a detection and correct classification of predefined key information in the call using various deep neural network models. The best model achieved a 0.909 F1 score for the person and a 0.758 F1 score for location extraction. Such models can be easily adapted to detect other key information based on emergency services needs. Additionally, multiple datasets were created from real emergency call data to improve the model and will be used to incorporate better neural network-based models as a helping hand for emergency line agents.

*xsarva00@vut.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

People calling the emergency line are often stressed and want to get help as soon as possible. To achieve this, the natural response when the emergency line agent picks up the call is to dump all the available information on the agent in a matter of seconds. However, this has the opposite effect, as human operators cannot process everything in such a short time. This leads to the agent asking questions even though the information may have already been given by the caller. The goal of this work is to automatically detect key information (known as Named Entity Recognition), such as the names of the callers or their location, in order to help the agent and speed up the emergency services response.

This work explores various neural network-based approaches, such as encoder-based, sequence-to-sequence, and Large Language models, and evaluates them on publicly available Czech coNLL-based CNEC2.0 [1] dataset, achieving the SOTA results. In addition, the encoder-based model achieved reasonably good results on the emergency data. Furthermore, several new emergency call datasets were created in the process.

2. Preliminaries

Named entity recognition (NER) is a Natural Language Processing (NLP) task where the system (in this work, it is a neural network model) detects and correctly classifies entities of interest into predefined categories. Examples of such categories are **person**, **location**, **organization** and other depending on the specified goal. This task operates at a word level, classifying each word as an entity of a specific type or with a general label that represents a word that is not part of any entity, usually with "**O**" tag. It is important to distinguish between the start of a new entity and an entity composed of multiple words. There are several formats for representing entities. The format used in this work is BIO2, which means that the tag of a first word of every entity has **B-** prefix while the tags of the remaining words have **I-** prefix.

2.1 Evaluation

In a NER task, entities usually represent only a minority of the data, and most of the words in train and test utterances are not considered entities. In such cases, the models are evaluated using F1, precision, and recall scores presented in the following equations 1,2, and 3 respectively.

$$F_1 = \frac{PR}{P+R} \tag{1}$$

$$P = \frac{|true\ positives|}{|true\ positives| + |false\ positives|} \tag{2}$$

$$R = \frac{|true\ positives|}{|true\ positives| + |false\ negatives|} \quad (3)$$

3. Experimental pipeline

The experimental pipeline presented in Figure 2 is made up of several parts that are subject to this work. Two main subparts are the creation of a new emergency line dataset and the NER model itself (presented in blue).

3.1 Models

Given the data sensitivity of emergency line calls and the fact that there was no dataset in the beginning, the models were first trained and evaluated on the coNLL-based CNEC2.0 [1] dataset. Three different approaches were tested, each with a different architecture. First, a token classification approach was used in which the model is a pre-trained encoder transformer (XLM-R [2]) with a linear classification layer on top. This is a very common approach for NER, classifying every word, which minimizes the post-processing of the model output. The second approach is a sequence-to-sequence model mT5 [3], where the model generates entity tags directly into the input text, as shown in the second example in the Anonymized examples of real data section.

Based on the rising popularity and performance of Large Language Models (LLMs) in recent years, this approach was also explored for the extraction of entities directly into the json format. However, at the time of making this poster, only inference with an instruction fine-tuned Mistral-7B [4] LLM was tested, resulting in a poor performance compared to other approaches.

3.2 Dataset

In the beginning, the only available data were .mp3 recordings of emergency calls with structured metadata stored in CSV files. The NER models operate on text data, therefore, the first step was to obtain transcriptions. This was done using the ASR model provided by Phonexia. A large amount of data were available from 2020 and 2022 (shown in Figure 1), but only a small subset was selected for further processing. The next step was to create NER labels to evaluate and fine-tune the models. One set of labels was created with the NER model pre-trained on the coNLL-based CNEC2.0 [1] dataset. Another set of labels was created by matching the entities from the metadata to the transcriptions. Both sets of labels were combined preferring the labels from metadata, in cases where the labels from metadata and model did not match.

Finally, the labels for this dataset were manually corrected and divided into train and test sets. Additionally, based on the first baseline results on this dataset, the train set was expanded with the augmented utterances. The augmentation consists mainly of adding utterances containing exclamations and special words such as "nashledanou" (meaning goodbye) that occur very frequently in real data and are miss-classified by the model in most cases.

4. Results

On the coNLL-based CNEC2.0 [1] dataset the token classification approach achieved the highest F1, precision, and recall scores, so it was selected for testing on emergency call datasets. Table 1 shows results on emergency call datasets obtained with XLM-R model [2] with multiple variations of training data. The **tc** model was trained on original truecase version of coNLL-based CNEC2.0 [1] dataset. The **tc, lc** was additionally trained on the combination of lowercase and truecase versions of the dataset. Truecase version has achieved overall better precision scores but much lower recalls than the mixed-case version. This may seem like the truecase model was performing better, but the low precision of the mixed-case model is due to the high rate of false positives introduced by training the model also on lowercase utterances. When comparing the performance of both models on the manually labeled dataset, shown in the second half of Table 1, the mixed-case version of XLM-R [2] surpassed the original version by a large margin. Finally, **finetuned** versions were trained on combination of truecase and lowercase versions of coNLL-based CNEC2.0 [1] followed by finetuning on train subset of emergency call data, achieved 0.909 F1 and 0.758 F1 for person and location entities, respectively. As expected, finetuning on in-domain data helps, but the performance is highly dependent on the quality of labels, which is mainly visible in the case of location entities.

5. Conclusions

I created a clean emergency call NER train and test datasets. In the process, tens of thousands of call transcriptions were created, which could be useful for training other models. In addition, I managed to train models that perform well on real-world data. At the time of submission of this article, the LLM approach was not yet fully explored, and it is a subject of next experiments.

References

[1] Michal Konkol and Miloslav Konopík. Crf-based czech named entity recognizer and consolidation of czech ner research. In Ivan Habernal and Václav Matoušek, editors, *Text, Speech, and Dialogue*, pages 153–160, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[2] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. *CoRR*, abs/1911.02116, 2019.

[3] Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. mt5: A massively multilingual pre-trained text-to-text transformer, 2021.

[4] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. Mistral 7b, 2023.