Towards the automatic extraction of procedural knowledge from robotic-assisted surgical texts

Bombieri, M.[1], Rospocher, M.[1], Dall'Alba, D.[1] & Fiorini, P.[1]
[1] *University of Verona, Italy.*

HAMLYN SYMPOSIUM ON MEDICAL ROBOTICS 2021

Introduction

The automatic extraction of knowledge about intervention execution from surgical manuals would be of the utmost importance to develop expert surgical systems and assistants. In this work we assess the feasibility of automatically identifying the sentences containing procedural information from a written description of a surgical intervention.

Thousands of different types of surgical procedures are performed daily in hospitals around the world. These procedures are typically described in detail in written resources, such as books, manuals, academic papers and online resources. Typically, the description of a procedure details:

- how to perform the intervention
- which anatomical structures to operate
- which tools to use.

The manual extraction of this information requires substantial human effort and expertise, hindering its application at scale. Towards the problem of automatic extraction of surgical procedure, we propose a first solution to the important sub-task of recognizing sentences containing procedural knowledge.

Method

We tackle the problem of procedural knowledge detection in surgical written text as a **sentence classification** task.
In order to design and test such classification approach, a dataset of sentences labeled as procedural/non-procedural is needed. We manually constructed and annotated a new public dataset, called *SPKS* (Surgical Procedural Knowledge Sentences) composed by 1,958 sentences) from different textual resources available in literature [1].

Exploiting SPKS, we proposed several procedural knowledge detection algorithms in order to make a comparative survey. In particular, we compared:

- Classical machine learning methods fed with TF-IDF features;
- FastText classifier [2] with subword embeddings;
- 1-Dimension Convolutional Neural-Network fed with FastText's embeddings
- Bi-directional Long short-term memory Neural-Network fed with FastText's embeddings
- Transformer-based classifiers using pre-trained language models (BERT [3] and ClinicalBERT [4]) and fine-tuning on SPKS training samples.

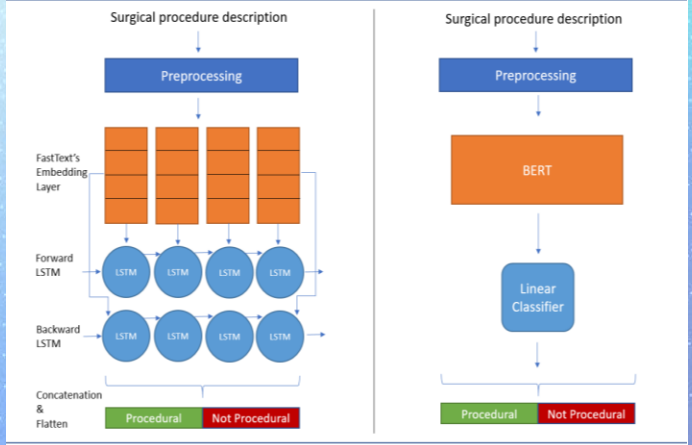


Figure 1 - Architectural schemes illustrating models used for detecting procedural sentences in texts exploiting Bi-LSTMs and Transformers' based models (B).

Results and discussions

In the first and second rows of Table 1 (in yellow) [5], we report the classification performances of classical machine-learning algorithms *Random-Forest* (Ra-Fo) and *Multinomial Naïve-Bayes* (Mul-NB) fed with TF-IDF features. The considered approaches have mediocre performances when used to solve this task.
The third row of Table 1 (in green) summarizes the performance of the FastText classifier. All scores demonstrate that it obtains much higher classification performance than the previous two. We then fed the FastText's word embeddings learned on the SPKS dataset to a 1D-CNN and a Bi-LSTM (blue rows in the Table). Using more complex classification models allows to substantially improve performances, but at the cost of more computational time required.
Finally, the pink rows of the table show that it is possible to achieve high classification performance, also using transformer-based pre-trained language models. In particular, ClinicalBERT performs slightly better than BERT, as somehow expected, given the characteristics of the source material used for pre-training the model: BERT is pre-trained on general domain texts, while ClinicalBERT is pre-trained on clinical notes and Electronic Health Records. While still different from surgical procedure descriptions, these texts are certainly closer to our domain than those used for training BERT.

Method	Acc.	M-Pre	M-Rec	M-F1	W-Pre	W-Rec	W-F1
Ra-Fo	0.740	0.743	0.678	0.686	0.741	0.740	0.721
Mul-NB	0.737	0.785	0.655	0.657	0.767	0.737	0.701
FastText	0.786	0.771	0.765	0.767	0.784	0.786	0.785
1D-CNN	0.829	0.816	0.828	0.820	0.835	0.829	0.831
Bi-LSTM	0.867	0.857	0.856	0.857	0.867	0.867	0.867
BERT	0.864	0.859	0.845	0.851	0.863	0.864	0.862
Cl- BERT	0.872	0.866	0.856	0.860	0.871	0.871	0.871

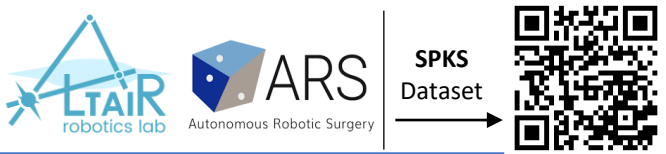
Table 1 - Classification performance of the tested methods.
Acc = Accuracy; M-Pre = Mean-Precision; M-Rec = Mean-Recall; M-F1 = Mean-F1; W-Pre = Weighted-Precision; W-Rec = Weighted-Recall; W-F1 = Weighted-F1

Conclusions and Future Works

The objective of this work was not to identify the best possible algorithm to tackle the problem of procedural knowledge detection in surgical texts. Our goal was indeed to provide a first assessment of the feasibility of this task using competitive methods. The obtained result can still be improved, and we identify the following directions for future works:

- Enlarging the dataset to allow more precise training.
- To integrate additional context-related features with information solely taken from sentences.
- To create a BERT model specifically trained from scratch on surgical procedural language.

Finally, we underline that this work is only a preparatory activity towards the long-term goal of extracting structured surgical intervention workflows from written procedural documents, a challenging and, to the best of our knowledge, never studied in the surgical domain.



This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 742671 "ARS")

References

[1] https://gitlab.com/altairLab/spks-dataset
[2] Joulin, A. et al. (2017). Bag of tricks for efficient text classification. *Proceedings of the 15th EACL Conference, Valencia, Spain.*
[3] Devlin, J. et al. (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the NAACL-HLT 2019 Conference*
[4] Alsentzer, E. et al. (2019). Publicly available clinical BERT embeddings. *Proceedings of the 2nd Clinical NLP Workshop Minneapolis, Minnesota, USA*
[5] Bombieri, M. et al. (2021). Automatic detection of procedural knowledge in robotic-assisted surgical texts. *Int. J. of CARS*