Text Classification Models for Form Entity Linking

María Villota, César Domínguez, Jónathan Heras, Eloy Mata, Vico Pascual
Department of Mathematics and Computer Science, University of La Rioja, Spain
{maria.villota, cesar.dominguez, jonathan.heras, eloy.mata, vico.pascual}@unirioja.es

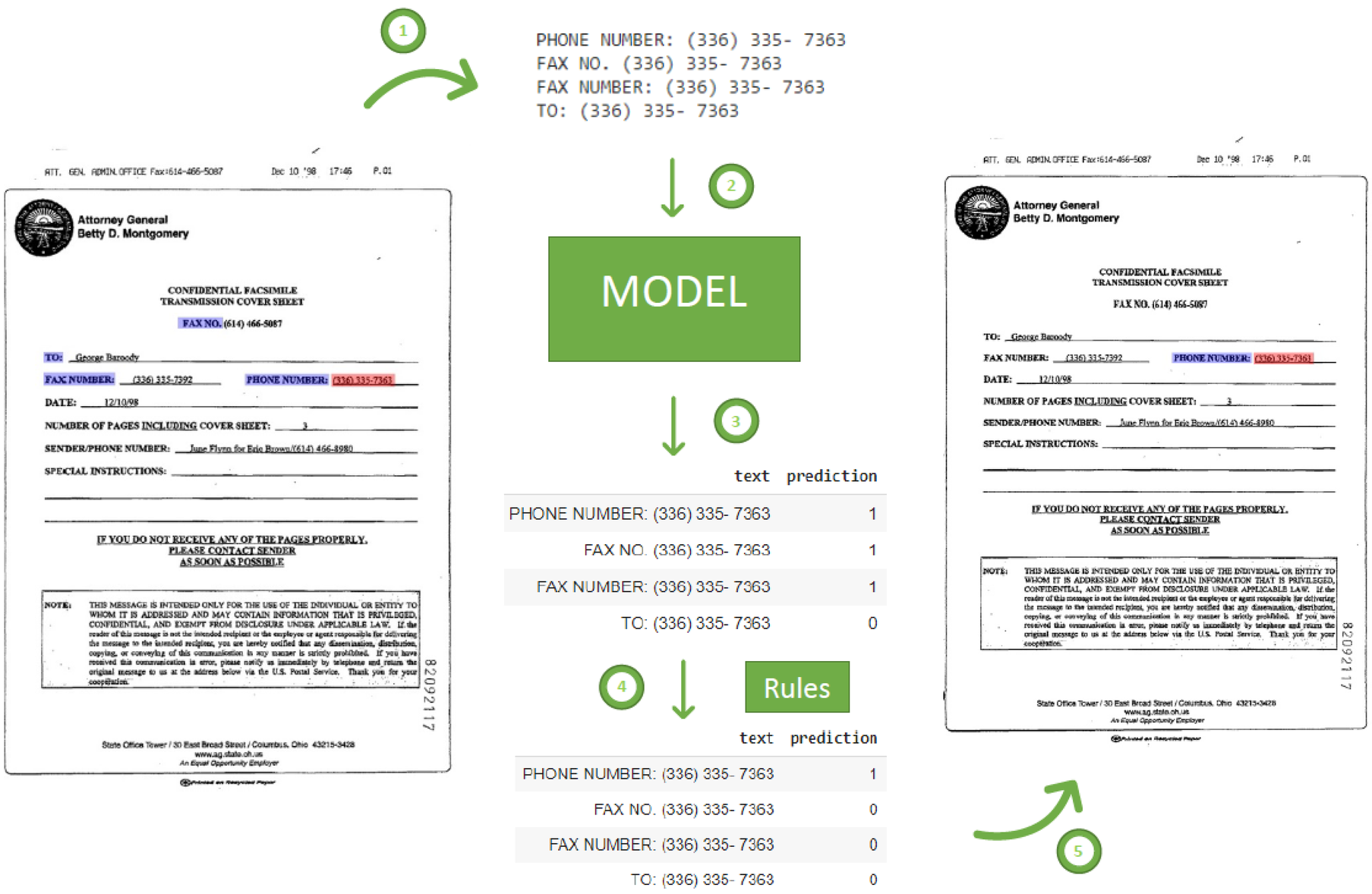
DAS 2022
May 22-25, La Rochelle

Abstract

Forms are a widespread type of template-based document used in a great variety of fields including, among others, administration, medicine, finance, or insurance. The automatic extraction of the information included in these documents is greatly demanded due to the increasing volume of forms that are generated in a daily basis. However, this is not a straightforward task when working with scanned forms because of the great diversity of templates with different location of form entities, and the quality of the scanned documents. In this context, there is a feature that is shared by all forms: they contain a collection of interlinked entities built as key-value (or label-value) pairs, together with other entities such as headers or images. In this work, we have tackled the problem of entity linking in forms by combining image processing techniques and a text classification model based on the BERT architecture. This approach achieves state-of-the-art results with a F1-score of 0.80 on the FUNSD dataset, a 5 % improvement regarding the best previous method. The code of this project is available at `https://github.com/mavillot/FUNSD-Entity-Linking`.

Method

The method proposed in this work consists of the following steps: (1) From an answer, a set of candidate questions are identified. (2) Each combination of candidate question-answer is fed to a text classification model that (3) identifies the valid combinations of question-answer. (4) If more than one combination is valid, we apply a set of rules based on the distance between the question and the answer. (5) Finally, the results are returned.



Results

Results achieved for entity linking by the tested text classification models. In bold face the best results.

	mAP	mRank	F1-score
BERT	**0.87**	**0.49**	**0.80**
DistilBERT	0.79	0.79	0.68
DistilRoBerta	0.76	0.95	0.65
LayoutLM	0.79	0.81	0.69
RoBerta	0.77	0.94	0.66

Comparison

Comparison of our approach with existing methods for entity linking. In bold face the best results.

	mAP	mRank	F1-score
BROS	-	-	0.67
Carbonell et al.	-	-	0.39
FUDGE	-	-	0.62
FUNSD paper	0.23	11.68	0.04
DocStruct Model	0.72	2.89	-
LayoutLM Word Level	0.47	7.11	-
MSAU-PAF	-	-	0.75
MTL-FoUn	0.71	1.32	0.65
Sequential Model	0.65	1.45	0.61
SPADE	-	-	0.41
Ours	**0.87**	**0.49**	**0.80**

Conclusions

- We have proposed a method for form entity linking based on the combination of image processing techniques and text classification models.
- We tested several transformer-based models for text classification and reached the conclusion that a model based on the BERT architecture produces the best results.
- This approach has achieved state-of-the-art results for form entity linking in the FUNSD dataset.
- We have shown the benefits of combining deep learning models with algorithms based on the existing knowledge about documents when working in contexts where annotated data is scarce.

This work was partially supported by Grant RTC-2017-6640-7; and by MCIN/AEI/10.13039/501100011033, under Grant PID2020-115225RB-I00.