ESAN: Automating medical scribing in Spanish

ESAN: Automatización de la toma de notas clínicas

Naiara Perez[1], Aitor Álvarez[1], Arantza del Pozo[1], Andrés Arbona[2], Oihane Ibarrola[2], Marta Suarez[2], Pedro de la Peña Tejada[3] and Itziar Cuenca[3]

[1]Fundación Vicomtech, Basque Research and Technology Alliance (BRTA), Donostia-San Sebastián, 20009, Spain
[2]Biokeralty Research Institute AIE, Vitoria-Gasteiz, 01510, Spain
[3]Instituto Ibermática de Innovación (i3B), Donostia-San Sebastián, 20009, Spain

Abstract

The ESAN research project aims at developing a Spanish digital scribe that reduces the administrative workload of clinicians and enhances the quality of the data collected in the medical records by automatically transcribing and structuring doctor-patient conversations. At present, the main goal of the consortium consists in collecting and annotating the data necessary for training and adapting speech and natural language processing models based on deep learning architectures.

Keywords

clinical data, EHR, speech recognition, data mining

1. Introduction

The past few decades have seen a worldwide, steady growth in the adoption of electronic health record (EHR) systems, with the ultimate goal of improving the efficiency and quality of the provided care. In spite of their many virtues, EHRs have also increased the administrative workload of healthcare professionals, to the point of having been identified as a direct cause of burnout and lack of meaningful doctor-patient eye contact [1, among others].

Meanwhile, the accumulation of massive amounts of digitised health records in the era of Big Data has boosted the pursuit of public policies aimed at accelerating the advent of new healthcare paradigms such as personalised medicine. Yet Big Data is no more profitable than the quality of the data allows. Currently, much of the data collected in EHRs is in the form of free text written in haste. It makes irregular use of grammar, standard medical terminology, and of the EHR structure itself. It may omit information that is not of evident immediate value. Moreover, it is barely codified (if at all), all of which hinders its automated exploitation.

More recently, the major and rapid advances of Deep Learning have prompted a surge of interest in the application of artificial intelligence to medical conversations, so much so that several tech giants have recently launched a workshop exclusively focused on this research topic [2, 3].

In this context we present the ESAN project (from "EStructuración de conversaciones en el ámbito SANitario" or *Structuring conversations in the health sector* in Spanish, but also "esan" or *say, tell* in Basque). ESAN is the first step of a joint, long-term effort towards alleviating the above introduced problems through the research and development of a Spanish digital scribe.

2. Consortium and funding body

ESAN is partially funded by the Basque Government through the Elkartek 2021 program of the SPRI Group under the grant agreement KK-2021/00117. It will run from 04/2021 to 12/2023.

The consortium includes the Vicomtech research centre, (https://www.vicomtech.org), Grupo Keralty's R&D division BioKeralty Research Institute (https://biokeralty.com), and Grupo Ibermática's R&D business unit and project leader Instituto Ibermática de Innovación or i3B (https://ibermatica.com/en/innovacion/).

SEPLN-PD 2022. Annual Conference of the Spanish Association for Natural Language Processing 2022: Projects and Demonstrations, September 21-23, 2022, A Coruña, Spain

nperez@vicomtech.org (N. Perez); aalvarez@vicomtech.org (A. Álvarez); adelpozo@vicomtech.org (A. d. Pozo); andres.arbona@keralty.com (A. Arbona); oihane.ibarrolv@keralty.com (O. Ibarrola); marta.suarez@keralty.com (M. Suarez); pm.delapena@ibermatica.com (P. d. l. P. Tejada); ia.cuenca@ibermatica.com (I. Cuenca)
0000-0001-8648-0428 (N. Perez); 0000-0002-7938-4486 (A. Álvarez)

3. Goals and expected results

The long-term, main technical objective of the ESAN consortium is to develop a Spanish digital scribe. A digital scribe is, in short, a program capable of documenting the encounter between a patient and their doctor or nurse. It involves Automatic Speech Recognition (ASR) to transcribe the conversations, and Natural Language Processing (NLP) to understand and transform those transcripts as necessary (e.g., extract relevant information and classify it into EHR sections).

At this early stage, the identified challenges of the project (see §4) point primarily to the need for problem-specific data and the lack thereof. Thus, the focus of this initial venture of the ESAN consortium is set on building a new corpus. The expected results of this line of work are:

- 150 hours of anonymised recordings (∼1K encounters) in 4 medical specialities, along with their manual, enriched transcripts and the corresponding written medical notes, all in Spanish.
- Guidelines for the annotation of the dialogues regarding the information extraction (IE) and classification tasks, as well as the manual annotations resulting from their application.

Second, we plan to train benchmark models for enriched ASR and IE adapted to the application scenarios of ESAN, exploiting primarily the aforesaid corpus and other publicly available data that might be considered beneficial.. The specific expected results in this regard are:

- Robust neural models for enriched ASR adapted to face-to-face clinician-patient conversations in Spanish, including automatic capitalisation and punctuation, and supervised diarisation.
- Initial neural IE and classification models to transform the dialogue transcripts into structured data that can be fed to an EHR.

The third and final major goal is to flesh out the next steps based on quantitative and qualitative evaluations of the obtained technology. The expected final outcome is then:

- A road map towards productisation, taking into account the performance of the ASR and NLP models and other aspects that are outside the current scope (e.g., usability, communication standards).

4. Challenges

The challenges faced by the ESAN research project are twofold because it must overcome major ethical and legal obstacles in addition to the scientific and technological.

Conversations between patients and their doctors are among the most sensitive pieces of information conceivable. Voice recordings alone qualify largely as personal data according to the many policies that we are subject to, from the international (e.g., the GDPR of the European Union) to the local (e.g., ethics committees). This means that there is no public dataset that we can leverage, and that we must overcome these ethical and legal barriers in order to collect it ourselves.

Regarding the scientific and technical challenges, at this stage of the project, the difficulties of developing a Spanish digital scribe stem also from the nature of the data to be processed, in all its facets:

Genre The input to the scribe is spontaneous speech produced in the context of a dialogue between two or more people. Current ASR technology still struggles in this scenario due to *a)* the difficulty to obtain quality audio, where all the interlocutors are recorded with optimal volume and energy and *b)* linguistic phenomena inherent to spontaneous speech (overlapping, false starts, repetitions, etc.). Human-human conversations are a serious challenge for NLP systems too for similar reasons. For example, questions may go unanswered or be answered at a later point in the dialogue, or relevant information may be transmitted through non-verbal means.

Domain Along with the genre, the highly specialised application domain constitutes the key defining challenge of ESAN. Out-of-the-box, generic ASR and NLP solutions are not viable here simply because they are not prepared to deal with the specialised vocabulary and the extraction or classification targets of the clinical domain. Further, building new solutions and resources requires at least the guidance of expert knowledge.

Register Conversations in consultations present the added difficulty that doctors tend to address their patients in technical terms, while the patients may be less formal and employ more colloquialisms. From the perspective of the technologies involved in the project, this discursive gap is translated into an increased range of vocabulary and semantic complexity that the automated systems must recognise and understand.

Language The ESAN consortium expects to gather data in—and, ultimately, be able to process—multiple varieties of the Spanish language, including the Colombian. The differences in pronunciation and vocabulary with standard Castilian Spanish pose an added important challenge both to ASR and NLP technologies and serve only to aggravate the problems listed above.

To these concerns, we must add the fact that the errors of the enriched ASR modules are cascaded down the pipeline to the text processing modules. In addition, it is noteworthy that the health sector is most demanding and intolerant of errors, due to the gravity of the consequences that could follow from decisions based on inaccurate data.

5. Approach

5.1. Audio collection

This is the most crucial yet sensitive task of the project. The strategy involves recording real doctor-patient encounters of at least 4 specialities in a private hospital.

Measures have been taken towards minimising the impact that this activity might have on the doctors' primary job, such as training dedicated staff responsible for informing the patients about ESAN and asking for their consent in the waiting rooms, prior to meeting their doctors.

In addition, we have already tested a variety of commercial microphone arrays both in terms of quality and user-friendliness, so as to ensure their suitability before starting the audio collection campaign. We will make the recordings with the audio software Audacity (https://www.audacityteam.org) in PCM WAV format at 44.1kHz and 24 bits.

5.2. Enriched ASR

The ASR models will be trained with the 150 hours of acoustic corpus to be recorded during the project.

This corpus will be manually annotated through the Transcriber 1.5.1 tool (http://trans.sourceforge.net) with spoken literal transcriptions and speaker turn information. The annotation process will be assisted by ASR technology, which will be iteratively enhanced as new annotated audio sets are generated: the first set of drafts to be post-edited will be created with generic Castilian Spanish recognition models; once each set is manually corrected, new adapted versions of the ASR models will be trained incrementally. This process, aimed at making the annotation task more productive, will be repeated until all hours are manually revised.

The ASR models will be built using the *nnet3* DNN setup of the Kaldi recognition toolkit [4] following our previous approach based on CNN layers and a TDNN-F network [5]. The ASR engine will also include n-gram language models for decoding and re-scoring the initial lattices. The transcriptions will be enriched with capitalisation and punctuation marks generated by the BERT-based AutoPunct system [6], which will be also adapted to the domain. Finally, new speaker diarisation models will be trained for the Kaldi X-Vectors-based system [7] to be developed.

5.3. From transcripts to the EHR

The corpus of transcribed dialogues will be manually annotated at a later stage to serve as training and testing data of IE and classification models.

The annotation policy, whose precise definition is another key task of ESAN, will be built around related efforts [8, 9, 10]. It will define guidelines for the annotation of information at different levels, including mentions of signs and symptoms, disorders, and medications, as well as related attributes (severity, location, dosage, etc.).

The models for the automatic detection and classification of this information will be based on the ubiquitous Transformers architecture [11]. We plan on exploiting the latest neural language models for Spanish and the biomedical domain [12, 13]. This line of work will also profit from previous work of consortium members on clinical IE [14, 15, 16].

5.4. Validation

Each of the above-mentioned technological modules will be assessed in isolation with gold standard data and the appropriate metrics (e.g., WER, F1-score) during their development. We will also measure the impact of the errors propagated from the ASR down the processing pipeline.

Equally, if not more, important in order to flesh out the productisation road map, we will carry out a qualitative evaluation of the technology as an integrated solution prototype. To that end, we intend to devise an initial integration of all the core modules, and to develop a graphic user interface for demonstration and testing purposes, through which expert testers will be able to identify potential areas of improvement.

6. Conclusions

We have presented the ESAN project, whose aim is to develop a Spanish digital scribe that reduces the

administrative workload of clinicians and enhances the quality of the data collected in the EHRs.

The envisaged solution consists of a neural enriched ASR component followed by IE and classification modules, based too on neural architectures. To that end, the consortium will devote significant resources and effort to gathering the data necessary for adapting this technology to the challenging domain that doctor-patient face-to-face conversations pose. This emphasis on data collection and domain adaptation sets ESAN apart from related projects [17, among others].

Acknowledgments

ESAN is partially funded by the Basque Business Development Agency, SPRI, under the grant agreement KK-2021/00117.

References

[1] C. Sinsky, L. Colligan, L. Li, M. Prgomet, S. Reynolds, L. Goeders, J. Westbrook, M. Tutty, G. Blike, Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties, Ann Intern Med 165 (2016) 753–760.

[2] P. Bhatia, S. Lin, R. Gangadharaiah, B. Wallace, I. Shafran, C. Shivade, N. Du, M. Diab (Eds.), Proceedings of the 1st Workshop on NLPMC, 2020.

[3] C. Shivade, R. Gangadharaiah, S. Gella, S. Konam, S. Yuan, Y. Zhang, P. Bhatia, B. Wallace (Eds.), Proceedings of the 2nd Workshop on NLPMC, 2021.

[4] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, K. Vesely, The kaldi speech recognition toolkit, in: Proceedings of IEEE ASRU, 2011, pp. 1–4.

[5] A. Álvarez, H. Arzelus, I. G. Torre, A. González-Docasal, Evaluating novel speech transcription architectures on the Spanish RTVE2020 Database, Appl. Sci. 12 (2022) 1–16.

[6] A. González-Docasal, A. García-Pablos, H. Arzelus, A. Álvarez, AutoPunct: A BERT-based automatic punctuation and capitalisation system for Spanish and Basque, Proces. de Leng. Nat. 67 (2021) 59–68.

[7] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, S. Khudanpur, X-Vectors: Robust DNN embeddings for speaker recognition, in: Proceedings of ICASSP, 2018, pp. 5329–5333.

[8] I. Shafran, N. Du, L. Tran, A. Perry, L. Keyes, M. Knichel, A. Domin, L. Huang, Y.-h. Chen, G. Li, M. Wang, L. El Shafey, H. Soltau, J. S. Paul, The Medical Scribe: Corpus development and model performance analyses, in: Proceedings of LREC, 2020, pp. 2036–2044.

[9] P. Chocrón, Á. Abella, G. de Maeztu, ContextMEL: Classifying contextual modifiers in clinical text, Proces. de Leng. Nat. 65 (2020) 45–52.

[10] B. Magnini, B. Altuna, A. Lavelli, M. Speranza, R. Zanoli, The E3C project: Collection and annotation of a multilingual Corpus of Clinical Cases, in: Proceedings of CLiC-it 2020, 2021, pp. 1–7.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of NIPS, 2017, pp. 6000–6010.

[12] G. López-García, J. M. Jerez, N. Ribelles, E. Alba, F. J. Veredas, Detection of tumor morphology mentions in clinical reports in spanish using transformers, in: Proceedings of IWANN, 2021, pp. 24—-35.

[13] C. P. Carrino, J. Llop, M. Pàmies, A. Gutiérrez-Fandiño, J. Armengol-Estapé, J. Silveira-Ocampo, A. Valencia, A. Gonzalez-Agirre, M. Villegas, Pretrained biomedical language models for clinical NLP in Spanish, in: Proceedings of BioNLP, 2022, pp. 193–199.

[14] N. Perez, P. Accuosto, À. Bravo, M. Cuadros, E. Martínez-Garcia, H. Saggion, G. Rigau, Cross-lingual semantic annotation of biomedical literature: experiments in Spanish and English, Bioinformatics 36 (2019) 1872–1880.

[15] S. Lima-López, N. Perez, M. Cuadros, G. Rigau, NUBes: A corpus of negation and uncertainty in Spanish clinical texts, in: Proceedings of LREC, 2020, pp. 5772–5781.

[16] A. García-Pablos, N. Perez, M. Cuadros, Vicomtech at eHealth-KD challenge 2021: Deep learning approaches to model health-related text in Spanish, in: Proceedings of IberLEF, 2021, pp. 712–724.

[17] P. J. Vivancos-Vicente, J. A. García-Díaz, J. S. Castejón-Garrido, R. Valencia-García, ISMR - Sistema basado en Deep Learning para la transcripción y extracción de conocimiento en entrevistas médico-paciente, in: Proceedings of SEPLN-PD, 2021, pp. 1–4.