Short paper abstract for DHBenelux 2022

Challenges on the Promising Road to Automatic Speech Recognition of Privacy-Sensitive Dutch Doctor-Patient Consultation Recordings

Berrie van der Molen (Utrecht University), Cristian Tejedor-García (Radboud University), Henk van den Heuvel (Radboud University), Roeland Ordelman (Sound & Vision), Toine Pieters (Utrecht University), Sandra van Dulmen (Nivel), Arjan van Hessen (University of Twente)

Keywords

Automatic Speech Recognition, domain adaptation, language modeling, machine learning, medical discourse, data sensitivity

Abstract

In this paper we present the currently running PDI-SSH project Homo Medicinalis (HoMed), in which we use machine learning to build an Automatic Speech Recognition (ASR) infrastructure for disclosing privacy-sensitive doctor-patient consultation recordings.

ASR using machine learning has been described as being in a ‘golden age’ (Vipperla 2020: 9). Advances in ASR technology make it possible to search culturally significant digitized audiovisual data archives on a spoken word level (Ordelman and Van Hessen, 2018). This has obvious benefits for the study of any topic over time in a changing socio-cultural context: the digital turn (Nicholson, 2013) that made it possible to study explore millions of newspaper pages using distant reading, might now be followed by a cross-media turn as our audiovisual cultural heritage data archives can be explored in the same manner using ASR (Van der Molen, 2022: 202). However, ASR performance decreases when, acoustically (recording conditions) and/or semantically, training and target domain do not match. Most of the current out-of-the-box commercial systems are intended for generic use (e.g. standard-Dutch, relatively good recording conditions), targeting the natural conversational speech scenario, lacking in technical terms and jargon (Litman et. al, 2018). Therefore, there is an evident need for improving ASR performance when used on in-house data-specific scenarios (Tejedor-García et. al, 2021). In HoMed we aim to do that in our work towards enabling such research beyond the public media data domain to audiovisual recordings from the highly privacy-sensitive medical-professional domain. Our understanding of the use and reputation of medicines can then be based on its coverage in the public sphere (e.g. newspaper, radio, television), but also on the way in which it has been discussed in the professional medical sphere by both medical professionals and patients (doctor-patient recordings).

In this project, the audiovisual health care related data collected over the years at the Netherlands Institute for Health Services Research (Nivel), are used to retrain the currently largest open-source general-purpose ASR system for Dutch (Kaldi NL) for the automatic

transcription of in-house privacy-sensitive doctor-patient consultation recordings. The ASR system will be adapted on both the semantic and the acoustic level, and in this process, the AV-recordings themselves will remain intramural at Nivel. The final ASR component will be made available at Nivel, for research on the Nivel material, and in the digital CLARIAH (Media Suite) research infrastructure, for use alongside the Media Suite's other digital media analysis tools and data collections (Ordelman et al, 2018), especially to adapt ASR for specialized domains with unique terms and jargon. The acoustic and language models developed in the project will also be made available open-source.

The privacy-sensitivity of this material raises specific challenges. The recordings can only be played within the viewing room of the Nivel institute, which means that the recordings and its transcriptions cannot be shared between institutes and project members. We addressed this in the first phase of HoMed by using training data that resembles our use case from a different source: the publicly available online news bulletins called *Medicijnjournaal* from the Dutch Institute for Rational Use of Medicine (IVM). Our current challenge is to develop a privacy-proof method for working with the privacy-sensitive data itself, which means we have to address the extremely limited software and hardware context in the viewing room at Nivel, and the many restrictions in terms of what we are allowed to add to this. A key question is also how we can securely use the sensitive data (or derivatives such as acoustic features) for training the language and acoustic models of the proposed ASR engines in a location external to Nivel in this process. In this process, we worked on versions of a protocol for orthographic transcription of: 1) medicine news bulletins, and 2) doctor-patient consultations.

The ultimate project goal is to deliver: 1) an ASR infrastructure tailor-made for doctor-patient consultation recordings that outperforms Kaldi_NL, and 2) a method on developing domain specific ASR engines with models that can be publicly shared based on sensitive audio data that cannot leave the premises of the owner. Another important aim is to accommodate and further develop the leveled approach (Van der Molen et al, 2017), a methodological approach that enables structural cross-media discourse analysis shifting between distant reading technologies and close reading, using keyword search, timeline graphs and word clouds. Challenges with regards to query building are expected when this approach is used to research doctor-patient consultations: even if the infrastructure will successfully recognize the often-mispronounced medical terms, it is not certain that using these terms in a keyword query is the most effective way to find out about the reputation of specific medicines in doctor-patient consultations. Strategies might have to incorporate colloquial language used to describe medicines (e.g. "pufje", "blauwe pillen") or queries built around likely effects (e.g. nausea to search for chemotherapy). We also see potential for incorporating a linked data approach to strengthen the leveled approach in this context.

References

Kaldi NL (Open Spraaktechnologie, 2022) <https://github.com/opensource-spraakherkenning-nl/Kaldi_NL> [accessed 11 February 2022]

Litman, Diane, Helmer Strik, and Gad S. Lim, 'Speech Technologies and the Assessment of Second Language Speaking: Approaches, Challenges, and Opportunities', *Language Assessment Quarterly*, 15.3 (2018), 294–309 <https://doi.org/10.1080/15434303.2018.1472265>

'Media Suite', *CLARIAH* <https://mediasuite.clariah.nl/> [accessed 8 April 2021]

van der Molen, Berrie, Lars Buitinck, and Toine Pieters, 'The Leveled Approach. Using and Evaluating Text Mining Tools AVResearcherXL and Texcavator for Historical Research on Public Perceptions of Drugs', *ArXiv:1701.00487 [Cs]*, 2017 <http://arxiv.org/abs/1701.00487> [accessed 24 January 2017]

van der Molen, Berrie, 'Talking XTC. Drug Discourse in Post-War Dutch Newspaper and Radio Debates.' (Universiteit Utrecht, 2022)

Nicholson, Bob, 'The Digital Turn: Exploring the Methodological Perspectives of Digital Newspaper Archives', *Media History*, 19.1 (2013), 59–73 <https://doi.org/10.1080/13688804.2012.752963>

Ordelman, R.J.F., and A.J. van Hessen, 'Speech Recognition and Scholarly Research: Usability and Sustainability', ed. by I. Skadina and M. Eskevich, *CLARIN 2018 Annual Conference*, 2018, 163–68

Ordelman, Roeland, Liliana Melgar, Jasmijn van Gorp, and Julia Noordegraaf, 'Media Suite. Unlocking Audiovisual Archives for Mixed Media Scholarly Research', *Selected Papers from the CLARIN Annual Conference 2018, Pisa, 8-10 October 2018*, Linköping Electronic Conference Proceedings, 2018, 133–43

Tejedor-García, Cristian, Valentín Cardeñoso-Payo, and David Escudero-Mancebo, 'Automatic Speech Recognition (ASR) Systems Applied to Pronunciation Assessment of L2 Spanish for Japanese Speakers', *Applied Sciences*, 11.15 (2021), 6695 <https://doi.org/10.3390/app11156695>

Vipperla, Ravichander, Samin Ishtiaq, Rui Li, Sourav Bhattacharya, Ilias Leontiadis, and Nicholas D. Lane, 'Learning to Listen... On-Device: Present and Future Perspectives of on-Device ASR', *GetMobile: Mobile Computing and Communications*, 23.4 (2020), 5–9 <https://doi.org/10.1145/3400713.3400715>