Speech-based Medical Decision Support in VR using a Deep Neural Network
(Demonstration)

Alexander Prange, Michael Barz, Daniel Sonntag
German Research Center for Artificial Intelligence, DFKI, Saarbrücken, Germany
alexander.prange@dfki.de, michael.barz@dfki.de, daniel.sonntag@dfki.de

Abstract

We present a speech dialogue system that facilitates medical decision support for doctors in a virtual reality (VR) application. The therapy prediction is based on a recurrent neural network model that incorporates the examination history of patients. A central supervised patient database provides input to our predictive model and allows us, first, to add new examination reports by a pen-based mobile application on-the-fly, and second, to get therapy prediction results in real-time. This demo includes a visualisation of patient records, radiology image data, and the therapy prediction results in VR.

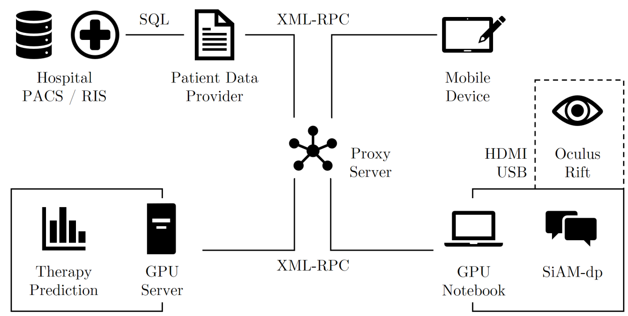

Figure 1: Architecture diagram

1 Introduction

In this demo paper, we present a novel real-time decision support system for the radiology domain, where the radiologist can visualise and interact with patient data in a VR environment by using natural speech and hand gestures. Our multimodal dialogue system is an extension of previous work [Luxenburger *et al.*, 2016] where we used an Oculus Rift DK2 with integrated eye tracking technology[1] in a remote collaboration setting. In this scenario the radiologist starts with a patient examination form he or she fills in by using a mobile tablet with integrated stylus. The data are transcribed using handwriting recognition, then analysed and stored based on common medical ontologies [Sonntag *et al.*, 2009]. In our scenario, the doctor uses the VR application and interacts with patient records using our dialogue system. The main contribution is the end-to-end system by providing a GPU-accelerated model for automated decision support that computes therapy predictions in real-time.

2 Medical Dialogue System

We implemented a *Patient Data Provider* service for connecting to *PACS* (picture archiving and communication system) and *RIS* (radiology information system). The mobile application retrieves patient data and medical images through this interface and sends back completed reports from the radiologist. As depicted in figure 1, all devices and services are connected through a central hub, the *Proxy Server*. Services can be registered through our API, and clients can use these services using the XML-RPC protocol. This video[2] shows the complete workflow.

For real-time data acquisition, we use a mobile device with a digital pen [Sonntag *et al.*, 2014]. The speech-based dialogue system supports three basic types of commands: interaction with the patient data shown on the virtual display (e.g., *"Open the patient file for Gerda Meier.", "Show the next page."*); question answering functionality about factiod contents of a patient record (e.g., *"When was the last examination?"*); and the RNN-based therapy prediction component (*"Which therapy is recommended?"*). In this demo, we implemented the following linguistic phenomena:

- Anaphoric reference resolution: *"What is her current medication?"*
- Elliptic speech input: *"... and the age?"*
- Multimodal deictic reference resolution: *"Zoom in here"* [user points on a region on the display]
- Cross-modal reference resolution (e.g., *"What is the second best therapy recommendation?"*)

We used SiAM-dp [Neßelrath, 2015], an open development platform and workbench for multimodal dialogue systems, for implementation. We integrated the dialogue system into the VR application and connected our backend services.

[1] Oculus Rift DK 2 with SMI Eye Tracking Upgrade

[2] link to video

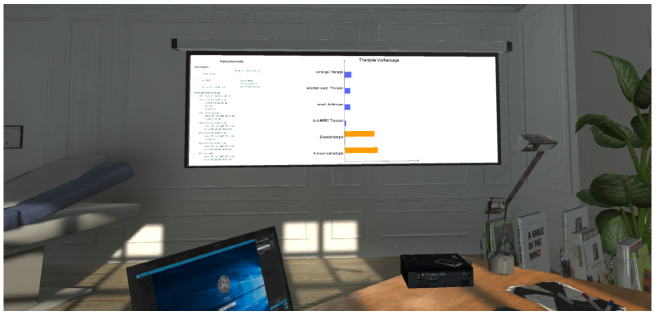

Figure 2: Screenshot of therapy prediction results in VR

3 Decision Support Model

Our medical dialogue system facilitates the doctor's decision about which therapy is most suitable for a given patient when examining the patient history. We integrated a prediction model for clinical decision support based on deep learning [Esteban *et al.*, 2016] as backend service, running on a dedicated GPU server. [Esteban *et al.*, 2016] presented a recurrent neural network (RNN) to include sequences of examinations which was modified to take new patient data as additional input. We used their proposed model (RNN/GRU + static input) trained on an anonymised dataset of 475 real patients, containing a total of 19438 diagnoses, 15352 procedures, 59202 laboratory results and 13190 medications from a clinical data intelligence project [Sonntag *et al.*, 2016]. Personal data, such as names, date of birth, and patient identifier were anonymised, and all date and time data were shifted.

For our speech-based multimodal dialogue system, fast response times are of particular interest. We use TensorFlow to enable GPU-accelerated predictions on a scalable platform. Tests show prediction time capabilities in the 100 milliseconds range. Our service runs on a high-performance computer and is accessible to the dialogue system through our Proxy Server.

In our current scenario we created a Unity3D application[3] that resembles a doctor's office. The user can move inside the room (by position tracking) and look around using head tracking. A virtual screen at the wall is used to show the patient file, the previously annotated digital form, or the therapy prediction (see figure 2). The image viewer is currently limited to 2D visualisations, however, for future work we plan to include 3D medical image data. Navigation inside the documents, like zooming or scrolling, is achieved by multimodal interaction, i.e., natural speech and the Oculus Touch controllers.

4 Conclusions and Future Work

A multimodal dialogue system in combination with data visualisation in virtual reality provides an intuitive and immersive user experience for the doctor. Currently, we are exploring how displaying complex 3D medical images (e.g., DICOM) in virtual reality can improve the diagnostic process. Future work includes additional modalities such as eye-tracking to improve the interaction in VR.

[3]http://unity3d.com/

Speech dialogue extensions include ambiguity resolution by asking clarification questions. If a patient name in a user input is ambiguous, the system could proactively ask for clarification (U: "Open the patient file of Mrs. Meier." S: "Gerda Mayer or Anna Maier?"). In addition, users should be able to change or add patient data through natural speech when applicable [Sonntag *et al.*, 2012].

Acknowledgments

This research is part of the project "clinical data intelligence" (KDI) which is founded by the Federal Ministry for Economic Affairs and Energy (BMWi). Further, we'd like to thank Yinchong Yang and Volker Tresp for providing their therapy prediction model and for their guidance during the model integration process.

References

[Esteban *et al.*, 2016] C. Esteban, O. Staeck, S. Baier, Y. Yang, and V. Tresp. Predicting clinical events by combining static and dynamic information using recurrent neural networks. In *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 93–101, Oct 2016.

[Luxenburger *et al.*, 2016] Andreas Luxenburger, Alexander Prange, Mohammad Mehdi Moniri, and Daniel Sonntag. Medicalvr: Towards medical remote collaboration using virtual reality. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, pages 321–324, New York, NY, USA, 2016. ACM.

[Neßelrath, 2015] Robert Neßelrath. *SiAM-dp : An open development platform for massively multimodal dialogue systems in cyber-physical environments*. PhD thesis, Universität des Saarlandes, Postfach 151141, 66041 Saarbrücken, December 2015.

[Sonntag *et al.*, 2009] Daniel Sonntag, Pinar Wennerberg, Paul Buitelaar, and Sonja Zillner. Pillars of ontology treatment in the medical domain. *J. Cases on Inf. Techn.*, 11(4):47–73, 2009.

[Sonntag *et al.*, 2012] Daniel Sonntag, Christian Husodo-Schulz, Christian Reuschling, and Luis Galárraga. Radspeech's mobile dialogue system for radiologists. In *17th International Conference on Intelligent User Interfaces, IUI '12, Lisbon, Portugal, February 14-17, 2012*, pages 317–318. ACM, 2012.

[Sonntag *et al.*, 2014] Daniel Sonntag, Markus Weber, Alexander Cavallaro, and Matthias Hammon. Integrating digital pens in breast imaging for instant knowledge acquisition. *AI Magazine*, 35(1):26–37, 2014.

[Sonntag *et al.*, 2016] Daniel Sonntag, Volker Tresp, Sonja Zillner, Alexander Cavallaro, Matthias Hammon, André Reis, Peter A. Fasching, Martin Sedlmayr, Thomas Ganslandt, Hans-Ulrich Prokosch, Klemens Budde, Danilo Schmidt, Carl Hinrichs, Thomas Wittenberg, Philipp Daumke, and Patricia G. Oppelt. The clinical data intelligence project. *Informatik-Spektrum*, 39(4):290–300, 2016.