zh aw **School of Engineering**

InIT Institut für angewandte Informationstechnologie

Spracherkennungssystem für SCAI

Conversational agents provide a dialog-based user interface using written or spoken natural language.

For spoken conversational agent, a reliable speech recognition system is essential. In this thesis different aspects of automatic speech recognition (ASR) for conversational agent for legal research are explored.

To train such ASR systems, huge data from the legal domain is necessary. As no such dataset exist for the Swiss legal domain two different ways to collect sample legal queries were explored. With a grammar-based approach, sample legal queries were randomly generated based on hand-crafted rules. Another approach investigated was neural question generation from a text body of knowledge. As this approach didn't generate too much senseless queries it was dropped.

With the first approach 3 datasets of sample queries for three types of legal queries were randomly generated and spoken by one speaker.

These dataset were used to compare the best available commercial online ASR system (Google) with two state-of-the-art open-source research systems (HMM/TDNN Kaldi, End2End wave2letter). The Google system performed with 17% and 24% WER (word error rate) quite well on two of the data sets comprising a lot of legal terms. On the 3rd data set with a lot of paragraph number sequences the system performed much worse with over 38% WER.

The offline research systems performed extremely poor with 80%-130% WER on all three data sets without a specific language model. When a specific n-gram language model was trained with the sample legal utterances the performance of the Kaldi systems improved drastically to below 17% for the 1st data set and 7.4% for the second data set which is ten times less then the WER of the Google system. The performance on the 3rd data set was with 34% still poorer than the Google system. The wave2letter system did much less benefit from the langage model with WER between 24%-51%.

In order to further compare hybrid HMM/DNN systems with End2End sytems different availavble German data sets (500h) were collected. The best Kaldi system trained on that dataset still outperformed the best Jasper system (End2End, character based, beam search, with 6-gram LM) trained of the same data by 10-30% relative lower WER. Byte-Pair-Encoding (BPE) did'nt further improve the Jasper system.

Diplomand/in
Matthias Büchi

Dozent/in
Hans-Peter Hutter

Bild klein 1.

Bild klein 2.