LVCSR-based Speech Analytics of a Hungarian Language Call-Center

Gellért Sárosi(1), Tibor Fegyó (1,2), Péter Mihajlik(1,3), Balázs Tarján (1), Judit Pancza (4), and Zoltán Hans (5)

(1) Department of Telecommunication and Media Informatics,
Budapest University of Technology and Economics, Budapest, Hungary
(2) Aitia International Inc., Hungary
(3) THINKTech Research Center, Nonprofit LLC, Hungary
(4) Statistical Products Hungary Ltd.
(5) AEGON Hungary

{sarosi, mihajlik, fegyo, tarjanb}@tmit.bme.hu, jpancza@spss.hu, hans.zoltan@aegon.hu

Abstract

Customer services record vast amount of voice conversations. This unstructured data has a great potential to predict customers intentions, identify dissatisfied customers, or to monitor the operators' effectiveness. In the exploitation of these spoken language resources speech recognition technology plays a key role. To our knowledge, in this paper the first operating Hungarian speech analytics solution is introduced that is based on LVCSR (Large Vocabulary Continuous Speech Recognition) technology. Our system integrates speech-to-text processing, text analytics and data mining technologies. We show that relevant business benefits can be gained by using rapidly applicable techniques.

Hungarian is considered a difficult language due to its agglutinations and inflections. The high morphological complexity can cause several known problems is LVCSR: high out of vocabulary rate, huge vocabulary size and weak language model parameter estimation. However, as our previous studies suggest [1], in case of spontaneous speech – especially with narrow topics – classical word-based acoustic and language modeling techniques can perform reasonably well even for morphologically rich languages. So, regarding specialized call center "speech-to-text" conversion tasks, the main problem can be language resource acquisition if a new language comes into picture. We decided to collect only in-domain data for our Hungarian language call center related task – about 40 hours were transcribed manually for speech recognizer training and 5 hours for test of the customer service conversations. Further 200 hours of untranscribed speech were used for unsupervised training.

The speech recognition system was based on cross-word tied-state triphone GMM (Gaussian Mixture Model) acoustic modeling, and word 3-gram language modeling. In our case no vocabulary cutoff could be applied (resulting in a 20k vocabulary size) since we wanted to preserve all the rare word forms which abounded in the lexicon. Pronunciations were derived using simple grapheme-phoneme rules and exception dictionary, although trivial grapheme-based acoustic modeling gave competitive performance on a similar task [2]. Customer service recordings usually contain large amounts of overlapping speech and various noises. Despite these facts and the obvious undertraining, we have achieved acceptable, 53.8% word and 75.1% letter baseline recognition accuracies. Using unsupervised language model adaptation the recognition accuracies significantly improved. Thanks to the relatively small number of model parameters and to our effective WFST (Weighted Finite State Transducer) based decoding approach, 4 minutes of speech could be processed by one CPU core in one minute in average.

Several information extraction applications were and could be fed by the output of speech-to-text module. Topic identification was performed on the speech-to-text output resulting in the following Precision/Recall values on three topics: claim – P:90%, R:86%, finance – P:75%, R:89%, and cancellation – P: 72%, R: 96%. So, for example, intentions regarding service cancellation can be predicted from complaints. Thus, based on automatic alerts, preventing steps can be made towards better customer retention.

The performance of operators was also monitored closely by the system in two layers. One layer is protocol compliance monitoring, where related events were identified beyond 95% accuracy. On the other layer the tone of the operator was controlled by observing negative textual terms and patterns. The applicability of operator performance measures were clearly validated by showing the positive effects of an operator training programme. Other possible applications are the identification of reversible terminations, real-time monitoring of special offer campaigns, daily, weekly, monthly reports of customer drop-out with or without operator rankings, telesales support, and fast feedback for crisis communication.

Index Terms: customer service, automatic speech recognition, text analytics, data mining, customer retention, protocol compliance

References

1. Tarján, B., Mihajlik, P., On Morph-based LVCSR Improvements. *In proc. SLTU 2010*, Penang, Malaysia, *pp. 10-16.*

2. Mihajlik, P., Tuske, Z., Tarjan, B., Nemeth, B., Fegyo, T., Improved Recognition of Spontaneous Hungarian Speech: Morphological and Acoustic Modeling Techniques for a Less Resourced Task. *IEEE Trans. of Audio Speech and Language Proc.* 18:(6) *pp. 1588-1600.* (2010)