Human Language Technologies – The Baltic Perspective
A. Utka et al. (Eds.)

doi:10.3233/978-1-61499-442-8-40

Medical – pharmaceutical information system with recognition of Lithuanian voice commands

Vytautas RUDŽIONIS[a,1], Gailius RAŠKINIS[b], Kastytis RATKEVIČIUS[c], Algimantas RUDŽIONIS[c], Gintarė BARTIŠIŪTĖ[c]
[a] *Vilnius University, Kaunas faculty*
[b] *Vytautas Magnus University, Kaunas*
[c] *Kaunas University of Technology*

Abstract. This paper presents a Lithuanian voice recognition system of medical - pharmaceutical terms. The system consists of two separate speech recognition modules working in parallel. One recognizer is a proprietary CD-HMM Lithuanian speech recognizer. The second recognizer is a Spanish speech recognizer adapted to recognize Lithuanian voice commands. The outputs of both recognizers are combined by the decision making block yielding the final decision. The decision making block was automatically derived by an induction algorithm that learns a set of symbolic rules. The investigations showed that both recognizers produce uncorrelated outputs and could complement each other. The investigations also showed that Lithuanian speech recognizer achieves higher accuracy (over 96 percent in a speaker independent mode) but the use of the adapted foreign language recognizer allows increase this baseline accuracy even further (over 98 percent in a speaker independent mode for 1000 voice commands). The voice recognition system is in the process of being embedded into several medical information systems which will be used by healthcare practitioners.

Keywords: speech recognition, foreign language recognizer, Lithuanian speech recognition system, software with voice user interface

Introduction

Healthcare industry is one of the areas in which the automatic speech recognition (ASR) systems show the most promising practical results. Healthcare industry requires many routine operations of documentation. Voice processing technologies allow save time of the highly qualified medical personnel which is spent on these operations. There are many successful examples of implementation of ASR systems in several countries. For instance, Providence Health & Services, which is among the 20 largest healthcare services providers in North America, is using speech recognition for everyday medical documentation processes in 27 hospitals and 250 clinics for approximately 8,000 clinicians since 2012 [1]. The financial benefits of voice recognition is well illustrated by the example of the Dragon Dictate software which

---

[1] Vytautas Rudzionis, Vilnius University Kaunas Faculty, Lithuania, vytautas.rudzionis@khf.vu.lt.

allowed to save the Norman Regional Health System (single medical center based in Oklahoma Cit) $1.8 million in transcription services in 2013 alone [2].

Analysis of the foreign experience and the desire to advance Lithuanian speech recognition suggested us to develop a Lithuanian ASR system for the medical – pharmaceutical domain. The objective was to achieve the level of accuracy necessary for practical implementations (at least 95% of correct recognition). As a compromise the vocabulary was limited to the set of 1000 phrases. It consisted of the titles of the most popular drugs, names of the most frequent diseases and patient complaints obtained from medical practitioners and pharmaceutical companies.

For this purpose we proposed and implemented a hybrid ASR system. Under the term hybrid we understand the system based on two or more constituent speech recognizers working in parallel. Our system was based on two recognizers: a proprietary continuous density Hidden Markov model based (CD-HMM) Lithuanian recognizer and a commercial foreign language recognizer adapted to recognize Lithuanian voice commands. The decisions of both recognizers were combined to produce the final decision of the hybrid ASR system. The rationale behind such design was the assumption that the proprietary Lithuanian recognizer should have better discrimination capability if trained properly whereas the foreign language recognizer allows exploiting information from huge amounts of training.

This paper is organized as follows. We present some considerations and design issues of the hybrid ASR system in the section 1. Some results of experimental evaluation of the system performance are presented in the section 2. Some conclusions and observations about the future work are given in the final section.

1. The architecture of the speech recognition system

Though medical information systems with ASR capabilities are widely used and may yield significant economic benefits, nearly all successful use cases are found in the countries with widely spoken languages (e.g. English, Spanish, Japanese, etc.). Successful practical implementations are much rare in the countries were lesser spoken languages are used. The primary reason of this is the limited availability of speech corpora to train acoustic model in these countries.

If some target language lacks extensive speech resources, two different approaches could be used for building an ASR system. The first approach is to develop acoustic models for the target language from scratch and to develop a proprietary ASR engine on their basis. The second approach is to adapt the existing ASR engine that is based on the acoustic models of some foreign language, to the target language.

The first approach requires the speech corpus with all necessary acoustic examples of the target language. This approach has potentially higher capabilities than the second one. However the main drawback of the first approach is the high associated cost. The second approach (called the crosslingual speech recognition) has the potential to achieve some practically acceptable results. The crosslingual speech recognition aims to transfer the existing source acoustic models from source language to the target language without using large speech corpora in target language and without full retraining of the speech recognition system. Several methods to adapt a foreign language ASR engine for the recognition of voice commands in a different target language were proposed. They are based on similarity measures and grouped into expert-driven and data-driven ones [3].

In expert-driven methods the mapping from one language to another is performed using human knowledge and is typically based on some acoustic-phonetic characteristics of sounds in both languages. One of the most frequently used methods is the use of so called IPA scheme. This method may be summarized as follows: for each phoneme in a target language an equivalent phoneme in the source language is selected. In the case when IPA equivalent is non-existent in the target language the most similar phoneme is selected. Data-driven crosslingual speech recognition approaches are based on data-driven similarity measures. In these methods the similarity measure is applied during mapping.

For several years we carried out a series of experiments using the adaptation procedure which is similar to the expert driven approach described above. We used Microsoft SAPI tools and several foreign language ASR engines trying to find the best set of transcriptions to recognize Lithuanian voice commands [4], [5]. Experiments showed that Spanish ASR engine was among the most accurate ones for Lithuanian commands recognition. The method showed the potential of rapidly developing speech recognition systems with limited vocabulary and high accuracy for some applications.

These investigations led to the next step of building a hybrid ASR system. The hybrid ASR system consists of two different speech recognizers working in parallel. One of those recognizers is the adapted Microsoft Spanish speech recognition engine. The second one is proprietary Lithuanian CD-HMM based recognizer. The decisions of both recognizers are combined together to get the final decision. It should be noted that many modern speech recognizers are using several recognizers working in parallel [6]. The well known Ripper automatic rule induction algorithm [7] has been selected as the robust and easy to use method to combine the outputs.

2. Experimental evaluation of Lithuanian medical terms recognition system

This chapter presents some of the experiments that were carried out to evaluate the possibilities of the hybrid ASR system and to demonstrate the efficiency of the proposed approach.

2.1. Speech corpora for experimental evaluation

It has been decided that voice user interface must be able to deal with the titles of the drugs, with the names of the illnesses and diseases as well as with some other words and phrases that are often met in the practice of healthcare professionals. The large part of medical terms used in practice by healthcare professionals is listed in the official register of diseases and disorders approved by the Ministry of Health (14179 diseases and disorders). It is composed of more than 88000 lexical tokens (not all are of medical origin) and has 10955 unique lexical types. 5991 lexical types cover 75 percent of the whole register list. Analysis showed that a large part of the daily voice requests could be successfully handled using a relatively small number of voice commands. We selected 731 names of diseases, patient complaints and drugs contained. Each voice command in the set has been recorded by 7 male speakers and 5 female speakers in a relatively silent environment. Each voice command was pronounced 20 times. Thus there were 240 utterances of each voice command. The size of the medical speech corpus was about 100 hours. In addition, we used the speech resources already at our possession (containing about 50 hours of speech) to train Lithuanian acoustic models.

2.2. Rationale for hybrid approach

It is reasonable to use hybrid approach if the performance of the constituent speech recognizers is uncorrelated. In the first stage we tried to compare the errors made by Spanish and Lithuanian speech recognizers without thorough optimization of their performance. Table 1 presents the results of this initial evaluation. 731 voice commands were used for testing recognition accuracy.

Table 1. Recognition accuracy of not optimized adapted Spanish and Lithuanian speech recognizers

Vocabulary	Number of utterances to test	Correct recognition of adapted Spanish recognizer, (%)	Correct recognition of Lithuanian speech recognizer, (%)
Diseases (217)	52080	91.8	99.65
Patient complaints (208)	49920	85.3	99.64
Drug names (306)	73440	74.5	98.24
average	175440	82.7	99.06

The most interesting result was the comparison of errors made by both recognizers. Table 2 breaks the combinations of decisions that are produced by both recognizers into subsets.

Table 2. Subsets of data used for decision rule training

Subset ID	Description	Number of utterances
T=T	Both recognizers produce the same decision and both decisions are correct	135898
F=F	Both recognizers produce the same decision and both decisions are incorrect	178
T-	Recognizer LT produces correct decision while recognizer SP doesn‘t produce any decision	3398
F-	Recognizer LT produces incorrect decision while recognizer SP doesn‘t produce any decision	48
-T	Recognizer SP produces correct decision while recognizer LT doesn‘t produce any decision	7
-F	Recognizer SP produces incorrect decision while recognizer LT doesn‘t produce any decision	1
--	Both recognizers do not produce any decision	1
TF	Recognizers produce different decisions, LT produces correct decision	33650
FT	Recognizers produce different decisions, SP produces correct decision	1357
FF	Recognizers produce different decisions, both are incorrect	902

Microsoft’s adapted Spanish recognizer was able to match 1357 erroneous decisions of the Lithuanian recognizer with correct decisions. This constituted about 70% of total errors of Lithuanian recognizer. Such situation was observed for 190 voice commands out of 731 used in tests.

2.3. Implementation of hybrid recognizer

The realization of the hybrid recognizer is an open and still not well investigated question. This is especially true when we need to combine the results obtained from so different recognizers as an adapted foreign language ASR engine and the CD-HMM based triphone Lithuanian recognizer. In these experiments enlarged up to 1000 Lithuanian medical voice commands vocabulary has been used.

The most important part of the hybrid ASR system is the block that makes the final decision by combining decisions of different constituent recognizers. In our case, this block consisted of a rule set. This rule set was learnt in a supervised manner on the basis of utterances for which the decisions of the adapted Spanish and proprietary Lithuanian recognizers differed. The supervised training problem was stated as the task to separate two classes: the class where Lithuanian ASR system was right and Spanish was wrong (TF class) and the FT class vice versa. The feature set included features describing the phonetic content of the utterance, logarithmic probabilities assigned to the output utterance by each of the recognizers, logarithmic probabilities assigned to the second best alternative. There were 35007 training instances in total.

This data set was characterized by the significant disproportion among the number of instances belonging to different classes. Class TF had 33650 instances, while the class FT had only 1357 instances. Consequently, the simple “blind” decision rule (if the outputs of the two recognizers differ then use the output of a “better”, i.e. Lithuanian, recognizer) would result in a 96,12% accuracy. Hybrid ASR approach would be useful only if its decision making block would surpass this accuracy level.

Though any supervised machine learning technique could be used to learn the decision making block, we used Ripper rule induction algorithm [7]. This algorithm was selected because of the acceptable time complexity and the intelligibility of its rule set. The ordered list of nearly 40 decision rules was obtained. The accuracy of this rule set was assessed by means of a standard 10-fold cross validation procedure. The recognitions accuracy of 98.73±0.24% was obtained.

2.4. Some applications of Lithuanian medical speech recognition

There are various plans and attempts to embed the Lithuanian speech recognition component for drug titles, patient complaints and diseases described above into practical applications. Together with industrial partners – the software development company developing IT systems for Lithuanian healthcare sector – voice recognition capabilities are planned to be used in the documentation system “Foxus”. Figure 1 presents the main window of the “Foxus” system. Red digits show the stages of introduction of the speech recognition capabilities.

At the first stage the possibility to record patient complaints by voice was added. This is particularly important for physicians working in psychiatry: they need to spend a lot of time recording complaints. Since our ASR system was trained to recognize a large number of the most frequent patient complaints, this subsystem was filled with speech recognition capabilities first. This system records and saves speech data gathered during real work too. Patient complaints, that are not recognized, are registered manually. These utterances will be used for further training of the recognizer. The second step that will be introduced into “Foxus” is the recording and recognition of patient diagnosis by voice using TLK – 10 (one letter and a sequence of digits) disease codes. This code enables the unique identification of patient diagnoses. At the third stage, speech recognition capabilities will be introduced to treatment recommendations. Treatment recommendations essentially represent a finite set of phrases though the word order within these phrases may vary. At the fourth stage it is planned to train the ASR subsystem to recognize phrases describing. The fifth stage is the preparation and printing of the prescription by voice. Currently used recognizer should be extended to recognize a larger set of names.

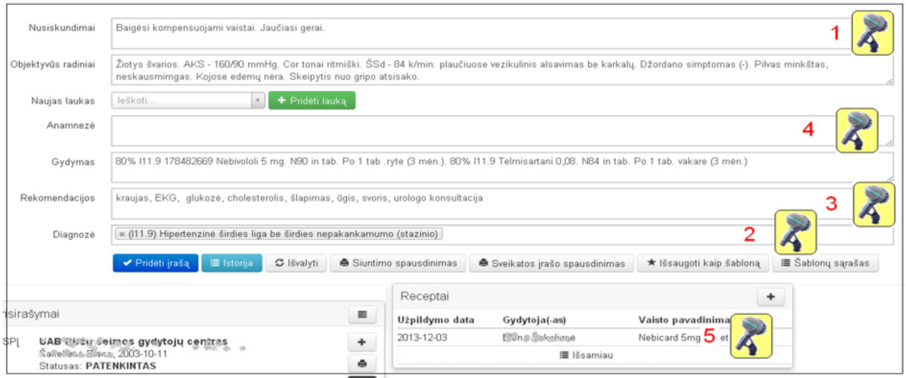

Figure 1. "Foxus" system main interface window.

3. Conclusions and further work

The healthcare is a very promising area to apply voice recognition based user interfaces. Many useful applications could be developed using voice commands recognition approach. Hybrid approach is one of the ways to achieve higher recognition accuracy of a speech processing system. The novel hybrid approach based on learning decision rules by Ripper induction algorithm has been proposed. This method achieved higher recognition accuracy with respect to the "blind" hybridization rule, i.e. with respect to the single best recognizer. The proper training and algorithms enabled to achieve the recognition level acceptable for practical applications for 1000 voice commands. Development of the voice applications in the real world environment was started.

References

[1] Nuance's Voice Recognition Software Deployed at Providence Health & Services' Hospitals, Clinics, *HealthTechZone, July 23, 2012*

[2] D. Carr, "Voice Recognition Speeds HER Use for Oklahoma Hospital", *Information Week Healthcare*, January 3, 2014, *http://www.informationweek.com/healthcare/mobile-and-wireless/voice-recognition-speeds-ehr-use-for-oklahoma-hospital/d/d-id/1113302*

[3] Zgank A. Data driven method for the transfer of source multilingual acoustic models to a new language. Ph.D. thesis, University of Maribor (2003)

[4] Rudzionis V., Maskeliunas R., Rudzionis A., Ratkevicius K: On the Adaptation of Foreign Language Speech Recognition Engines for Lithuanian Speech Recognition. In *Lecture Notes in Business Information Processing* (2009), 21: 113-118

[5] Maskeliunas, R., Ratkevicius, K., Rudzionis, V. Some Aspects of Voice User Interfaces Development for Internet and Computer Control Applications. *Elektronika ir elektrotechnika*,(2013), 19(2), 53-56

[6] G. Saon, J.-T. Chien: Large-Vocabulary Continuous Speech Recognition Systems: A Look at Some Recent Advances. *IEEE Signal Process. Mag*azine, (2012). 29(6): 18-33

[7] Cohen, W.W. Fast Effective Rule Induction. In: *Proceedings of the Twelfth International Conference on Machine Learning*, (1995), pp. 115-123