Speech recognizer for preparing medical reports:

Development experiences of a Hungarian speaker independent continuous speech recognizer

KLÁRA VICSI, SZABOLCS VELKEI, GYÖRGY SZASZÁK, GÁBOR BOROSTYÁN, GÉZA GORDOS

BME, Dept. for Telecommunication and Mediainformatics, Laboratory of Speech Acoustics
{vicsi,szaszak,gordos}@tmit.bme.hu

Reviewed

Keywords: automatic speech recognition, HMM models, n-gram models, bi-gram models, perplexity

A development tool (MKBF 1.0) for constructing continuous speech recognizers has been created under Windows XP. The system is based on a statistical approach (HMM phoneme models, and bi-gram language models with non linear smoothing) and works in real time. The tool is able to construct a middle sized speech recognizer with a vocabulary of 1000-20000 words. New solutions have been developed for the acoustical pre-processing, for the statistical model building of phonemes, and in syntactic level. Through our examination, different training sets were used with different vocabularies. Hungarian is a strongly agglutinative language, in which the number of the word forms is very high. This is the reason why two forms of bi-gram linguistic model were constructed: one is the traditional word forms based and the other is the morpheme based model, in which the vocabulary is much smaller. In this article, test results and the experiences drawn from them are presented. Recognition accuracy has been considerably increased using perplexity based linguistic adaptation.

Introduction

Hungarian belongs to the Finno-Ugric Language family, and – like the other members of this family – is a strongly agglutinative language. The number of different word forms is about hundreds of millions. Word forms are composed by oblique stem and suffixes. In addition, suffixes influence the form of stem in many cases.

The phonetic transcription of written form to spoken one can be easily generated using some rules, but the pronunciation of most words starting or ending with a consonant depends on the adjacent words, because difficult consonant combinations are replaced by simpler ones by a hierarchy of the phonological rules. The situation is more complicated in case of linking morphemes.

In the Laboratory of Speech Acoustics of the Budapest University of Technology and Economics a Hungarian continuous speech recognizer (ASR) has been developed according to the standard knowledge components in a state-of-the art ASR system. These components, the acoustic pre-processing, the acoustic-phonetic model [4] and the syntactic, morpho-syntactic models have been optimalized.

The acoustic pre-processing is the following: the sampling rate is set to 16 kHz, data is coded on 16 bits. The frequency analysis was done in Bark scale (Bark filterbank using 17 bands). The observation sequence vectors, including first order time derivatives are calculated every 10 ms: 17 delta frequency Bark coefficients +17 delta time frequency coefficients +1 energy coefficient are used alltogether.

Phoneme based *acoustic-phonetic models* were used for modeling the Hungarian phonemes. These models are Quasi Continuous Hidden Markov Models (QCMM) with 24 steps, and 5 states. These models were trained and tested by using the Hungarian Reference Speech Database [9]. The test results of the acoustic pre-processing and the phoneme based acoustic-phonetic models were presented in our earlier work [8].

In this article the development of *language models* in syntactic/morpho-syntactic level is presented. Bi-gram models were constructed in two different ways: in the first experimental setup, the basic linguistic units are the word forms; in the second setup, morpheme based bi-gram models were constructed. The training corpus consisted of medical reports collected from the Semmelweis University of Budapest (4000 records) and from the Medical University of Szeged (6365 records) in the field of endoscopy. In the first setup, the vocabulary of the word forms (with 14 331 words) and in the second one, the vocabulary of morphemes (with 6 824 morphemes) together with their pronunciation were prepared based on this corpus. The HUMOR morpheme analyser [5] was used to split the words into morphemes. Medical reports were composed automatically by the computer by recognition of the utterances pronounced by physicians during the examination of patients. These examination reports had been recorded from 5 speakers (4 records from each of these 5 physicians were used). These reports were recorded at the Semmelweis University of Budapest.

1. The bi-gram language model

1.1. Description of the language model

We used a probabilistic language model (LM) based on the assumption that the probability of a word occurence depends on the words preceeding it. If the language model computes the probability of a word

occurrence using the previous n-1 words, it is called an n-gram LM. In practice, language models are usually bi-grams or tri-grams.

The probability of an n-gram is computed from its frequency within a training text, or corpus. In most cases, corpus must be very large.

If the probability of a sequence of words is referred to as $\hat{P}(w_1, w_2, ..., w_m)$ then:

$$P(w_1, w_2, ...w_m) = P(w_1)\prod^{m} P(w_i | w_{i-1} ... w_1) \quad (1)$$

By limiting the context this can be replaced by the following approximation:

$$P(w_1, w_2, ...w_m) \cong P(w_1)\prod_{i=2}^{m} P(w_i | w_{i-1} ... w_{i-n+1}) \quad (2)$$

where $n>0$ is an arbitrary selected integer number.

The probability of a word occurrence using the previous n-1 words:

$$P(w_i | w_{i-1} ... w_{i-n+1}) = \frac{N(w_i ... w_{i-n+1})}{N(w_{i-1} ... w_{i-n+1})} \quad (3)$$

where *N(.)* is the number of occurence for a given word in the training set. We have used bi-gram model in our experiments.

1.2. Smoothing of an N-gram model

The correct estimation of the probability of rare word events is a primary concern in building language models. Generally, the training corpus must be very large in order to ensure that rare words appear at least some times. Instead of increasing the size of the text corpus, different smoothing technics can be used to compensate for data sparsity and to generalize the LM to better model unseen events [6].

We used a non-linear smoothing technique. This smoothing method, based on *absolute discounting,* generally outperforms the others proposed in the literature [7]. Formula applied to the evaluation of the conditional bi-gram probabilities becomes (4):

$$\hat{P}(w_j | w_i) = \max\left\{\frac{N(w_j, w_i) - D_i}{N(w_i)}, 0\right\} + D_i \frac{|V| - n_0(w_i)}{N(w_i)} P(w_i)$$

where $|V|$ is the size of the vocabulary, $n_0(w_i)$ is the number of bi-grams that have the predecessor word w_i and that never occur during the training, $P(w_j)$ is the probability of the unigram w_j, $0 \le D_i \le 1$ is a constant value.

The non-linear interpolated model has some noteworthy properties, interesting by modeling the conditional probabilities. Indeed, if a certain predecessor word is followed by a single word or by a few different words, the effect of the smoothing will be less than in case if the word is followed by many different words. If $D=1$, the events seen only once are handled in the same way as the unseen events.

$$D_i = \frac{|V| \cdot b}{n_0(w_i)}, \qquad \text{where } b = \frac{n_1}{n_1 + 2n_2} \quad (5)$$

Here n_1 and n_2 are the number of bi-grams detected exactly one and two times in the training set. We note that D has an index depending on the predecessor word, i.e. it is constant for all the bi-grams that have the same predecessor.

If the factor $\frac{|V|}{n_0(w_i)}$ in (5) is neglected then D becomes independent of the predecessor word [6].

2. Testing the language model, examination of perplexity

For the training of bi-gram LM models we used the training corpus mentioned in the Intruduction. Training texts were composed of corrected, annotated and phonetically transcribed medical reports collected from two hospitals. This training corpus was divided into 4 groups:

- **G.1** – Gastroscopy reports from Semmelweis University of Budapest, Faculty of Medicine, II. Department of Medicine (Budapest gastroscopy)
- **G.2** – Gastroscopy reports from University of Szeged, Faculty of Medicine, (Szeged gastroscopy)
- **G.3** – Colonoscopy reports from Semmelweis University of Budapest, Faculty of Medicine, II. Department of Medicine (Budapest colonoscopy)
- **G.4** – Colonoscopy reports from University of Szeged, Faculty of Medicine, (Szeged colonoscopy)

These four groups and their combinations were used for the training. For training of the acoustic-phonetic models, the Hungarian Reference Speech Database [9] was used.

2.1. Training conditions

Before training the LM, the vocabulary of the training texts of colonoscopy and gastroscopy were examined. It was found, that only a small part of the vocabularies of Budapest and Szeged reports was common, as it can be seen in the *Table 1* (on next page). The reason for this relatively poor coverage between Budapest and Szeged corpus groups is the use of different expressions for preparing medical reports in the two institutions. Finally, all reports included in the four groups were used together for LM training.

Our later analysis also showed that the vocabularies of the materials given for testing the recognizer contained some new words which were not included in the training corpus.

2.2. Perplexity based WER estimation

The recognition accuracy of speech recognition systems is usually characterized by the *Word Error Rate* (WER) in scientific literature.

The calculation of WER is an expensive process, moreover, it would be very practical to introduce such an indicator that can estimate the recognition accuracy irrespectively of the acoustic-phonetic level. Perplexity is such an estimation method which can help to examine

the language model separated from the acoustic one. The calculation of the perplexity is given by the following equation:

$$PP = \left(\prod_{i=1}^{N} P(w_i \mid w_{i-1}) \right)^{-\frac{1}{N}} \quad (6)$$

where N is the number of words in the test corpus, w_i and w_{i-1} are the i^{th} and $(i-1)^{th}$ words of the test corpus.

In case of a morpheme based LM, words are replaced by morphemes in the above formula. The value of perplexity is a real number greater than 1. The closer this value to 1, the better the coverage of the language model on the given corpus. Too high values refer to a language model that does not really cover the selected test corpus.

3. Test results

To evaluate recognition results, we use some metrics and abbreviations which are:

- ***Ref**: number of units (words or morphemes) to be recognized*
- ***Rec:** number of units recognized*
- ***Corr**: number of units correctly recognized*
- ***Ins:** number of units inserted*
- ***Del:** number of units deleted*
- ***Subs:** number of units substituted*
- ***Accuracy:*** $Acc = \frac{Corr - Ins}{ref}$ (7)
- ***Word Error Rate:*** $WER = 1 - \frac{Corr}{ref}$ (8)

Standard recognition results for word-based LM trained on the whole medical corpus (G.*) are shown in *Tables 2 and 3*. The reason for the poor recognition performance (high word error rates) was found to be that although the LM training corpus covers well the test corpus in terms of vocabulary, but not in terms of bi-gram entries. Typically conjunction words were mostly confused, which is explained by the different reporting standards in the 2 hospitals. By training the LM on the mixed corpus, we also incorporated a high "noise" into the bi-gram field.

If we have a look at Table 1, the differences between the vocabularies of Budapest and Szeged hospitals are obvious. Since speech data used for testing was recorded in Budapest, we trained the LM by using only the G.1 Budapest gastroscopy corpus. Results for this setup are shown in *Table 4*.

Table 2.
Test results for gastroscopy with word unit based bi-gram LM trained on G.1–G.2–G.3 and G.4 mixed; tested on recorded medical reports spoken by physicians.

Ref [# of units]	Rec [# of units]	Corr [# of units]	Ins [# of units]	Del [# of units]	Subs [# of units]	Acc [%]	WER [%]
1173	1580	750	451	22	401	25.4	36.1

Table 3.
Test results for colonoscopy with word unit based bi-gram LM trained on G.1- G.2 - G.3 and G.4 mixed; tested on recorded medical reports spoken by physicians

Ref [# of units]	Rec [# of units]	Corr [# of units]	Ins [# of units]	Del [# of units]	Subs [# of units]	Acc [%]	WER [%]
890	1326	504	822	8	370	-35.7	43.4

Ref [# of units]	Rec [# of units]	Corr [# of units]	Ins [# of units]	Del [# of units]	Subs [# of units]	Acc [%]	WER [%]	PP
1150	1417	799	283	8	343	44.8	30.5	73.59

Table 4.
Test results for gastroscopy with word unit based bi-gram LM trained on G.1, tested on recorded medical reports spoken by physicians

Results presented in Tables 2-4, were obtained by using the utterances of physicians. These records were however relatively noisy, articulation was also poor. Hence, we re-recorded the same 20 reports in a low-noise environment with accurate, standard articulation in order to examine the effect of acoustic quality on recognition performance. By using the same LM trained on G.1 corpus, results are shown in *Table 5.* As it can be seen, WER was reduced considerably. (The reason for the little increase in the number of reference units in *Table 5.* compared to Table 4. is explained by the accurate articulation.)

Ref [# of units]	Rec [# of units]	Corr [# of units]	Ins [# of units]	Del [# of units]	Subs [# of units]	Acc [%]	WER [%]	PP
1173	1451	922	280	1	250	54.7	21.3	73.59

Table 5.
Test results for gastroscopy with word unit based bi-gram LM trained on G.1, tested on medical reports recorded with accurate articulation in low-noise environment

For a special case, we evaluated the recognition performance in case of utterances included in LM training corpus. For this purpose we have chosen 10 reports included in the Budapest gastroscopy (G.1) training corpus, we recorded them in low-noise environment with accurate articulation. Results are shown in *Table 6.*, as expected, WER is much lower, perplexity is also very low. These results can be regarded as theoretical optimum, in this case it was excluded that a missing word from vocabulary, or a forbidden bigram entry influence recognition performance.

Ref [# of units]	Rec [# of units]	Corr [# of units]	Ins [# of units]	Del [# of units]	Subs [# of units]	Acc [%]	WER [%]	PP
416	444	380	28	0	36	84.6	8.6	9.36

Table 6.
Test results for gastroscopy with word unit based bi-gram LM trained on G.1, tested on medical reports included in G.1, recorded with accurate articulation in low-noise environment

3.1. Conclusion for word based tests

Summarizing the results for word based LM testing, it is obvious that Budapest gastroscopy and Budapest colonoscopy reports are very different to the ones of Szeged in terms of expression syntax. A bi-gram LM based on a mixed Budapest-Szeged corpus is more robust and includes more words in the vocabulary, but in practice, recognition performance is worse by 5.54% than in case of a bi-gram trained only on Budapest gastroscopic reports.

Another crucial point is a relatively correct, accurate articulation and a lowered noise level. We should remark that 1 physician of the 5 asked to test the recognizer spoke very low. If we eliminate his 4 reports from the test set WER decreases to 24.27% from 30.52%.

The bi-gram LM trained on G.1 does not cover sufficiently the area recommended by a physician user. This

is verified in the last experiment and seen in Table 6. Moreover, a detailed analysis of errors in the 20 test utterances recorded from physicians and in the 20 utterances recorded with accurate articulation, a very high correlation was found between the errors in the corresponding poor articulation – good articulation speech utterances, which refers to data sparsity problem concerning the LM training data.

Finally, a low rate for *Acc* might refer to an improper coverage of the LM, since in this case the number of insertions increases radically. Inserted units are usually frequent words consisting of one or two syllables, whose vowels correspond to the vowels of the original word misrecognized.

3.2. Perplexity analysis

Relying on perplexity defined in section 1.2., we have seen in section 1.3. that perplexity is occasionally a good predictor of word error rates during recognition. In this section we would like to examine if this prediction is feasible or not. Perplexity is usually used to characterize the language model, but does not deal with the acoustic level: hence, in theory, it is possible that a LM with relatively low perplexity on a corpus yields worse recognition results than a LM of higher perplexity because of eventual acoustic similarity of vocabulary elements.

On *Figure 1*, correlation between perplexity and word error rate is illustrated, measured on G.1 corpus for the original test set of 20 medical reports. As it can be seen, a linear dependence of WER on perplexity can be assumed [1]. The variation of values on Fig.1. is explained by the fact, that only a subset of the Hungarian language is investigated. The other reason for this has been already mentioned above: for some test utterances, acoustical distance between vocabulary elements is various, the closer they are in spectral properties, the more likely is the confusion. This causes the right "shift" of perplexity – WER value pairs seen on Fig.1 [4].

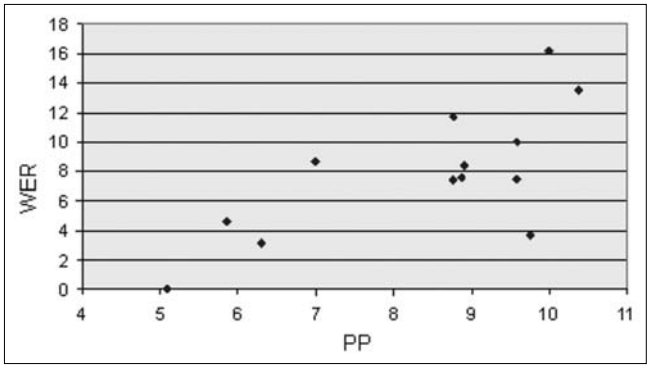

Figure 1. Correlation between perplexity and WER (LM trained on G.1, low noise utterances)

3.3. Word- or morpheme-based LM?

Beyond the "traditional" word-based bi-gram LM, a morpheme-based bi-gram was also prepared. Recognition tests for the morpheme-based LM were carried out with the same conditions as for the word-based one, the obtained results were quite similar for these two cases *(see Table 7)*.

The advantage of the morpheme-based language modeling is a considerable reduction in the size of vocabulary, hence the size of the bi-gram field to store and to use by recognition is also reduced, which is a critical issue, since bi-gram fields are usually stored in the memory during recognition to ensure a real-time operation. Based on the whole LM training corpus, the number of distinct words was found to be 14 331, but the number of distinct morphemes that covered fully the same corpus was only 6 706, less than the half of the number of words. Since the size of the bi-gram field is proportional to $|V|^2$ (the square of vocabulary size), this means that the bi-gram LM for words needs ~4.5 times more storage capacity in our case, moreover, the operation will also be slower. The values of bi-gram probabilities are higher in case of a minor vocabulary, which is also advantageous.

	Ref [# of units]	Rec [# of units]	Corr [# of units]	Ins [# of units]	Del [# of units]	Subs [# of units]	Acc [%]	WER [%]	PP
Word-based LM	1631	2045	1241	778	9	355	28.3	23.9	27.31
Morpheme-based LM	1173	1451	922	280	1	250	54.7	21.3	73.59

Table 7. Recognition results with word-based and morpheme-based LM trained on G.1, using test reports with accurate articulation

The disadvantage of morpheme-based language modeling is the difficult handling of assimilation phenomena across morpheme boundaries. The correct description of these events is not automatically feasible currently. Another problem might be the existence of some very short morphemes difficult to model (e.g. suffix *-t* in Hungarian to express accusative).

4. Expanding the language model based on perplexity measures

Perplexity can be used to predict recognition accuracy with the restrictions presented in section 2.2. We have also shown the difference between test data included directly or not included in the LM (see Tables 6 and 5). The results in Table 6 are obviously better, since the coverage of the LM was ensured for test corpus. In this section we would like to investigate, whether it is possible to increase recognition accuracy by including into the LM not the test corpus, but some typical word sequences from it. We would like to know also, how many times a sequence should be included to get a LM with corresponding weights for these bi-gram entries.

By training of missing word connections we were looking for word sequences that had not been included in the original training corpus G.1. This can be described by the following formalism (here '+' refers to one or more repetitions as in regular expressions):

```
<word included in G.1>
     <word not included in G.1>+
            <word included in G.1>          (9)
```

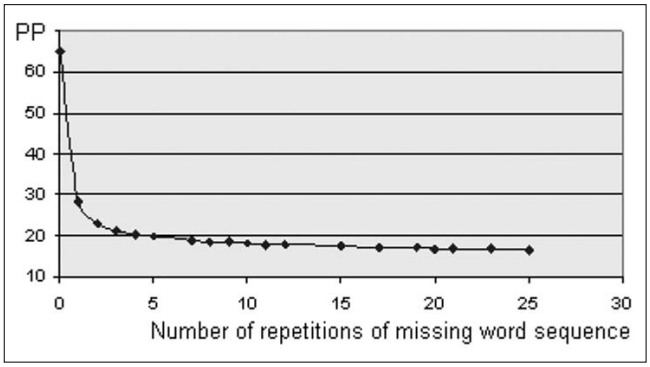

Figure 2.1.
Perplexity for test set depending on the number of repetition of missing word sequences

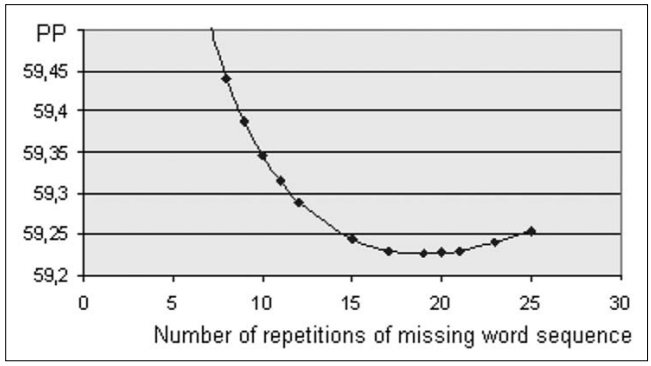

Figure 2.2.
Perplexity for control set depending on the number of repetition of missing word sequences

Our aim is to incorporate the missing bi-gram entries into the training corpus without distorting the actual LM. This explains why each word sequence selected for incorporation begins and ends with words already included in the corpus. Hereby, the context of new items will also be added.

4.1. Bi-gram weight optimization for new items

To determinate the number of times a selected word sequence should be added to the training corpus, we assume that relying on perplexity, we are able to predict recognition performance with the given LM training corpus.

Please note, that perplexity values are not comparable normally, but in this special case, it can be a good predictor because we carry out only minor modifications on the LM training corpus, preserving all other characteristics of data. The expanding process can be formalized like:

```
<original training corpus (G.1)>
    <selected word sequences>*
```
(10)

Table 8.1.
Recognition results for the selected 4 test reports with original (non-expanded) LM

Recognition of reports with LM from Original G.1 training corpus								
Report ID	Ref	Rec	Corr	Ins	Del	Subs	Acc	WER
3	95	116	67	21	0	28	48.4	**29.4**
33	63	80	48	17	0	15	49.2	**23.8**
53	55	68	50	13	0	5	67.2	**9.1**
92	55	62	46	7	0	9	70.9	**16.3**
Average WER:	**19.6%**		Average Acc:	**58.9%**				

Table 8.2.
Recognition results for the selected 4 test reports with expanded LM (20 repetitions of missing word sequences)

Recognition of reports with LM from expanded G.1 training corpus								
Report ID	Ref	Rec	Corr	Ins	Del	Subs	Acc	WER
3	95	110	78	17	1	16	64.2	**17.8**
33	63	69	61	6	0	2	87.3	**3.1**
53	55	58	53	3	0	2	90.9	**3.6**
92	55	61	49	6	0	6	78.1	**10.9**
Average WER:	**8.9%**		Average Acc:	**80.1%**				

Table 9.1.
Recognition results for 3 control test reports with original (non-expanded) LM

Recognition of reports with LM from Original G.1 training corpus								
Report ID	Ref	Rec	Corr	Ins	Del	Subs	Acc	WER
3	51	64	41	13	0	10	54.9	**19.6**
33	34	41	11	11	2	21	0.0	**67.6**
53	118	167	70	49	0	48	17.7	**40.6**
Average WER:	**42.6%**		Average Acc:	**24.2%**				

Table 9.2.
Recognition results for 3 control test reports with expanded LM (20 repetitions of missing word sequences)

Recognition of reports with LM from expanded G.1 training corpus								
Report ID	Ref	Rec	Corr	Ins	Del	Subs	Acc	WER
3	51	64	41	13	0	10	54.9	**19.6**
33	34	41	11	11	2	21	0.0	**67.6**
53	118	166	68	48	0	50	16.9	**42.3**
Average WER:	**43.2%**		Average Acc:	**23.9%**				

where '*' refers to the number of times a word sequence was added to the training corpus.

For testing, 4 medical reports were chosen from the test set. The other 16 test reports were also kept to control whether the LM becomes distorted. After the determination of missing word sequences included in these 4 reports, but missing form the LM training corpus, these word sequences were added, progressively increasing the number of times they were repeated, and controlling whether perplexity measures for the rest of the test reports (16) do not become worse. Our aim is to increase recognition accuracy (predicted by perplexity) for the 4 reports selected and keep or even improve perplexity of the rest 16 test reports.

In *Figure 2.1.* it can be seen that perplexity for the 4 test reports improves by adding the missing word sequences again and again. In parallel, in *Figure 2.2.* perplexity for the 16 control reports improves until 18-20 repetitions, but decreases after. According to Figures 2.1. and 2.2., the final repetition number was set to 20. To control recognition performance, a full test process was carried out again, after the expansion of LM training set. Results can be seen in *Tables 8.1-2. and 9.1-2.*

As expected, recognition results for the 4 selected reports are radically increased. Recognition results without incorporation of missing word sequences are presented in *Table 8.1.*, while after incorporation with 20 repetitions they change according to *Table 8.2.*

In *Tables 9.1. and 9.2.*, recognition performance for some control test reports are presented. As it can be seen, they are only slightly influenced by LM expansion, but there is no evidence of any LM distortion.

5. Conclusion

According to our investigations reviewed in this section, language model expansion (and its implicit re-weighting) is feasible, and by this expansion, the number of times a missing item should be repeated can be determined using perplexity. This procedure does not decrease perplexity and recognition performance for LM. This method can be used in the future to expand LM fast and efficiently. An implementation of a self adaptation (LM tuning for user's profile) algorithm might be a result of our current investigations.

On the other hand, the method we evaluated is not a universal solution to the problem of LM updating. By highly agglutinative languages, like Hungarian, usage of new items by the ASR user is always possible, but for a restricted area, like medical solutions for diagnostics, the method can be suitable to ensure better performance.

References

[1] Máté Szarvas, Sadaoki Furui:
Evaluation of the Stochastic Morpho-syntactic Language Model on a One Million Word Hungarian Dictation Task. Eurospeech, Genova, 2003. pp.2297–2300.

[2] Stanley Chen, Dougles Beeferman, Ronald Rosenfeld:
Evaluation Metrics For Language Models,
In: DARPA'98, National Institue of Standards and Technology (NIST), accessible:
www.nist.gov/speech/publications/darpa98/html/lm30/lm30.htm

[3] Philip Clarkson, Tony Robinson:
Towards improved language model evaluation measures, accessible:
http://Citeseer.ist.psu.edu/clarkson99toward.html

[4] Yonggang Deng, Milind Mahajan, Alex Acero:
Estimating Speech Recognition Error Rate without Acoustic Test Data, accessible:
http://research.microsoft.com/srg/papers/2003-milindm-eurospeech.pdf

[5] HUMOR: Hungarian Morpheme Analyser, accessible:
http://www.morphologic.hu/en_humor.htm

[6] Claudio Becchetti, Lucio Prina Ricotti:
Speech Recognition, Theory and C++ implementation, Fondazione Ugo Bordoni, Rome, 1999.
ISBN 0-471-97730-6

[7] Ney, H., Essen, U., Kneser, R.:
On Structuring Probabilistic Dependencies in Stochastic Language Modeling,
Computer Speech and Language, 1994/8. pp.1–38.

[8] Velkei Szabolcs, Vicsi Klára:
Beszédfelismerő modellépítési kísérletek akusztikai, fonetikai szinten, kórházi leletező beszédfelismerő kifejlesztése céljából
(ASR Model Building Experiments on Acoustic-phonetic Level for a Medical ASR Application),
in Hungarian: II. Magyar Számítógépes Nyelvészet Konferencia 2004.
pp.307–315.

[9] Vicsi Klára, Kocsor András, Teleki Csaba, Tóth László:
Beszédadatbázis irodai számítógép-felhasználói környezetben,
(A Speech Database for Office Environment),
in Hungarian: II. Magyar Számítógépes Nyelvészet Konferencia 2004.
pp.315–319.